The optimal pH of AID is skewed from that of its catalytic pocket by DNA-binding residues and surface charge

Atefeh Ghorbani\textsuperscript{1,2}, Justin J. King\textsuperscript{1,2}, Mani Larijani\textsuperscript{2*}

\textsuperscript{1}Program in Immunology and Infectious Diseases, Department of Biomedical Sciences, Faculty of Medicine, Memorial University of Newfoundland, St. John’s, Newfoundland A1B 3V6, Canada.

\textsuperscript{2}Department of Molecular Biology and Biochemistry, Faculty of Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.

*To whom correspondence should be addressed: Email: Atefeh_Ghorbani@sfu.ca, Justin_King@sfu.ca, Mani_Larijani@sfu.ca
Abstract

Activation-induced cytidine deaminase (AID) is a member of the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) family of cytidine deaminases. AID mutates immunoglobulin loci to initiate secondary antibody diversification. The APOBEC3 (A3) sub-branch mutates viral pathogens in the cytosol and acidic endosomal compartments. Accordingly, AID functions optimally near neutral pH, while most A3s are acid-adapted (optimal pH 5.5-6.5). To gain a structural understanding for this pH disparity, we constructed high-resolution maps of AID catalytic activity vs pH. We found AID’s optimal pH was 7.3 but it retained most (>70%) of the activity at pH 8. Probing of ssDNA-binding residues near the catalytic pocket, key for bending ssDNA into the pocket (e.g R25) yielded mutants with altered pH preference, corroborating previous findings that the equivalent residue in APOBEC3G (H216) underlies its acidic pH preference. AID from bony fish exhibited more basic optimal pH (pH 7.5-8.1) and several R25-equivalent mutants altered pH preference. Comparison of pH optima across the AID/APOBEC3 family revealed an inverse correlation between positive surface charge and overall catalysis. The paralogue with the most robust catalytic activity (APOBEC3A) has the lowest surface charge, most acidic pH preference, while the paralogue with the most lethargic catalytic rate (AID) has the most positive surface charge and highest optimal pH. We suggest one possible mechanism is through surface charge dictating an overall optimal pH that is different from the optimal pH of the catalytic pocket microenvironment. These findings illuminate an additional structural mechanism that regulates AID/APOBEC3 mutagenesis.

Abbreviations: Activation-induced cytidine deaminase (AID), Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC)

Keywords: AID, APOBEC, pH preference, DNA-binding, deaminase
Introduction

Activation-induced cytidine deaminase (AID) is a member of the AID/APOBEC family that converts deoxycytidine (dC) to deoxyuridine (dU). The enzyme family is comprised of eleven paralogues in humans: AID, APOBEC1, APOBEC2, APOBEC3A (A3A), APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3D (A3D), APOBEC3F (A3F), APOBEC3G (A3G), APOBEC3H (A3H), and APOBEC4 (1–3). AID and APOBEC2 are considered the ancestral members of the family, whereas APOBEC1 and APOBEC3 appeared as later derivatives (2,4). AID is primarily expressed in activated B cells and mutates the immunoglobulin (Ig) loci, initiating somatic hypermutation and class switch recombination of antibodies (1,5–7). APOBEC1 mutates apolipoprotein B100 mRNA, introducing a premature stop codon resulting in a truncated Apolipoprotein B48 involved in lipid uptake from the liver and small-intestine (8,9). APOBEC2 expression has been shown to be crucial for muscle development (10,11). APOBEC4 is expressed in the testes and although one study has shown a pro-viral role in HIV-1 infection, its physiological function remains unknown (12,13). Moreover, in contrast to their APOBEC paralogues, APOBEC2 and 4 have failed to demonstrate catalytic activity in eukaryotic/prokaryotic expression systems (13–15) despite APOBEC4’s capacity to bind single-stranded DNA (ssDNA) (13). The APOBEC3 sub-branch (A3) function as restriction factors against viral genomes and mobile endogenous retroelements (16–18). AID/A3s bind ssDNA and RNA, but only deaminate DNA (19–23), except for A3A and A3G which also deaminate RNA in specific cells (24,25). Despite AID/APOBEC beneficial mutations in the context of adaptive and innate immune responses, they are loosely targeted throughout the genome and mutate accessible ssDNA regions liberated during transcription and/or replication (26–29). As such, aberrant AID/APOBEC activity instigates genome-wide mutations and double-stranded breaks
It is well established that off-target AID mutations are involved in tumor development, aggressiveness, and adaptation to chemotherapeutics (30,36–42). More recently, A3A and A3B mutations have been identified in various cancers, including lung- and breast-cancer (43–49). Thus, outside their normal physiological role, AID/A3 expression and activity is associated with tumor initiation and propagation.

AID/APOBEC3 paralogues can be further classified based on the number of zinc-binding catalytic domains. AID, A3A, A3C, and A3H have single zinc-binding domains, while A3B, A3D, A3F, and A3G have two zinc-binding domains (4). The AID/APOBEC3 structures all conserve the same basic core folding of a central β-sheet encompassed by 6-7 α-helices (2,3,50). The catalytic pocket is composed of variable secondary catalytic loops (L1, L3, L5, and L7) (51,52) that surround the catalytic residues. These residues include an arrangement of 3 Zn-coordinating residues (2 Cys, and 1 His) and a catalytic glutamic acid (E) that acts as a critical proton-donor during catalysis (53). Though the core catalytic pocket arrangement and overall folding of AID/APOBEC3s is well conserved, the enzymes differ highly in their surface residue composition and dynamics. One such difference is the shape and location of ssDNA binding grooves: A3s typically have one well-defined ssDNA binding groove (54,55), while AID has evolved several nucleic acid binding regions (19,51,52,56–59). AID contains two ssDNA binding grooves that pass over the catalytic pocket (ssDNA binding groove 1 and 2) (51,52,56,57), an assistant patch located distally from the catalytic pocket that enhances structured DNA binding (57), and an RNA binding region involved in binding switch-region RNA transcripts (58). The DNA/RNA binding regions of AID are predicted to work in concert for binding different structures of ssDNA, ssRNA, G4-quadruplex DNA and RNA, DNA/DNA, and DNA/RNA hybrids (19,51,52,56–58,60). At pH 7 the majority of A3s are near neutral
charge, in contrast to the overwhelmingly positive surface of AID (+1; (51,52,57,61–63). These surface differences underlie various affinities for ssDNA binding and different ssDNA substrate shape preferences across the family (29,51,52,55,61–64,64–69).

Surrounding the active site, the highly variable secondary catalytic loops dictate sequence preference and catalytic rate. While AID prefers to mutate a broad range of WRC (W=A/T, R=A/G) sequences, A3s typically prefer to mutate more specific motifs, such as a TC (A3A/A3B/A3F) and CCC (A3G) (29,54,55,64,64–75). This substrate specificity within the AID/APOBEC family has been largely attributed to L7 with minor contribution from L1 (73,75). Recently, several studies have shown that the catalytic pocket of AID/APOBEC3s can undergo dynamic pocket closure/opening, a phenomenon termed “Schrödinger’s CATalytic pocket” that acts as a structurally inherent regulator of mutagenic activity (51,52,76–78). Thus, within the AID/APOBEC3 family, many of the structural features underlying their role in immunity vs cancer (substrate preference, shape, DNA binding, and dC catalysis) are largely attributed to the surface structure of the variable secondary catalytic loops.

Outside their role in immunity and cancer, the surface composition of the AID/APOBEC3s has also been attributed to other biochemical features within the family. Several studies have examined the pH preference within the AID/APOBEC3 family. A3s were found to have a considerably lower optimal pH (pH 5.5-6.5), characterized by dramatically lower activity at physiological pH of 7.4 (61,79–82), except for A3F catalytic domain (A3F-CTD) which demonstrated a high catalytic rate across pH 5.5-9.5 (81). In contrast, AID has been shown most active near physiological pH (80). The optimal pH of A3G-CTD (pH 5.5-6.5) is influenced by the surface residue H216 on L1 (79). H216R (equivalent to R25 in AID) largely erased this dependency on acidic pH (79), while in A3F, N214H (equivalent to H216 in A3G)
increased activity at pH 5.5 by ~27-fold (81). Cumulatively, this suggests that protonation of this residue (Histidine pK_a = 6.0) is important for binding substrate and catalysis under acidic conditions. In AID and other A3s, this equivalent-residue is located at the “mouth” of the catalytic pocket and plays a pivotal role in anchoring the negatively-charged backbone of ssDNA into a deamination-feasible complex (51,55,55,57,79–81). Similarly, A3A was also shown to have an optimal pH of 5.5, demonstrating a dramatically higher binding affinity and catalysis in acidic vs physiological pH (61,80,82). It was hypothesized that the acidic optimal pH of A3A represents adaptation for enzymatic activity in acidic compartments of the cell, such as lysosomes and phagolysosomes where A3A targets HIV-1 (61).

The aforementioned studies have established that A3s are more acid-adapted in comparison to AID and suggested substrate binding as a major mechanism underlying optimal pH in the AID/APOBEC3 family. However, these studies used pH increments of 0.5-1 across activity buffers and we sought to establish a higher-resolution map of AID’s pH preference. In so doing, we pinpointed the optimal pH of AID and demonstrated it retains most activity at more basic pH. DNA-binding groove mutants localized at the mouth of the catalytic pocket were found to shift the optimal pH and dampen AID’s activity at basic pH. Mutation of the equivalent DNA-binding groove residues in AID orthologs demonstrated more dramatic shifts in optimal pH. And finally, we found AID chimeras involving ssDNA-binding groove 2 and the assistant patch that shift the optimal pH and dramatically alter activity at more basic pH. This work highlights the network of structural regions that determine the optimal pH of AID, with emphasis on the DNA binding grooves. More broadly speaking, we suggest AID/APOBEC3s modify their optimal pH through differences in isoelectric point. These results are important for understanding the evolution of pH adaptation within the AID/APOBEC3 family, amongst
evolutionary-distant AID orthologs, and for engineering acid- or base-stable AID variants suitable for use in cytidine base-editing technologies.

Materials and methods

AID and APOBEC expression and purification:

Wildtype human and orthologous GST-AIDs as well as domain-swapped and point mutant AIDs were expressed and purified as described before (51,64,83–85). Briefly, EcoRI-flanked coding sequences were synthesized by Genescript and cloned into pGEX-5X-3 vector (GE Healthcare, USA) to generate GST-AID expression constructs. GST-AID was induced in DE3 E.coli cells by the addition of 1mM IPTG at 16°C for 16 hours. Bacterial cells were lysed using a French pressure cell press (Thermospectronic) and the expressed GST fusion protein was purified using glutathione Sepharose high performance beads (Amersham). The yield and purity of protein preparations were determined through SDS-PAGE. GST-AID of ~80-90% purity were dialyzed in 100 mM NaCl, 20 mM Tris-HCl pH 7.5, and 1 mM dithiothreitol and stored at -80°C. AID mutants were generated through site-directed mutagenesis. APOBEC3 enzymes were also expressed as GST-fusion proteins by expression of the pcDNA3.1-based vectors in HEK293T cells, as previously described (86,87). Briefly, 25x10 cm plates each containing 5x10^5 cells were transfected with 5 µg of expression plasmid per plate using Polyjet transfection reagent (Froggabio), incubated for 48 hours, collected, and resuspended in 500mM NaCl, 50mM Phosphate Buffer pH 8.2, 0.2mM PMSF, and 50 µg/ml RNAse A. Following lysis in a French pressure cell press (Thermospectronic), centrifugation was carried out to clear the lysate and the supernatant was purified using Glutathione Sepharose high-performance beads (Amersham). Yields and relative purity of the enzymes used were assessed on SDS-PAGE using bovine-serum albumin (BSA) standard curves. APOBEC3 enzymes of ~90% purity were dialyzed in 100mM
NaCl, 20mM Tris-HCl pH 7.5, 1mM DTT, 5% Glycerol, and 50 µg/ml BSA and stored at -80°C.

Two to four independent preparations of each enzyme were purified to obtain average results from multiple independently purified preparations of each enzyme tested.

**Substrate preparation:**

Single-stranded and partially single-stranded bubble oligonucleotides for AID and APOBEC3s were prepared as described previously (64,83–85,87). Briefly, 2.5 pmol of the target strand was 5'-radiolabeled with [γ-^32^P] dATP using PNK enzyme at 37°C for one hour in a volume of 10 µL. To remove the excess free [γ-^32^P] dATP, reactions were purified through mini-Quick spin DNA columns (Roche, Indianapolis, IN, USA). For partially single-stranded bubble oligos, 2-fold excess (5pmol) of the partially complementary strand was annealed to the target strand in a total volume of 50 µL in TE buffer with 100mM KCl to generate the partially single-stranded bubble substrate. AID substrates contained WR_C in the context of bubble substrates, while A3A and A3G contained TT_C and CC_C motifs in single-stranded DNA, respectively.

**Alkaline cleavage deamination assay:**

The standard alkaline cleavage activity assay has been described previously for AID and APOBEC3s (19,51,64,83–85,87). Briefly, 1 µL of substrate (17.5 fmol), 3 µL of AID (~100 ng) and 6 µL of 100 mM phosphate buffer of varying pH (described below) were mixed to a final volume of 10 µL. For the pH range comparisons, the individual reactions only varied in the activity buffer pH. To determine enzyme velocity at several different pH, increasing [TGCbub7 substrate: a 56 nt-long bubble substrate with a 7 nt-long bubble AID-target region, containing the WRC motif TGC] of 1.25, 2.5, 5, 10, 15, 20, 25, 30, 40, 50, 65 and 80 nM were incubated with AID in pH buffer 6.7 (acidic pH representative), 7.99 (basic pH representative) and
corresponding optimal pH, using the same reaction composition as described above. Depending on the optimal activity parameters of each AID/APOBEC, reactions were incubated at its optimal temperature and time, ranging from several minutes to 4 hours, followed by heat inactivation of AID/APOBECs (85°C for 20 minutes) to ensure all measured activity occurred at the precise pH point rather than subsequent uracil glycosylation step which would slightly alter the reaction pH. To remove AID-generated uracil from substrate, 10 μL of UDG reaction including 0.2 μL of UDG enzyme (NEB, USA), 2 μL of UDG buffer and 7.8 μL of milliQ water was added to each reaction followed by incubation at 37°C for 30 minutes. The abasic site was cleaved by incubating the reactions at 96°C for 10 min by addition of 2 μL of 1 M NaOH to a final [NaOH] of 100 mM. To separate the substrate from product, reactions were electrophoresed on a 14% denaturing acrylamide gel. To visualize the result, gels were exposed to a Kodak Storage Phosphor Screen GP (Bio-Rad) and imaged using a PhosphorImager (Bio-Rad, Hercules, CA, USA).

Preparation of enzyme activity assay buffers of finely varying pH increments:

The AID activity buffers are based on the previously described standard reaction buffer for optimal AID activity in the alkaline cleavage assay, which is 100 mM Phosphate buffer, pH 7.2 (19,51,83,84,87). 100 mM Phosphate buffer with pH ranging from 5.8 to 8.0 at 0.1 intervals were prepared using 0.2 M sodium phosphate monobasic and 0.2 M sodium phosphate dibasic solutions. All solutions were made in RNase/DNase free water and filter sterilized. To determine the effective pH in the final alkaline cleavage reaction, phosphate buffer, TE buffer and AID or APOBEC storage buffer were mixed to their final ratio of 6:1:3, to reflect the exact ratio of the three components of the reaction mixture (phosphate activity buffer + substrate in TE + AID/APOBEC) and final pH was measured. Thus, by using reaction buffers with pH ranging
from 5.8 to 8.0 at 0.1 intervals, the final effective alkaline cleavage reaction pH was determined to range from 5.9 to 8.2.

Electromobility Shift Assay (EMSA)

EMSA has been described previously for DNA-binding experiments with AID (19,64,88). Briefly, 1 µL of labelled 7-bub substrate as described above (0.15-30 fmol; see figure S1) was incubated with 3 µL (~100 ng) of Hs-AID or mutants in a buffer containing 50 mM Tris, pH 7.5, 2.0 µM MgCl₂, 50 mM NaCl, and 1 mM DTT in a final volume of 10 µL at 37 °C for 45 minutes. DNA+buffer alone was used as a negative control to contrast from complex formation in the presence of AID. After incubation, samples were immediately cross-linked (Stratagene) on ice at 2 cm from the UV source with 100 mJ and an irradiation time of 30s. Samples were electrophoresed at 4°C on an 8% native gel (0.5xTBE, 6% glycerol, 8% acrylamide: Bisacrylamide [19:1]) with a running buffer of 0.5xTBE at 300V for 3 hours. Gels were dried and visualized using a Phosphorimager (Molecular dynamics).

Data collection and quantification:

Image Lab or Quantity One Software (Bio-Rad, Hercules, CA, USA) were used to perform band densitometry of the product and substrate bands in alkaline cleavage gels and for the bound and unbound bands of EMSA gels. For pH range plots, the alkaline cleavage deamination assay was carried out for two to four independently purified preparations of each enzyme in duplicate resulting in 4 to 8 values represented by each graphed data point. Data in pH range plots were “lightly” smoothed according to GraphPad Prism’s XY smoothing function, using 4 neighboring data points. For the enzyme velocity kinetic plots at given pH, two independent preparations of each AID were tested in two parallel experiments for each preparation, totaling 4 independently-
measured values represented by each graphed data point. Enzyme kinetic data points were plotted according to GraphPad Prism’s Michaelis-Menten equation. For EMSA binding kinetic graphs, 2 independent enzyme preparations were tested across two parallel experiments at pH 7.5, totaling 4 values represented by each graphed data point. Using GraphPad Prism Software, data were plotted as deamination activity of AID (the percentage of the substrate which was converted to product) versus effective pH. To facilitate comparison between various AID mutants, chimeras and orthologs with different overall activity levels, maximum percentage of deamination activity was calculated by dividing each data point by the maximum absolute value for each data set. EMSA data points were plotted according to GraphPad Prism’s one site-specific binding equation. In each graph, the error bars represent standard deviations.

**Structural modeling and derivations of surface charge, isoelectric point and catalytic residue pKa:**

AID structure modeling and AID-ssDNA docking was described previously (51,84,87). Briefly, APOBEC3G (PDB: 3E1U; (89), and the X-ray crystal structure of the near native AID structure (PDB: 5W1C; (57) were retrieved from the protein databank (https://www.rcsb.org/) and used as templates for modelling using the default parameters of I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/ ; (90,91). The entire protein was homology modelled, except for the C-terminal amino acids, which were modeled *ab initio*. In total 5 catalytically-accessible AID conformations (51) were used for analysis and docking with ssDNA. ssDNA substrate 5’-TTTGCTT-3’ was constructed in Marvin Sketch v.5.11.5 with surface topology and docking parameters generated using Swiss-Param (https://www.swissparam.ch/); (91). AID conformations were docked with ssDNA using SwissDock (http://www.swissdock.ch/); (92,93), generating 5000-15000 binding modes, with 8
recurring binding modes (based on RMS) constituting a low-energy cluster, totaling 32 low-energy clusters for each AID-sDNA complex. AID-sDNA docked complexes were analyzed using UCSF chimera v1.7 (https://www.cgl.ucsf.edu/chimera/); (94). All AID and AID-sDNA complexes were visualized using PyMOL v1.7.6 (https://pymol.org/2/). The $E_{\text{Catalytic}}$ effective pKa, isoelectric point and surface charges of non-complexed enzymes were calculated using PDB2PQR web server (https://server.poissonboltzmann.org/pdb2pqr; (95,96).

**Results and Discussion**

**R25H acid-shifts the optimal pH of AID and dampens activity at basic pH**

It was previously reported that the majority of active A3 paralogues (A3A, A3B, A3C, A3D, A3G and A3H) deaminate cytidine efficiently at acidic pH (pH 5.5-6.5) but quite poorly at neutral or slightly basic pH (pH 7.0-8.0). It has been suggested that this difference in enzymatic activity was due to a net positive charge at acidic pH and net negative (or near zero) charge at neutral pH ranges, thus favoring binding to the negatively charged substrate ssDNA at lower pH. We and others have previously described the surface charge of AID at neutral pH and showed that it is highly positively charged particularly in the ssDNA binding grooves (19,51,56,57,84,84). Indeed, a previous report showed that AID was active at pH ranges 6.5-8.5 with the highest activity levels at pH 7.5 (80); However, this study examined activity at only 4 pH points (pH 5.5, 6.5, 7.5, and 8.5). Therefore, to establish a structural understanding of AID’s pH preference, we began by ensuring that our AID and A3 enzyme preparations behave according to the pH hierarchy established by several previous reports. To this end, we compared AID, A3A and A3G (Figure 1A) and examined their enzymatic activity at several pH in the standard alkaline cleavage assay (Figure 1B). We confirmed the previous observations that the A3s were adapted to more acidic pH than AID, with respective optimal pH at ~ 6 and ~7.2,
respectively. To ensure that AID’s optimal pH was not dependent on a particular substrate used, we also compared activity on three WRC motifs (TGC, TAC and AAC) and a non-WRC target (GGC) and observed the same optimal pH, suggesting this preference is a property of AID (Figure S1A). Interestingly, we noted a narrowing of AID substrate specificity at acidic pH, in agreement with A3A’s narrowed motif specificity at low pH (61).

Given the moderate degree of homology between Hs-AID and A3A/A3G (Figure 1A) we wondered about the structural basis for this difference in pH preference within the AID/APOBEC3 family. Due to the dramatic charge differences predicted at pH 7 (Figure S1B), we rationalized that there must be a network of ionizable residues that modulate substrate binding and influence catalysis. We examined surface exposed residues in the ssDNA binding groove of AID, near the catalytic pocket. We previously presented AID-ssDNA binding simulations identifying residues predicted to bind ssDNA on the surface in deamination-feasible complexes (51). This study proposed 50 DNA binding residues localized along two predicted ssDNA binding grooves (grooves 1 and 2) intersecting at the catalytic pocket. Subsequent crystal structures of AID and AID-DNA complexes (PDB: 5W0R) confirmed that groove 1 was the dominant DNA-binding groove in AID, thereby strengthening our original predictions (57). Of these DNA-binding residues, R25 was identified as the most frequent ssDNA-binding residue in 94% of AID-ssDNA complexes that positioned dC for catalysis. Intersecting both ssDNA binding grooves and located at the apex of the catalytic pocket, R25 was found to bind nucleobases and electrostatically bind 2-5 phosphodiester groups of ssDNA (Figure 1C). Indeed, the R25 equivalent residue in A3G (H216) (Figure 1A) was previously identified as being a key residue responsible for the acidic optimal pH of A3G, hypothesized to occur due to the protonation of its imidazole ring contacting ssDNA (79). Comparing several A3s and AID
orthologs we found that the R25 equivalent in Hs-AID is remarkably variable within the AID/APOBEC3 family and within AID orthologs as either an R, H or N (Figure S1C). Given this variability, its prominent role in ssDNA binding near the catalytic pocket and demonstrated role of its equivalent residue in A3G’s pH adaption, we wondered if mutation of R25 in Hs-AID could alter pH preference and overall activity.

Previous studies of AID/APOBEC3 pH preference typically tested activities at 3 or 4 pH points, often leaving wide gaps in the tested pH. To precisely pinpoint optimal pH and to also compare the extent of acid, neutral and basic pH stability of each enzyme, we designed experiments to test dC deamination at 24 different pH points, at fine increments of 0.07-0.12, ranging from pH 5.94-8.20 (Figure 1D-F). This analysis pinpointed the exact optimal pH of Hs-AID to 7.3 and showed that it maintains near-optimal activity (>90% catalysis) at a pH range of 7.1-7.6. We then examined the activities of a panel of Hs-AID R25 mutants including R25H, R25N, R25A, and R25D (Figure 1E). Compared to Hs-AID, we found R25H had comparable levels of activity, while R25N had retained approximately 50% activity. R25D was inactive, while R25A retained minimal activity. Despite their severe impairment of catalysis, R25D and R25A displayed a high affinity for binding ssDNA, characteristic of Hs-AID ($K_d = 0.33$ nM, 0.02 nM, and 0.33 nM, respectively; Figure S1D), suggesting these mutants have maintained the integrity of their ssDNA-binding grooves. These data are consistent with overall ssDNA attachment to AID’s surface globally (as measured by EMSA) being a feature of its very highly positive surface charge, rather than dependent on one specific residue (51). At the same time the data suggest AID requires a positive/hydrophilic residue at this position in order to direct the path of ssDNA specifically towards the catalytic pocket in deamination-feasible complexes (51). Besides differences in overall activity, mutation of R25 also altered the pH preference as
hypothesized. Whereas Hs-AID has an optimal pH range of 7.2-7.6 with stable activity at higher pH (70% of peak activity at pH 8.0), R25H and R25N exhibit slightly lower optimal and unaltered optimal pH (7.1 and 7.33, respectively) (Figure 1F). Thus, R25H lowers the optimal pH of AID and dramatically decreases activity at even slightly more basic pH (10-20% of max activity of R25H at pH 8.0-8.2, compared to 50-70% for Hs-AID), likely due to a loss of local ssDNA binding near the catalytic pocket via deprotonation of the imidazole ring. In contrast, R25N is closer to Hs-AID both in its optimal pH (7.33 vs. 7.45) and with respect to retaining much of its activity at basic pH (50-60% of max activity at pH 8.20; Figure 1F). The activity profiles obtained thus far were based on using single AID:substrate ratios across many fine pH increments as shown in Figure 1D. To validate the optimal pH results of these activity assays across pH 5.94-8.20, we carried out enzyme kinetics to compare initial deamination velocities of Hs-AID, R25H and R25N. For each of the three enzymes, we measured velocity at Hs-AID’s optimal pH (pH 7.21), an acidic pH below optimal (pH 6.7), and a basic pH above optimal (pH 7.99) (Figure 1G). The initial velocity kinetics supported the results obtained by pH range experiments, first in that all three enzymes exhibited the highest velocity at their optimal pH, and second, in that Hs-AID maintains a higher level of activity at the basic range past its optimal pH, as compared to R25 mutants. From these analyses, we make three conclusions: first, Hs-AID retains most of its activity at basic pH; second, homologous mutations of R25 can dramatically change overall activity levels of AID and its pH preference patterns; third, in line with previous finding that the R25 equivalent in A3G (H216) underlies the acidic pH preference of A3G, Hs-AID R25H also exhibits a shift to lower optimal pH with significantly lowered activity at higher basic pH ranges.

**E26 stabilizes the ssDNA binding grooves, while E26R base-shifts the optimal pH**
Having demonstrated the role of R25 in modulating AID’s pH sensitivity, we continued along the same rationale to explore the role of other ionizable residues neighboring R25 in the ssDNA binding grooves. Adjacent to R25, E26 has also been predicted to contact ssDNA in a large proportion of AID:ssDNA complexes (75% of complexes; (51). Unlike R25, which binds ssDNA via the phosphodiester backbone and nucleobases, E26 binds nucleobases, but appears to electrostatically repel the DNA phosphodiester groups. Instead, E26 hydrogen bonds with Q14 and Y28, and electrostatically binds ssDNA binding residues R25, R50 and K52 (R25, R50 and K52 bind ssDNA in 93%, 31% and 69% of complexes, respectively; (51); Figure 2A-B). In the case of R25H, at pH 8-8.5 there is a local negative charge created proximal to the opening of the catalytic pocket (neutral H25 + negative E26), thus requiring a lower pH to protonate H25 and promote ssDNA binding. By that same logic, E26R creates a positively-charged side-chain capable of hydrogen bonding with Q14 and Y28, but disrupts R25, R50 and K52 stabilization, thereby hindering ssDNA binding near the catalytic pocket. We found that E26R retained catalysis near that of Hs-AID, but unexpectedly increased the optimal pH to 7.77 (Figure 2D). We suspect the disrupted DNA binding grooves of E26R are optimized at higher pH, resulting in near-native levels of catalysis.

Similar to R25, R24 was predicted to bind ssDNA in 50% of AID:ssDNA complexes (51). However, due to a large portion of its side-chain buried between L1/α1, R24 plays a more critical role in stabilizing the DNA binding grooves (Figure 2A; (51,57). We therefore speculated that mutation of R24 would result in a severe impairment of activity. Indeed, R24D resulted in a complete loss of enzymatic activity, while R24A produced a catalytically hindered mutant with no change in pH preference (Figure 2C). Like the impairment of catalysis caused by mutation of R25, R24D and R24A hampered catalysis whilst maintaining a high affinity for binding ssDNA.
(Figure S1D), suggesting these mutations, like R25H, maintain the integrity of the ssDNA binding grooves, but compromise deamination-specific ssDNA binding. We also probed the mutational effect of several hyper-IgM mutations and other mutations on the pH preference of AID. Most mutants were catalytically inactive and those that were enzymatically active showed no change in pH preference in comparison to Hs-AID (Figure S1E). We conclude that R24 and E26 stabilize deamination-feasible conformations of the ssDNA binding grooves and that E26 influences pH through stabilization of electrostatic topology near the catalytic pocket.

**Orthologous AIDs have variable optimal pH, largely due to differences in a single surface residue**

We examined the pH preference of AID from evolutionary-distant species. We have previously demonstrated that AID from zebrafish (*Danio rerio*; Dr-AID) and channel catfish (*Ictalurus punctatus*; Ip-AID) display biochemical differences compared to Hs-AID, such as catalytic rates (Dr-AID >> Hs-AID >> Ip-AID) and optimal temperatures (fish AID are adapted to colder temperatures than mammalian AID) (83, 85). We wondered if some of these structural differences also mediated differences in optimal pH. Like AID/A3s from humans, we observed the same variability of the Hs-AID R25-equivalent residues amongst orthologous AID (R, H or N) in that Dr-AID and Ol-AID both had H29 whilst Ip-AID had N28 (Figure 3A). We next examined the ssDNA binding interactions of the Hs-AID R25 equivalent residues in Dr- and Ip-AID (H29 and N28, respectively) (Figure 1C and 3B). Comparing R25 (Hs-AID), H29 (Dr-AID) and N28 (Ip-AID), we found stabilization of 3-5, 3-4 and 1-2 phosphodiester groups, respectively. This difference in phosphodiester group stabilization is likely reflected by the +1, partial positive charge and neutral charge of R, H and N, respectively. Regardless of the residue present, a conserved R and E neighbour this position on either side in orthologous AID.
However, there are many sequence differences between Hs-, Dr-, Ip- and Ol-AID, notably the secondary catalytic loops, ssDNA-binding grooves, and overall surface charge (Figure 3A and S1B), all of which could influence pH preference. Additionally, the A3s were shown to be acid-adapted (optimal pH 5.5-6.5; (61,79–81) despite varying between R, H or N at this position (Figure S1C). Therefore, it is unlikely that the optimal pH of orthologous AID is dictated solely by the identity of this residue. To test this hypothesis, we first examined the enzymatic activity of Dr-, Ol- and Ip-AID in comparison to Hs-AID at different pH (Figure 3C). In terms of optimal pH, Dr-AID and Ol-AID were quite close to Hs-AID (pH 7.6 vs 7.3) whilst Ip-AID was more base-adapted (pH of 7.9). Like Hs-AID, the orthologous fish AID also exhibited a stable pH profile at higher pH. The near-optimal pH, (>90% of optimal activity window) is 7.1-7.6 for Hs-AID, 7.2-7.8 for Dr-AID, 7.2-8.1 (broadest) for Ol-AID and 7.5-8.2 (most basic) for Ip-AID.

We next probed the effects of reciprocal mutations of this position in Dr- and Ip-AID (Figure 3D and E). Dr-AID H29R and H29N both showed a shift towards a more basic optimal pH of 8.1 in comparison to Dr-AID. On the other hand, Ip-AID N28R and N28H showed a shift towards a more acidic optimal pH of 7.1 and 7.33, respectively, compared to Ip-AID. Thus, AID from different species demonstrate differences in optimal pH, and reciprocal mutation of the Hs-AID R25-equivalent residue between Dr-AID (H29) and Ip-AID (N28), imparts the donor enzyme’s optimal pH to the recipient (7.3 vs. 8.1). These results are consistent with the analyses of Hs-AID mutants in which R25N was slightly more basic shifted in its optimal pH than R25H. Even though the reciprocal mutation of Hs-AID’s R25 equivalents in Dr-AID and Ip-AID were sufficient to impart the optimal pH of the donor, two lines of evidence suggested that other residues or regions in AID/APOBEC3s are likely involved in determination of optimal pH. First, R25N in Hs-AID did not mediate a drastic basic pH shift, as it did in the case of Dr-AID’s
H29N, and second, A3F with an N in the equivalent position (N214) represents an oddity within the AID/APOBEC family as indicated by its retention of catalysis across a broad pH range of 5.5-9.5 (81). Thus, we next sought to examine three different factors that could influence optimal pH: first, the optimal pH for the deamination reaction in the catalytic pocket, second, the overall surface charge and isoelectric point of each protein in its native folded state, and third, specific physical regions that when transplanted into chimeric enzymes could alter optimal pH.

The catalytic pocket microenvironment is not a determinant of optimal pH in AID/APOBEC3s

In the catalytic pocket of AID/APOBEC3s, the catalytic glutamic acid (E58 in human AID, henceforth referred to as ECatalytic) donates a proton during the cytidine deamination reaction, followed by re-protonation, thus regenerating the catalytic pocket for the next deamination reaction (53). The transition between the de-protonated and re-protonated states would be quickest at pH ranges nearest to the effective pKa of ECatalytic. Unlike intrinsic pKa, which is suitable for surface residues, the effective pKa is determined by the microenvironment of the ionizable group. For example, many enzyme active sites utilize a catalytic glutamic acid with an elevated effective pKa (e.g. intrinsic pKa = 4.5 vs. effective pKa = 6.7; (97). Thus, in the case of AID, the ECatalytic effective pKa is influenced by its solvent accessibility and the composition of the secondary catalytic pocket residues. Given the catalytic pocket homology within the AID/APOBEC3 family (Figure 1A), we wondered if optimal pH of AID/APOBEC3s would reflect the ECatalytic effective pKa. Due to A3F’s unique and wide-ranging optimal pH of 5.5-9.5, we suspect unique structural features outside the scope of this study explain its pH preference and have therefore excluded A3F from further analysis. We predicted the ECatalytic effective pKa of AID/APOBEC3 structures and found it was well conserved within a narrow range (ECatalytic...
effective pKa = 2.7-3.5) (Figure S1F). The average E_{\text{Catalytic}} effective pKa amongst the A3s was only slightly lower than that of the AID orthologs (3.0 vs. 3.5). Therefore, it appears unlikely that this factor is responsible for the differences in optimal pH. However, an interesting trend can be noted in the difference between the E_{\text{Catalytic}} effective pKa and optimal pH. If the most efficient de-protonation/re-protonation cycle of E_{\text{Catalytic}} occurs at or near its effective pKa, it would stand to reason that the further the enzyme’s optimal pH is from its E_{\text{Catalytic}} effective pKa, the less optimal the de-protonation/re-protonation regeneration of E_{\text{Catalytic}} between deamination reactions, and hence the slower the catalytic rate. In support of this notion, we observed that the difference between optimal pH and E_{\text{Catalytic}} effective pKa was greater in AID (all orthologs) vs. APOBEC3s (Figure 4A), a trend that correlates with their respective catalytic rates. These findings suggest that E_{\text{Catalytic}} effective pKa does not directly alter the optimal pH of AID/APOBEC3s; however, it might be suggestive of a possible mechanism used by these enzymes to limit mutagenic activity: to evolve structural features outside the catalytic pocket that positions the enzyme’s overall optimal pH at a greater distance from its E_{\text{Catalytic}} effective pKa. Thus, we sought to investigate the structural features beyond the catalytic pocket that determine overall pH optima.

**Computed surface charge correlates with optimal pH in the AID/APOBEC3 family**

To examine other possible attributes that may determine the optimal pH of AID/APOBEC3 paralogues, we probed the relationship between the optimal pH and isoelectric points. When the correlation between isoelectric points and optimal pH was examined, a linear relationship was observed where the APOBECs with lower isoelectric points had lower optimal pH (Figure 4B). Thus, optimal pH within the AID/APOBEC3 family is highly dependent upon isoelectric point. When we examined surface charge vs. optimal pH or near-optimal pH (>90% catalysis) (Figure
4C, Figure S1B) we noted all AID/APOBEC3s exhibited a positive surface charge. All AID orthologs had a higher positive charge (average +11) vs. the A3s (average +6) at their optimal pH. Consistent with its quick on/off rate of ssDNA binding, A3A exhibited the lowest positive charge (+3.5) at its optimal pH 5.5, while its surface charge became negative near neutral pH (Figure S1B) explaining its requirement for acidic pH. This indicates that an overall positive surface charge is a key determinant of optimal pH within the AID/APOBEC3 family, likely for optimal ssDNA binding/catalysis. This idea is consistent with a previous study that demonstrated the optimal pH for A3A catalysis correlated with the optimal pH of substrate binding, namely that A3A’s binding affinity for ssDNA was several-fold diminished when tested at near neutral pH vs. its optimal pH 5.5 (61). In addition, we note that the most catalytically robust AID/APOBEC3 parologue (A3A) has the lowest optimal pH and lowest surface charge at its optimal pH, while the most catalytically lethargic enzyme (AID) has the highest optimal pH and surface charge at optimal pH. Thus, we speculate that higher surface charge in the AID/APOBEC3 family contributes to catalytic lethargy through two mechanisms: first by facilitating slower on/off binding events, and second, through resulting in an optimal pH that is further displaced from the $E_{\text{Catalytic effective pKa}}$, as discussed in the preceding section.

The optimal pH and base-stability of AID is dependent on the ssDNA binding grooves and assistant patch

To test whether there are distinct structural motifs in the AID structure that alter pH sensitivity, we constructed multiple chimeras incorporating regions from Dr-AID into a Hs-AID scaffold (Figure 5A and S2) and compared optimal pH to Hs-AID (Figure 5). We began by targeting loop 2 (L2) and loop 4 (L4; Hs/Dr-AID-5 and Hs/Dr-AID-6, respectively), regions not involved in...
binding catalytically-productive ssDNA complexes and thus, unlikely to perturb optimal pH. As expected, Hs/Dr-AID-5 and Hs/Dr-AID-6 did not perturb the optimal pH (Figure 5B). We also targeted α3 (Hs/Dr-AID-3, Hs/Dr-AID-4, and Hs/Dr-AID-1), a region that composes ssDNA-binding groove 2 (19,51,56) and would likely affect optimal pH (Figure 5C). Analogous to R25H in Hs-AID, we found Hs/Dr-AID-3 and Hs/Dr-AID-4 both acid-shifted the optimal pH slightly (optimal pH= ~7.1), while substantially reducing activity at more basic pH (30-40% and 20-30% of max activity for Hs/Dr-AID-3 and Hs/Dr-AID-4 at pH 8.0-8.2, respectively, compared to 50-70% for Hs-AID). Inspection of the Hs/Dr-AID-3 sequence (Figure S2) revealed major differences from Hs-AID, including the rearrangement of several positive/negative residues, as well as the loss and gain of histidine residues (H93Q and D96H, respectively) that may impact ssDNA-binding across pH. As a follow-up, we also examined the optimal pH of L2+L4+α3 (Hs/Dr-AID-1), a chimera we previously reported with enhanced catalytic pocket accessibility and catalytic rate (51). Surprisingly, Hs/Dr-AID-1 had retained the same optimal pH and base-stability as Hs-AID, suggesting the reduced activity at basic pH offered by α3 was masked by L2 and L4. Although not directly involved in binding ssDNA complexes near the active site, L2 and L4 are heavily involved in the predicted dimerization interface of AID (56,98), whereby L2 and L4 from the neighboring AID forms an extended ssDNA binding groove 1. Thus, chimeras involving L2 and L4 may indirectly affect pH stability through changes of the ssDNA binding groove-1 of an AID dimer, or the formation of AID dimers altogether.

Next, we examined chimeras altering the N-terminal region (N, Hs/Dr-AID-8), loop 7 (L7, Hs/Dr-AID-9) and the C-terminal region (C, Hs/Dr-AID-11; Figure 5D). The N-terminal region alters the N-termini, α1 and part of L1 (retains R25), which contains some residues involved in ssDNA binding, but acts primarily to stabilize the conformation of L1 residues involved in
deamination-specific ssDNA binding (e.g., R25 and E26). L7 contains many ssDNA contact residues, most notable for its vital role in substrate specificity (75) and dC binding in the catalytic pocket amongst the AID/APOBEC family (51,52,57,63,66,73,78). The C-terminal region alters α6 and α7. α6 houses the assistant patch, which mediates structured DNA-binding in AID (57), while α7 is unique to AID in that it shares no homology with other APOBECs (Figure 1A). Additionally, α7 was predicted to adopt several low-energy conformations relative to the core of AID, some of which contact L7 and may impact the ssDNA binding grooves (51). Thus, we hypothesized that alteration to these regions alone or in combination would modulate pH sensitivity. Examining Hs/Dr-AID-8 and Hs/Dr-AID-9, we find both base-shift the optimal pH to 7.56, while retaining a similar level of activity at more basic pH. In contrast, Hs/Dr-AID-11 is most active across a broadened pH range (optimal pH= 7.56-8.2; Figure 5D, right panel). Given the interplay between these regions, we next examined chimeras involving N+L7, N+C and N+L7+C (Hs/Dr-AID-10, Hs/Dr-AID-12, Hs/Dr-AID-13, respectively; Figure 5E). Hs/Dr-AID-10 and Hs/Dr-AID-12 both demonstrated base-shifted optimal pH (optimal pH 7.56 and 7.77, respectively) and stability at basic pH like that of Hs/Dr-AID-8, suggesting the N-terminal region plays a dominant role in these chimeras. Interestingly, Hs/Dr-AID-13 demonstrated an acid-shifted optimal pH of 6.99 with retention of >70% activity across a broadened pH range of 6.8-8.2 (Figure 5E, right panel). The observations that the N, L7 and C chimeras alone constitute a base-shifted enzyme, but a chimera with all three combined demonstrate an acid-shifted enzyme with activity across a broadened pH reinforce the notion that the pH sensitivity of AID is dependent upon a network of DNA binding residues.

Conclusion
Using a high-resolution map of catalysis vs. pH, we have pinpointed the optimal pH of AID and considered it in the context of its APOBEC3 paralogues. We found a correlation between optimal pH and surface charge in AID/APOBEC3s and noted that catalytic rate has an inverse relationship with both surface charge and optimal pH. The family member with the lowest catalytic rate (AID) has the most positive surface, highest optimal pH (neutral-basic), and largest difference between its optimal pH and the optimal pH of its catalytic pocket microenvironment ($E_{\text{catalytic effective pKa}}$). In contrast, the family member with the most robust catalytic rate (A3A) has the least positively charged surface, lowest optimal pH (acidic), and smallest difference between its optimal pH and the optimal pH of its catalytic pocket microenvironment ($E_{\text{catalytic effective pKa}}$). Thus, we suggest that in AID/APOBEC3s, surface charge regulates enzyme activities through two mechanisms: first, through allowing for faster on/off binding rates of negatively charged ssDNA by the enzymes with the least positively charged surface; and second, through modulating the magnitude of the difference between the enzyme’s overall optimal pH and the optimal pH of the catalytic microenvironment ($E_{\text{catalytic effective pKa}}$), as the further the enzyme’s optimal pH is from the $E_{\text{catalytic pKa}}$, the less efficient the process of catalytic regeneration (de-protonation/re-protonation cycle) between successive cytidine deamination reactions.

We previously showed that structural restriction of the catalytic pocket through spontaneous occlusion/opening (Schrödinger’s CATalytic pocket) is a structurally built-in mechanism that regulates the mutagenic activities of AID/APOBEC3s. Here, our in-depth study of pH optima across AID/APOBEC3s led us to recognize yet another structurally built-in regulatory mechanism that limits the mutagenic activities of these enzymes.
Acknowledgements

This work was supported by operating funds from the Canadian Institutes of Health Research (CIHR, MOP111132) and a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant (2015-047960). JK was supported by a CIHR doctoral fellowship and is currently supported by a CIHR post-doctoral fellowship.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and/or its supplementary materials. Raw data that support the findings of this study are available from the corresponding author, upon reasonable request.
Figure 1
Figure 1: R25 mediates ssDNA binding and pH sensitivity in AID. A) Sequence alignment of AID and several APOBEC3 paralogues, highlighting R25 in AID and its homologous residues. B) General activity of AID, A3A and A3G, demonstrating AID is more base-adapted. C) Surface structure of a representative AID-ssDNA complex. The AID surface is colored beige, with purple representing the Zn-coordinating and catalytic glutamic acid residues, blue as R25 and red as the bound ssDNA in a deamination-feasible conformation. Inset: Zoomed in interaction, demonstrating the electrostatic stabilization of R25 and 2-5 phosphodiester groups of ssDNA with only the phosphodiester backbone shown for clarity. D) Representative alkaline cleavage activity assay gel of AID. E) Catalytic activity vs pH of AID and several R25 mutants. F) Catalytic activity (%max) vs pH of Hs-AID, R25H and R25N. G) Deamination velocities of Hs-AID (left), R25H (middle) and R25N (right) at different pH.
Figure 2
Figure 2: E26 indirectly mediates ssDNA binding and pH sensitivity in AID. A) Surface structure of AID illustrating R24, E26 and the residues bound by E26. The Zn-coordinating and catalytic glutamic acid residues are colored purple, positively charged residues blue, negative red and residues hydrogen-bonded by E26 colored green. B) E26 stabilization of neighboring residues across two different conformations of catalytically-accessible AID (left and right). C) Catalytic activity vs pH of R24A and E26R. R24D was catalytically dead at all pH tested and omitted from the graph. D) Catalytic activity (%max) vs pH comparing Hs-AID, R25H and E26R and demonstrating changes in optimal pH.
Figure 3
Figure 3: ssDNA binding and pH preference in orthologous AID. A) Sequence alignment of Hs-, Dr-, Ip-, and Ol-AID, highlighting the sequence homology of R25 in Hs-AID and orthologous AID. B) Surface structure of a representative Dr-AID-ssDNA complex (left) and Ip-AID-ssDNA complex (right). The AID surface is colored beige, with purple representing the Zn-coordinating and catalytic glutamic acid residues. Left and right inset: Zoomed in interaction demonstrating the role of the Hs-AID R25 equivalent residues in Dr-AID (H29) and Ip-AID (N28) with only the phosphodiester backbone shown for clarity. For Dr- and Ip-AID, H29 and N28 bind ssDNA through hydrogen bonding of the phosphodiester backbone. C) Catalytic activity pH (top) and Catalytic activity (%max) (bottom) vs pH for Hs-, Dr-, Ol- and Ip-AID. D) Catalytic activity (top) and catalytic activity (%max) (bottom) vs pH for Dr-AID and homologous mutants H29N and H29R. E) Catalytic activity (top) and catalytic activity (%max) (bottom) vs pH for Ip-AID and homologous mutants N28H and N28R.
Figure 4: Orthologous AID and APOBEC3 characteristics that relate to optimal pH. A) Optimal pH-E\textsubscript{Catalytic} effective pKa across APOBEC3s and orthologous AIDs. On average, APOBEC3 enzymes have a lower difference compared to orthologous AID. B) Optimal pH vs Isoelectric point (R\textsuperscript{2}=0.85, p < 0.05). In general, orthologous AID have a higher isoelectric point and optimal pH as compared to APOBEC3s. C) Enzyme surface charge vs optimal pH of orthologous AID and APOBEC3s. At their respective optimal pH, AID has a higher positive surface charge as compared to APOBEC3s. The horizontal range bars indicate the pH range at which the enzyme’s activity is within the >90% of the maximal activity measured at optimal pH. The vertical range bars indicate the charge range at this pH range (derived from Figure S1B).
Figure 5
Figure 5: Hs-AID with different regions of Dr-AID grafted to localize which regions affect pH sensitivity. A) General structure of Hs-AID depicting which regions apply to the Hs/Dr-AID chimeras. N-to-C terminal progression is colored from Blue to red, with the coordinated-Zn in the catalytic pocket depicted as a purple sphere. B) Catalytic activity (%max) vs pH for chimeras involving loop 2 (L2), loop 4 (L4), α3, L2+α3, L2+L4+α3, D) N-terminal region (N), loop 7(L7), C-terminal region and E) N+L7, N+C and N+L7+C. In panel B-E, the inset AID structure is colored beige, with regions grafted from Dr-AID colored red.
References

1. Larijani M, Martin A. The biochemistry of activation-induced deaminase and its physiological functions. Semin Immunol. 2012 Aug;24(4):255–63.

2. Conticello SG. The AID/APOBEC family of nucleic acid mutators. Genome Biol. 2008;9(6):229.

3. Conticello SG, Langlois MA, Yang Z, Neuberger MS. DNA deamination in immunity: AID in the context of its APOBEC relatives. Adv Immunol. 2007;94:37–73.

4. Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol. 2005 Feb;22(2):367–77.

5. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, et al. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem. 1999 Jun 25;274(26):18470–6.

6. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell. 2000 Sep 1;102(5):553–63.

7. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, et al. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell. 2000 Sep 1;102(5):565–75.

8. Navaratnam N, Morrison JR, Bhattacharya S, Patel D, Funahashi T, Giannoni F, et al. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J Biol Chem. 1993 Oct 5;268(28):20709–12.

9. Teng B, Burant CF, Davidson NO. Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science. 1993 Jun 18;260(5115):1816–9.

10. Liao W, Hong SH, Chan BH, Rudolph FB, Clark SC, Chan L. APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochem Biophys Res Commun. 1999 Jul 5;260(2):398–404.

11. Sato Y, Ohtsubo H, Nihei N, Kaneko T, Sato Y, Adachi S-I, et al. Apobec2 deficiency causes mitochondrial defects and mitophagy in skeletal muscle. FASEB J Off Publ Fed Am Soc Exp Biol. 2018 Mar;32(3):1428–39.

12. Rogozin IB, Basu MK, Jordan IK, Pavlov YI, Koonin EV. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. Cell Cycle. 2005 Sep;4(9):1281–5.
13. Marino D, Perković M, Hain A, Jaguva Vasudevan AA, Hofmann H, Hanschmann K-M, et al. APOBEC4 Enhances the Replication of HIV-1. PloS One. 2016;11(6):e0155422.

14. Mikl MC, Watt IN, Lu M, Reik W, Davies SL, Neuberger MS, et al. Mice deficient in APOBEC2 and APOBEC3. Mol Cell Biol. 2005 Aug;25(16):7270–7.

15. Lada AG, Krick CF, Kozmin SG, Mayorov VI, Karpova TS, Rogozin IB, et al. Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. Biochem Biokhimia. 2011 Jan;76(1):131–46.

16. Chiu YL, Greene WC. APOBEC3 cytidine deaminases: distinct antiviral actions along the retroviral life cycle. J Biol Chem. 2006 Mar 31;281(13):8309–12.

17. Harris RS. Enhancing immunity to HIV through APOBEC. Nat Biotechnol. 2008 Oct;26(10):1089–90.

18. Stavrou S, Ross SR. APOBEC3 Proteins in Viral Immunity. J Immunol. 2015 Nov 15;195(10):4565–70.

19. Abdouni HS, King JJ, Ghorbani A, Fifield H, Berghuis L, Larijani M. DNA/RNA hybrid substrates modulate the catalytic activity of purified AID. Mol Immunol. 2017 Nov 18;93:94–106.

20. Nabel CS, Lee JW, Wang LC, Kohli RM. Nucleic acid determinants for selective deamination of DNA over RNA by activation-induced deaminase. Proc Natl Acad Sci U S A. 2013 Aug 27;110(35):14225–30.

21. York A, Kutluay SB, Errando M, Bieniasz PD. The RNA Binding Specificity of Human APOBEC3 Proteins Resembles That of HIV-1 Nucleocapsid. PLoS Pathog. 2016 Aug;12(8):e1005833.

22. Svarovskaia ES, Xu H, Mbisa JL, Barr R, Gorelick RJ, Ono A, et al. Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs. J Biol Chem. 2004 Aug 20;279(34):35822–8.

23. Stauch B, Hofmann H, Perkovic M, Weisel M, Kopietz F, Cichutek K, et al. Model structure of APOBEC3C reveals a binding pocket modulating ribonucleic acid interaction required for encapsidation. Proc Natl Acad Sci U S A. 2009 Jul 21;106(29):12079–84.

24. Sharma S, Patnaik SK, Taggart RT, Kannisto ED, Enriquez SM, Gollnick P, et al. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun. 2015 Apr 21;6:6881.

25. Sharma S, Patnaik SK, Taggart RT, Baysal BE. The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. Sci Rep. 2016 Dec 15;6:39100.

26. Peters A, Storb U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. Immunity. 1996 Jan;4(1):57–65.
27. Branton SA, Ghorbani A, Bolt BN, Fifield H, Berghuis LM, Larijani M. Activation-induced cytidine deaminase can target multiple topologies of double-stranded DNA in a transcription-independent manner. FASEB J Off Publ Fed Am Soc Exp Biol. 2020 Jul;34(7):9245–68.

28. Siriwardena SU, Chen K, Bhagwat AS. Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases. Chem Rev. 2016 Oct 26;116(20):12688–710.

29. Love RP, Xu H, Chelico L. Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. J Biol Chem. 2012 Aug 31;287(36):30812–22.

30. Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, et al. Two levels of protection for the B cell genome during somatic hypermutation. Nature. 2008 Feb 14;451(7180):841–5.

31. Klasen M, Spillmann FJ, Marra G, Cejka P, Wabl M. Somatic hypermutation and mismatch repair in non-B cells. Eur J Immunol. 2005 Jul;35(7):2222–9.

32. Kasar S, Kim J, Improgo R, Tiao G, Polak P, Haradhvala N, et al. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. Nat Commun. 2015 Dec 7;6:8866.

33. Hoopes JI, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA, Roberts SA. APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. Cell Rep. 2016 Feb 16;14(6):1273–82.

34. Adolph MB, Love RP, Chelico L. Biochemical Basis of APOBEC3 Deoxycytidine Deaminase Activity on Diverse DNA Substrates. ACS Infect Dis. 2018 Mar 9;4(3):224–38.

35. Henderson S, Fenton T. APOBEC3 genes: retroviral restriction factors to cancer drivers. Trends Mol Med. 2015 May;21(5):274–84.

36. Robbiani DF, Bothmer A, Callen E, Reina-San-Martin B, Dorsett Y, Difilippantonio S, et al. AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. Cell. 2008 Dec 12;135(6):1028–38.

37. Robbiani DF, Bunting S, Feldhahn N, Bothmer A, Camps J, Deroubaix S, et al. AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. Mol Cell. 2009 Nov 25;36(4):631–41.

38. Ramiro A, Reina San-Martin B, McBride K, Jankovic M, Barreto V, Nussenzweig A, et al. The role of activation-induced deaminase in antibody diversification and chromosome translocations. Adv Immunol. 2007;94:75–107.

39. Pasqualucci L, Bhagat G, Jankovic M, Compagno M, Smith P, Muramatsu M, et al. AID is required for germinal center-derived lymphomagenesis. Nat Genet. 2008 Jan;40(1):108–12.
40. Klemm L, Duy C, Iacobucci I, Kuchen S, von Levetzow G, Feldhahn N, et al. The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. Cancer Cell. 2009 Sep 8;16(3):232–45.

41. Compagno M, Wang Q, Pighi C, Cheong TC, Meng FL, Poggio T, et al. Phosphatidylinositol 3-kinase delta blockade increases genomic instability in B cells. Nature. 2017 Feb 23;542(7642):489–93.

42. Fruman DA, O’Brien S. Cancer: A targeted treatment with off-target risks. Nature. 2017 Feb 23;542(7642):424–5.

43. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 2013 Feb 21;494(7437):366–70.

44. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013 Aug 22;500(7463):415–21.

45. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet. 2013 Sep;45(9):970–6.

46. Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: footprints and mechanisms. Nat Rev Cancer. 2014;14(12):786–800.

47. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016 Jun 2;534(7605):47–54.

48. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet. 2015 Sep;47(9):1067–72.

49. Knisbacher BA, Gerber D, Levanon EY. DNA Editing by APOBECs: A Genomic Preserver and Transformer. Trends Genet TIG. 2016 Jan;32(1):16–28.

50. Smith HC, Bennett RP, Kizilyer A, McDougall WM, Prohaska KM. Functions and regulation of the APOBEC family of proteins. Semin Cell Dev Biol. 2012 May;23(3):258–68.

51. King JJ, Manuel CA, Barrett CV, Raber S, Lucas H, Sutter P, et al. Catalytic pocket inaccessibility of activation-induced cytidine deaminase is a safeguard against excessive mutagenic activity. Struct Lond Engl 1993. 2015 Apr 7;23(4):615–27.

52. King JJ, Larijani M. A Novel Regulator of Activation-Induced Cytidine Deaminase/APOBECs in Immunity and Cancer: Schrodinger’s CATalytic Pocket. Front Immunol. 2017;8:351.

53. Conticello SG, Langlois MA, Neuberger MS. Insights into DNA deaminases. Nat Struct Mol Biol. 2007 Jan;14(1):7–9.
54. Kouno T, Silvas TV, Hilbert BJ, Shandilya SMD, Bohn MF, Kelch BA, et al. Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. Nat Commun. 2017 Apr 28;8:15024.

55. Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nat Struct Mol Biol. 2017 Feb;24(2):131–9.

56. King JJ, Larijani M. Structural plasticity of substrate selection by activation-induced cytidine deaminase as a regulator of its genome-wide mutagenic activity. FEBS Lett. 2021 Jan;595(1):3–13.

57. Qiao Q, Wang L, Meng FL, Hwang JK, Alt FW, Wu H. AID Recognizes Structured DNA for Class Switch Recombination. Mol Cell. 2017 Aug 3;67(3):361-373 e4.

58. Zheng S, Vuong BQ, Vaidyanathan B, Lin JY, Huang FT, Chaudhuri J. Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA. Cell. 2015 May 7;161(4):762–73.

59. Mondal S, Begum NA, Hu W, Honjo T. Functional requirements of AID's higher order structures and their interaction with RNA-binding proteins. Proc Natl Acad Sci U S A. 2016 Mar 15;113(11):E1545-54.

60. Pucella JN, Chaudhuri J. AID Invited to the G4 Summit. Mol Cell. 2017 Aug 3;67(3):355–7.

61. Pham P, Landolph A, Mendez C, Li N, Goodman MF. A biochemical analysis linking APOBEC3A to disparate HIV-1 restriction and skin cancer. J Biol Chem. 2013 Oct 11;288(41):29294–304.

62. Xiao X, Yang H, Arutiunian V, Fang Y, Besse G, Morimoto C, et al. Structural determinants of APOBEC3B non-catalytic domain for molecular assembly and catalytic regulation. Nucleic Acids Res. 2017 Jul 7;45(12):7494–506.

63. Pham P, Afif SA, Shimoda M, Maeda K, Sakaguchi N, Pedersen LC, et al. Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification. DNA Repair Amst. 2016 Jul;43:48–56.

64. Larijani M, Petrov AP, Kolenchenko O, Berru M, Krylov SN, Martin A. AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. Mol Cell Biol. 2007 Jan;27(1):20–30.

65. Pham P, Calabrese P, Park SJ, Goodman MF. Analysis of a Single-stranded DNA-scanning Process in Which Activation-induced Deoxycytidine Deaminase (AID) Deaminates C to U Haphazardly and Inefficiently to Ensure Mutational Diversity. J Biol Chem. 2011 Jul 15;286(28):24931–42.

66. Hou S, Silvas TV, Leidner F, Nalivaika EA, Matsuo H, Kurt Yilmaz N, et al. Structural Analysis of the Active Site and DNA Binding of Human Cytidine Deaminase APOBEC3B. J Chem Theory Comput. 2019 Jan 8;15(1):637–47.
67. Yang H, Ito F, Wolfe AD, Li S, Mohammadzadeh N, Love RP, et al. Understanding the structural basis of HIV-1 restriction by the full length double-domain APOBEC3G. Nat Commun. 2020 Jan 31;11(1):632.

68. Byeon IJ, Ahn J, Mitra M, Byeon CH, Hercik K, Hritz J, et al. NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. Nat Commun. 2013;4:1890.

69. Mitra M, Hercik K, Byeon I-JL, Ahn J, Hill S, Hinchee-Rodriguez K, et al. Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties. Nucleic Acids Res. 2014 Jan;42(2):1095–110.

70. Sohail A, Klapacz J, Samaranayake M, Ullah A, Bhagwat AS. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. Nucleic Acids Res. 2003 Jun 15;31(12):2990–4.

71. Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc Natl Acad Sci U S A. 2003 Apr 1;100(7):4102–7.

72. Yu K, Huang F-T, Lieber MR. DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. J Biol Chem. 2004 Feb 20;279(8):6496–500.

73. Carpenter MA, Rajagurubandara E, Wijesinghe P, Bhagwat AS. Determinants of sequence-specificity within human AID and APOBEC3G. DNA Repair Amst. 2010 May 4;9(5):579–87.

74. Pham P, Afif SA, Shimoda M, Maeda K, Sakaguchi N, Pedersen LC, et al. Activation-induced deoxycytidine deaminase: Structural basis for favoring WRC hot motif specificities unique among APOBEC family members. DNA Repair. 2017 Jun;54:8–12.

75. Kohli RM, Abrams SR, Gajula KS, Maul RW, Gearhart PJ, Stivers JT. A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. J Biol Chem. 2009 Aug 21;284(34):22898–904.

76. Shi K, Carpenter MA, Kurahashi K, Harris RS, Aihara H. Crystal Structure of the DNA Deaminase APOBEC3B Catalytic Domain. J Biol Chem. 2015 Nov 20;290(47):28120–30.

77. Shi K, Demir Ö, Carpenter MA, Banerjee S, Harki DA, Amaro RE, et al. Active site plasticity and possible modes of chemical inhibition of the human DNA deaminase APOBEC3B. FASEB BioAdvances. 2020 Jan;2(1):49–58.

78. Shi K, Demir O, Carpenter MA, Wagner J, Kurahashi K, Harris RS, et al. Conformational Switch Regulates the DNA Cytosine Deaminase Activity of Human APOBEC3B. Sci Rep. 2017 Dec 12;7(1):17415.

79. Harjes S, Solomon WC, Li M, Chen KM, Harjes E, Harris RS, et al. Impact of H216 on the DNA binding and catalytic activities of the HIV restriction factor APOBEC3G. J Virol. 2013 Jun;87(12):7008–14.
80. Ito F, Fu Y, Kao S-C, Yang H, Chen XS. Family-Wide Comparative Analysis of Cytidine and Methylcytidine Deamination by Eleven Human APOBEC Proteins. J Mol Biol. 2017 Jun 16;429(12):1787–99.

81. Wan L, Nagata T, Katahira M. Influence of the DNA sequence/length and pH on deaminase activity, as well as the roles of the amino acid residues around the catalytic center of APOBEC3F. Phys Chem Chem Phys PCCP. 2018 Jan 31;20(5):3109–17.

82. Silvas TV, Hou S, Myint W, Nalivaika E, Somasundaran M, Kelch BA, et al. Substrate sequence selectivity of APOBEC3A implicates intra-DNA interactions. Sci Rep. 2018 May 14;8(1):7511.

83. Dancyger AM, King JJ, Quinlan MJ, Fifield H, Tucker S, Saunders HL, et al. Differences in the enzymatic efficiency of human and bony fish AID are mediated by a single residue in the C terminus modulating single-stranded DNA binding. FASEB J Off Publ Fed Am Soc Exp Biol. 2012 Apr;26(4):1517–25.

84. Quinlan EM, King JJ, Amemiyah CT, Hsu E, Larijani M. Biochemical Regulatory Features of Activation-Induced Cytidine Deaminase Remain Conserved from Lampreys to Humans. Mol Cell Biol. 2017 Oct 15;37(20).

85. Abdouni H, King JJ, Suliman M, Quinlan M, Fifield H, Larijani M. Zebrafish AID is capable of deaminating methylated deoxycytidines. Nucleic Acids Res. 2013 May 1;41(10):5457–68.

86. Borzooee F, Larijani M. Pichia pastoris as a host for production and isolation of mutagenic AID/APOBEC enzymes involved in cancer and immunity. New Biotechnol. 2019 Jul 25;51:67–79.

87. Diamond CP, Im J, Button EA, Huebert DNG, King JJ, Borzooee F, et al. AID, APOBEC3A and APOBEC3B efficiently deaminate deoxycytidines neighboring DNA damage induced by oxidation or alkylation. Biochim Biophys Acta Gen Subj. 2019 Nov;1863(11):129415.

88. Larijani M, Martin A. Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. Mol Cell Biol. 2007 Dec;27(23):8038–48.

89. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, et al. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. Nature. 2008 Nov 6;456(7218):121–4.

90. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9:40.

91. Roy A, Yang J, Zhang Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res. 2012 Jul;40(Web Server issue):W471-7.

92. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res. 2011 Jul;39(Web Server issue):W270-7.

93. Grosdidier A, Zoete V, Michielin O. Fast docking using the CHARMM force field with EADock DSS. J Comput Chem. 2011 May 3;
94. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605–12.

95. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W665-667.

96. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. J Chem Theory Comput. 2011 Feb 8;7(2):525–37.

97. Harris TK, Turner GJ. Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. IUBMB Life. 2002 Feb;53(2):85–98.

98. Prochnow C, Bransteitter R, Klein MG, Goodman MF, Chen XS. The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature. 2007 Jan 25;445(7126):447–51.
Figure 1

A) Sequence alignment of Hs-AID, A3A, and A3G-CTD.

B) pH dependence of deamination activity.

C) 3D model of the enzyme and its interaction with ssDNA.

D) SDS-PAGE gel showing protein bands.

E) Deamination activity at different pH levels.

F) Deamination activity at different pH levels for Hs-AID variants.

G) Velocity-substrate concentration plots for Hs-AID and Hs-AID variants.
Figure 3
Figure 4
Figure 5