University of Massachusetts Medical School

eScholarship@UMMS

Schiffer Lab Publications

Biochemistry and Molecular Pharmacology

2018-11-20

Structural analysis of the active site and DNA binding of human cytidine deaminase APOBEC3B

Shurong Hou
University of Massachusetts Medical School

Et al.

Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/schiffer

🟡 Part of the Amino Acids, Peptides, and Proteins Commons, Biochemistry Commons, Cancer Biology Commons, Enzymes and Coenzymes Commons, Medicinal Chemistry and Pharmaceutics Commons, Medicinal-Pharmaceutical Chemistry Commons, Molecular Biology Commons, and the Structural Biology Commons

Structural analysis of the active site and DNA binding of human cytidine deaminase APOBEC3B

Shurong Hou, Tania Silvas, Florian Leidner, Ellen A. Nalivaika,
Hiroshi Matsuo, Nese Kurt Yilmaz, and Celia A. Schiffer

Just Accepted

Structural analysis of the active site and DNA binding of human cytidine deaminase APOBEC3B

Shurong Hou[1], Tania V. Silvas[1], Florian Leidner[1], Ellen A. Nalivaika[1], Hiroshi Matsuo[2], Nese Kurt Yilmaz[1] and Celia A. Schiffer[1,]*

[1]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01655, USA

[2]Basic Research Laboratory, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA

*Corresponding author: celia.schiffer@umassmed.edu

ABSTRACT

APOBEC3s proteins (A3s), a family of human cytidine deaminases, protect the host from endogenous retro-elements and exogenous viral infections by introducing hypermutations. However, overexpressed A3s can modify genomic DNA to promote tumorigenesis, especially A3B. Despite overall similarity, A3 proteins have distinct deamination activity. Recently determined A3 structures have revealed the molecular determinants of nucleotide specificity and DNA binding. However, for A3B, the structural basis for regulation of deamination activity and the role of active site loops in coordinating DNA had remained unknown. Using advanced molecular modeling followed by experimental mutational analysis and dynamics simulations, we investigated molecular mechanism of DNA binding by A3B-CTD. We modeled fully native A3B-DNA structure, identified Arg211 in loop 1 as the gatekeeper coordinating DNA and critical residues for nucleotide specificity. We also identified a unique auto-inhibited conformation in A3B-CTD that restricts access and binding of DNA to the active site. Our results reveal the structural basis for DNA binding and relatively lower catalytic activity of A3B and provide opportunities for rational design of specific inhibitors to benefit cancer therapeutics.

KEYWORDS

APOBEC3, structural analysis and modeling, DNA binding, molecular dynamics simulations

INTRODUCTION

APOBEC3s (A3s) are a family of cytidine deaminases that catalyze a zinc-dependent cytidine to uridine reaction on single strand DNA (ssDNA) or single strand RNA (ssRNA) [1-4]. The family comprises of seven members that have either one (A3A, A3C, A3H) or two (A3B, A3D, A3F, A3G) zinc-binding domain [5]. The two-domain A3s have a pseudo-catalytic N-terminal domain (NTD) and a catalytically active C-terminal domain (CTD). A3s play a key role in innate immunity by protecting the host cell from exogenous viral infections and endogenous retro-elements through introducing G to A hypermutations [6-11]. However, when overexpressed, their mutagenic activity can also cause modification of genomic DNA and thus promote tumorigenesis [12-14]. A3B has been identified as a significant enzymatic source of mutagenesis in a variety of cancers [15]. Endogenous A3B is involved in the restriction of retro-element LINE-1 [16] and HBV [17-18]. However when overexpressed, A3B can mutate the host genome to trigger cancer phenotypes [12]. The up-regulation of A3B in tumors is correlated with both dispersed and clustered high occurrence of cytidine mutations, p53 (tumor protein 53) inactivation, and poor patient outcome in cancer treatment [12, 19-22]. In addition, the genomic mutations preferentially occur at 5' -TCA, 5' -TCG, and 5' -TCT trinucleotide motifs, which resemble the substrate preference of A3B in biochemical assays [19, 21, 23]. Unlike other cancer sources, A3B can actively create genomic mutations, which means that a growing number of DNA mutations will be created in cancer cells. This will further benefit cancer evolution, for instance, to help escape immune monitoring, outgrow, metastasize, and potentially acquire resistance to therapeutic treatments [24]. Hence, in addition to its non-essential nature [25], A3B represents a promising target for novel anti-cancer drug development.

Over the past several years, crystal and NMR structures of human A3s (A3A, A3C; CTDs of A3B, A3F, A3G) in the apo state have been determined by our group [26-32] and others [33-46]. In general, A3 proteins are structurally highly similar despite their distinct deamination activities. Even though full-length A3B has 5-6 fold higher activity compared to A3B-CTD, A3B-CTD alone can deaminate cytidine in ssDNA but A3B-NTD is catalytically inactive [45, 47-48]. Among all human A3s, A3B-CTD and A3A share the highest sequence identity (**Figure 1A**); however, A3A is about 15-fold more active compared to A3B-CTD [45]. The overall A3 domain structure consists of six alpha-helices and five beta-strands with the zinc-binding region in the middle. In fact, based on the amino acid sequence or even the available structures, it is not apparent what molecular mechanisms are responsible for varied ssDNA binding affinity and deamination activity among A3 domains, including the catalytically inactive NTDs. Recently, our laboratory [49], along with two other groups, have determined the crystal structures of three A3–DNA complexes (A3A-DNA, chimeric A3B-DNA and rA3G-DNA) [46, 50]. When A3A binds to DNA, two major changes occur at the active site involving the side chain of Tyr132, which stacks against the DNA, and the gatekeeper His29, which locks the DNA in the active site. To facilitate crystallization of DNA-bound A3B, loop 1 of A3A was swapped into A3B to determine the structure. However, differences in loop 1 between the two A3s are mainly responsible for the difference in catalytic activity, as swapping loop 1 of A3B-CTD by that of A3A increases A3B-CTD activity by 10-fold [45]. Loop 1 also exhibits the largest amino acid sequence difference between A3A and A3B-CTD (**Figure 1A**) and is longer in A3B with a three residue insertion $_{206}$PLV$_{208}$. In addition, the DNA gatekeeper residue His29 in A3A[49] is missing in A3B and is likely replaced by one of the arginines within the unique triple arginine patch $_{210}$RRR$_{212}$. However, the role of loop 1, and identifying whether Arg211 or Arg212 might be the gatekeeper

residue, in coordinating DNA and in regulating A3B's catalytic activity could not be revealed by crystal structures determined to date.

To elucidate the molecular mechanism of A3B-DNA recognition and how A3B-CTD structurally regulates its catalytic activity compared to A3A, we used a combination of molecular modeling with molecular dynamics (MD) simulations and experimental mutational analysis followed by fluorescence-anisotropy based DNA binding assays. We identified the key role of loop 1 in A3B binding to DNA and down-regulating its activity. We present a structural model of the A3B–DNA complex that elucidates the molecular mechanism and determinants of DNA binding to A3B-CTD. The model and mutational verification identified Arg211 as the gatekeeper for locking DNA into the active site, which is further stabilized by Arg212. We also identified an auto-inhibited conformation in A3B-CTD that is unique among human A3s, resulting from differences in loop 1 length and sequence, which explains the relatively low catalytic activity of A3B. Overall, our results shed light into the structural regulation of A3 activity and differences in loop 1 coordination around the bound DNA, which may potentially lead to discovering anti-cancer drugs to benefit cancer therapeutics.

RESULTS AND DISCUSSION

Despite overall similarities, the length and sequence differences of loop 1 between A3B-CTD and A3A (**Figure 1**) are responsible for alterations in the active site and likely the differences in DNA binding and deamination activity. To elucidate the mechanism of DNA binding by A3B-CTD, multiple MD simulations were performed of both apo and DNA-bound structures of fully wild-type (WT) A3B-CTD, and compared with A3A (**Table S1**). These

mechanisms were further validated by a series of experimental fluorescence anisotropy-based DNA binding assays of A3B-CTD variants (**Table 1**).

Molecular mechanism of A3B-DNA recognition

The role of loop 1 and molecular mechanism of DNA binding had remained unknown for A3B, as recently determined A3B-CTD DNA co-crystal structure is a chimera with the crucial loop 1 swapped from A3A [50]. Here careful molecular modeling was used using available crystal structures to answer for A3B: 1. How does DNA bind to A3B? 2. Which residue is the gatekeeper for latching DNA in the active site? 3. How does A3B define its substrate specificity for thymidine over cytidine at -1 position? To address these questions, WT A3B-CTD bound to substrate DNA containing a TCG trinucleotide motif was modelled based on the crystal structures of apo A3B-CTD and A3A–DNA complex (see Methods for details). The quality of the complex models was further examined through both computational analysis of 100 ns MD simulations (**Table S1**), and experimental DNA binding assays of inactive A3B-CTD variants (**Table 1**). All the MD simulations of DNA-bound structures converged and were stable over the 100 ns trajectory time.

Arg211 is the gatekeeper residue sequestering DNA in the active site

As A3B-CTD has a unique $_{210}RRR_{212}$ patch in loop 1 instead of $_{28}RH_{29}$ compared to A3A (**Figure 1A**), either Arg211 or Arg212 could be the gatekeeper for DNA binding. To identify the critical residue, DNA-bound models were generated with either Arg211 or Arg212 latching DNA in the active site and performed and analyzed triplicate 100 ns MD simulations. The Arg212 model DNA complex was much less stable during the MD simulations compared to either A3A or Arg211 model as indicated by the considerably larger root-mean-square

fluctuations (RMSFs) of bound DNA, especially at the two termini (**Figure 2A; Movie S1**). In both A3A structure and Arg211 model, Glu255 consistently interacted with substrate cytidine 98.94% and 96.29% of the simulation time, respectively (**Figure 2B**). However in the Arg212 model, the contact frequency was decreased to only 60.21%, which suggests poor quality of the model. Thus, the stability over MD simulations indicated that Arg211 rather than Arg212 is the gatekeeper for DNA binding in A3B-CTD.

To further verify our results from computational analysis, we experimentally generated A3B-CTD R210A, R210K, R211A, R211H, R212A and R212H inactive variants and measured DNA binding using fluorescence anisotropy-based assay (**Table 1; Figure S1**). The low binding affinity and catalytic activity of WT A3B-CTD poses challenges in assessing changes in deamination rates and DNA binding. Hence, based on previous studies [31, 51], both TAMRA-labeled linear (in poly A background) and hairpin DNA with A3B preferred sequence TCG in the stem loop were tested in order to minimize background non-specific binding and promote binding affinity. Despite low affinity, the A3B-CTD inactive variant could bind to both linear and hairpin DNA. Both R210A and R210K variants were able to bind DNA but with decreased binding affinity, which is in agreement with Arg210's role in stabilizing overall structure through the conserved hydrogen bond network in apo crystal structure (**Figure S2; Figure S5C**). R212A variant bound to DNA with same affinity as wild type A3B, which confirmed that Arg212 is not the gatekeeper for DNA binding. In contrast, R211A mutant lost binding completely to both ssDNA and hairpin DNA. In agreement with DNA binding assay results, which we have shown to correlate with catalytic activity [51], recent A3B catalytic activity studies have also shown that R211A mutant lost deamination activity (**Table 1**). Thus experimental DNA binding assay data were in agreement with our model and that Arg211 is the critical residue for DNA binding.

In A3A, His29 is the gatekeeper for DNA binding through both hydrogen bond interactions to DNA backbone and stacking interactions to +1 base [49]. A3B-CTD R211H variant, however, showed no binding to DNA. Rather the two types of interactions His29 makes with bound DNA in A3A can be assigned separately to Arg211 and Arg212 in A3B, as Arg211 hydrogen bonds to DNA in the active site while Arg212 stabilizes DNA binding through either stacking or hydrogen bond interactions with the +1 DNA base. In agreement with this mode of binding, in contrast to R211H, the R212H variant was able to bind DNA with comparable affinity as WT A3B-CTD (**Table 1**). During the MD simulation for R212H in complex with DNA, residue R212H formed pi stacking and hydrogen bond interactions with +1 G base 49% and 47% of the time, respectively, which is actually slightly higher than Arg212 (43% and 44%, respectively). The cyclic histidine side chain may facilitate better stacking interactions, leading to the slightly improved binding and catalytic activity (**Table 1**). Therefore, our computational model of fully native DNA-bound A3B-CTD structure was verified by both MD simulations and experimental mutational analysis.

ssDNA binding to A3B-CTD

The DNA-bound model of A3B-CTD revealed the molecular mechanism as well as the role of loop 1 for DNA binding to A3B-CTD. Overall, ssDNA bound to A3B-CTD in a U-shape similar to A3A (**Figure 3A**). Compared to the apo crystal structure, loop 1 underwent major conformational changes to open up the active site for DNA binding (**Figure 3B**), especially at Arg211 which stacks against Tyr315 to close the active site in apo structure (**Figure 1B**). In addition, the side chain of Tyr315 rotated from a dihedral angle $\chi1$ of ~180° to ~60° as in A3A to accommodate DNA binding (**Figure 3B**). In general, our model overlaid well with chimeric A3B-CTD DNA co-crystal structure, but provided critical information on loop 1 conformation in

DNA-bound form (**Figure 3C**), and conformational changes needed to open up the active site to allow DNA binding. The critical DNA binding residues were also examined in terms of intermolecular van der Waals (vdW) interactions (**Figure 3D, E**) and hydrogen bonds (**Figure 4**). The total vdW contacts between A3B-CTD and ssDNA is about -84.3 kcal/mol, which is predominantly contributed from the loops around the active site, loop 1 (-37.8 kcal/mol), loop 3 (-15.9 kcal/mol) and loop 7 (-15.9 kcal/mol). Significantly, loop 1, which is missing in the chimeric A3B-CTD DNA crystal structure, contributed 45% of the total vdW contacts. The most critical intermolecular interaction between A3B-CTD and ssDNA involved the gatekeeper Arg211. Arg211 coordinated DNA binding through both hydrogen bond interactions with the phosphate backbone of -1 T, 0 C and +1 G bases and hydrophobic interactions with DNA backbone (**Figure 4A, C**). Arg212 instead stabilized DNA binding through either stacking (**Figure 4A, B**) or hydrogen bond interactions with the +1 G base.

A3B should be able to bind pre-bent or hairpin DNA based on our model as observed for A3A [51] and likely with higher affinity as the entropic cost of bending the DNA would be decreased. Accordingly, we observed higher binding affinity to hairpin DNA (K_d = 2.0 μM) compared to linear DNA (K_d = 5.4 μM) for A3B (**Table 1**). Interestingly, unlike A3A [51-52], A3B-CTD showed no binding to RNA hairpin (data not shown). These structural and functional differences between A3A and A3B-CTD despite their high sequence similarity might have implications for their biological function as well as cellular localization.

D314 defines substrate specificity for thymidine over cytidine at -1 position

In A3A, the substrate specificity for thymidine over cytidine at -1 position is determined through hydrogen bond interactions with Asp132 [49]. In our A3B–DNA model, the same hydrogen-bonding pattern between -1 T base and Asp314 as in A3A (**Figure 4A,D; Figure S3A**)

was observed. Specifically, O2 atom of -1 T formed a direct hydrogen bond with Asp314 backbone, while OD1 and OD2 atoms of Asp314 had both direct and water-mediated hydrogen bonding with N3 and O4 of -1 T base. All these hydrogen bond interactions were stable during MD simulations (**Figure S3B**). In contrast, when thymidine was changed to a cytidine, the side chain hydrogen bond interactions between the -1 C and Asp314 were disrupted and the DNA was destabilized as indicated by increased dynamics in the active site (**Figure S3C,D; Movie S2**). Similarly, in the Arg212 model, which we deduced to be poor based on dynamics above, the hydrogen bonds of Asp314 with -1 T were destabilized and not reproducible among the replicate simulations (**Figure S4A,B; Movie S1**). These findings suggest that A3B likely uses the same molecular mechanism to determine the substrate specificity as A3A at -1 position since Asp314 is conserved and the DNA interactions maintained between the two A3s.

Structural mechanism of auto-inhibited conformation of apo A3B-CTD

The crystal structure of apo A3B-CTD has a closed active site conformation that results from the stacking interactions between Arg211 or Arg212 of loop 1 and Tyr315 of loop 7, in contrast to A3A [43, 53] (**Figure 1B**). Therefore, the active site of A3B-CTD has to open up from the closed conformation to accommodate DNA binding (**Figure 3B**). To elucidate the structural basis of the closed active site conformation in A3B-CTD, which can modulate DNA binding and thus deamination, we performed detailed structural analysis on the apo forms of A3B-CTD and A3A.

Closed active site conformation correlates with lower DNA affinity

A structural inspection of A3B-CTD in comparison to the closely related A3A and other A3 domains revealed that $_{206}$PLV$_{208}$ insertion in loop 1 forms a unique hydrogen bond network

with Arg311 and Asp205 (**Figure 1C**). Specifically, the backbone carbonyl oxygen of Pro206 makes a hydrogen bond with NH2 atom of Arg311; Val208 forms two backbone hydrogen bonds with the backbone of Asp205 and NH1 atom of Arg311. The backbone carbonyl oxygen of Asp205 also has a hydrogen bond with NH1 atom of Arg311. Arg311 is conserved among all A3 domains (**Figure S5A**). However, when we superimposed all the available active apo A3 structures (A3A, A3B-CTD, A3C, A3F-CTD and A3G-CTD), the side chain of this conserved Arg was locked in a hydrogen bond network that was distinct from the conformation of Arg311 observed in A3B. (**Figure 1C; Figure S5B, C**). This network involves primarily the backbone atoms of conserved Ser97, Ser99, Asn21 (Gln in A3C and His in A3F-CTD and A3G-CTD) and Asn23 (Lys in A3C, A3F-CTD and A3G-CTD). In A3A, the side chain of Ser97 forms an additional hydrogen bond with this conserved Arg. Rather than with Arg311 in A3B-CTD, these residues form an analogous hydrogen bond network with Arg210 in loop 1 (**Figure S2**). The distinct conformation of Arg311 in A3B-CTD and hydrogen bonding with the $_{206}PLV_{208}$ in loop 1 may contribute to the closed active site conformation of A3B-CTD as well as A3B's lower activity.

To investigate $_{206}PLV_{208}$'s role in regulating activity, 1 μs MD simulations were performed on WT A3A and A3B-CTD as well a variant where the $_{206}PLV_{208}$ sequence was deleted (A3B-CTD-ΔPLV). From the MD simulation trajectories, the stability of the closed active site conformation (**Figure 5**) was monitored. The closed active site conformation in WT A3B-CTD was stable during the MD simulations, as the distance between the side chain of Arg211 in loop 1 and Tyr315 in loop 7 varied around 6 Å, which is within the range of stacking interactions that close the active site. In contrast, the equivalent residues in wild type A3A, Arg28 and Tyr132 had a distance distribution around 15 Å, which indicates the active site is in

the open conformation (**Figure 5A,C**). Interestingly, in A3B-CTD-ΔPLV, Arg211 lost the

stacking interactions with Tyr315. The distance between the side chains of Arg211 and Tyr315

was more than 12 Å during the MD simulations. As a result, the active site conformation was

altered into the open conformation, analogous to that observed in A3A. The more open active

site correlates with higher activity in A3A and A3B-CTD-ΔPLV compared to WT A3B.

The side chain conformation of Tyr315, which is analogous to Tyr132 in A3A, was also

monitored during the MD simulations as an indicator for the compatibility to bind DNA (**Figure

5B, C**). The Tyr side chain has to undergo a conformer change to accommodate binding of DNA

at the active site [49-50] (**Figure 3B**). In WT A3B, the side chain dihedral angle $\chi 1$ of Tyr315

remained around 180° in 99.8% of the time during the MD simulations, in agreement with our

finding that the closed active site conformation in A3B is not compatible for DNA binding. The

same dihedral angle $\chi 1$ of Tyr132 in A3A changed from about 180° to 60° upon DNA binding in

the crystal structure [49]. In the MD simulation, the side chain of Tyr132 in A3A sampled between

the apo and DNA-compatible conformations (68% of the time in DNA-compatible

conformation). In A3B-CTD-ΔPLV, the side chain of Tyr315 sampled two conformations (73%

of time in DNA-compatible conformation) as in A3A. Thus the high sampling frequency of

DNA-compatible side chain conformation of Tyr315 in A3B-CTD-ΔPLV and A3A correlates

with the higher DNA affinity and activity compared to WT A3B.

We also observed auto-inhibited conformation of WT A3B-CTD in the 1 μs MD

simulation, which involved the $_{206}PLV_{208}$ hydrogen bond network with Tyr315 (**Figure S6A**).

Specifically, the OH atom of Tyr315 interacted with both NH2 atom of Arg311 and backbone

carbonyl oxygen of Pro206 through direct and water-mediated hydrogen bonds. As a result, the

side chain of Tyr315 was locked in the DNA incompatible conformation (**Figure 5B, C**). These

hydrogen bonds were stable throughout the MD simulations (**Figure S6B**). To verify the role of this auto-inhibited conformation in down-regulating A3B's activity, the Y315F variant was modeled and subjected to the same 1 μs MD simulations. Phe315 in the Y315F variant lost the ability to interact with $_{206}$PLV$_{208}$ hydrogen bond network, as the hydroxyl group was lost, and was released from the auto-inhibited conformation. As a result, the side chain dihedral angle χ1 of Phe315 sampled the DNA-compatible conformation with 46.4% frequency in MD simulation (**Figure 5B**). The active site of Y315F variant somewhat opened up as the stacking interactions between Arg211 and Phe315 was disrupted during the MD simulation (**Figure 5A, C**). However, the extent of active site opening was less compared to A3B-CTD-ΔPLV and A3A. The distance between Arg211 and Phe315 (~10 Å) was smaller than that in A3B-CTD-ΔPLV (~18 Å) and A3A (~15 Å) (**Figure 5A, C**). Overall, this result suggests that the hydrogen-bonding network involving Tyr315 helps stabilize the closed active site conformation. In agreement with the MD results, the Y315F variant slightly gained DNA affinity in experimental binding assay relative to WT protein, especially for DNA hairpin (**Table 1**). Hence, disrupting the hydrogen-bonding network between residue 315 and $_{206}$PLV$_{208}$ destabilizes the closed active site conformation but is not enough to shift to a fully open active site, as the longer loop 1 with the PLV insertion is critical for the closed active site conformation and thus down-regulating A3B-CTD's activity. Recent studies have shown that removing PLV from loop 1 in A3B increases the enzyme's activity, similarly to the chimeric A3B-CTD with the whole loop 1 swapped from A3A (estimated from gel band intensities from reference [53]); this is in agreement with the more open active site conformation that we observed for A3B-CTD-ΔPLV. Thus, the closed active site conformation observed in modeling and simulations were in complete agreement with

experimental binding and catalytic activity. Together these findings strongly suggest that the PLV insertion in loop 1 is the key for restricting and regulating A3B's deamination activity.

Proline stabilizes the conformation of the longer loop 1 in A3B-CTD for DNA binding

Among all A3s, A3A, A3B-CTD and A3G-CTD share the highest sequence similarity and belong to the Z1 group. Both A3B-CTD and A3G-CTD have a longer loop 1 that includes a proline residue, compared to A3A. Considering proline's unique geometry and rigidity compared to other amino acids, this extra residue may help stabilize the conformation of a longer loop 1. To test this hypothesis, we modeled A3B-CTD P206G variant and examined the RMSF of loop 1 as a measure of dynamics (**Figure 5**). Despite the differences in activity, WT A3B-CTD and A3A have similar levels of loop 1 dynamics with RMSF values varying around 4 Å, as well as WT A3G-CTD. The RMSF of loop 1 in A3B-CTD Y315F mutant and A3B-CTD-ΔPLV were also within 4 Å during the MD simulations. Loop 1 in A3B-CTD P206G variant, however, is highly dynamic with RMSF varying from 1 up to 8 Å indicating that loop 1 bearing the P206G mutation may not be able to stably coordinate DNA. Experimentally, the P206G variant lost the ability to bind both linear and hairpin DNA (**Table 1**). These results indicate the importance of proline in a longer loop1 and consistency of loop 1 dynamics in DNA binding.

Conclusions and implications for DNA binding to other A3s

There are still several A3 domains whose structures are unknown, and no full-length two-domain A3 structure has been determined. The low solubility and DNA affinity of certain A3 proteins have required introducing mutations to be able to structurally and biochemically characterize these proteins in vitro [26-28, 30-32, 39, 43, 46, 49-50], or altogether prevented such characterization especially for NTDs. For instance, WT A3B-CTD has poor solubility and low

binding affinity towards DNA, which makes crystalizing native A3B-DNA complex extremely challenging. The available DNA-bound structure was that of an A3B chimera engineered to increase affinity and promote crystallization. While this structure did not inform on the role of loop 1 in DNA binding, the apo structure of A3B with the native loop1 and A3A–DNA structures enabled computational modeling of WT A3B bound to substrate DNA. Future studies focusing on the active site loops may elucidate differences between A3 family members, including the catalytically active and pseudo-catalytic A3 domains. Similarly, combining experimental structures with computational modeling, verified by simulations and experimental mutational analysis as in this study, can provide insights into the function and DNA binding of other A3s.

A3 proteins have the same overall fold, with highly conserved active sites and yet neither the available structures nor the amino acid sequences offer obvious insights into why they have highly varying catalytic activity, from totally inactive pseudo-catalytic NTDs to the highly active A3A. Instead the seemingly minor diversity in the loops 1, 3 and 7 around the active site may be responsible for regulating A3 activity, which could have implications in regulating the biological function in innate immunity and cancer development. Our results suggest that the length and sequence differences in loop 1, which were missing in the DNA-bound A3B crystal structure [50], are key in regulating activity of Z1 A3 domains (**Figure 7**). A short loop 1, as in A3A or A3B-CTD-ΔPLV, results in high catalytic activity. A longer loop 1 as in A3G-CTD, which includes a proline to stabilize the overall conformation, can form molecular interactions with loop 7 to close the active site, and results in medium activity. Finally, having the auto-inhibited conformation due to the $_{206}PLV_{208}$ hydrogen bond network with Arg311 in addition to a longer loop 1 in A3B-CTD further restricts deamination activity. Thus, the detailed analysis of A3B-CTD structure

here revealed insights into how amino acid differences in loops around the active site can

structurally regulate the relative catalytic activity of A3s despite highly similar overall structure

and conserved active site.

To date, design and development of inhibitors or activators for A3s has proven to be

extremely challenging. Our results provide opportunities for drug design to specifically target

A3B and thus benefit cancer therapeutics. Small molecules that stabilize the unique auto-

inhibited mode of A3B might be able to allosterically inhibit A3B without cross-reacting with

other A3s. Besides, the residue-specific information on regulation of auto-inhibition and closed

active site conformation provides the starting point for engineering A3 domains to achieve

varying catalytic efficiencies or distinct substrate specificity.

MATERIAL AND METHODS

Molecular modeling

A3B-CTD wild type apo structure was modeled based on human A3B-CTD isoform A

sequence and A3B-CTD crystal structure (PDB: 5CQH) through program Modeller 9.15 using

basic modeling. All DNA-bound structures were modeled based on both the apo crystal structure

(PDB: 5CQH for A3B; PDB: 4XXO for A3A) and A3A–DNA co-crystal structure (PDB: 5KEG)

through program Modeller 9.15 using advanced modeling. 5'-TCG motif was used in both

modeling and fluorescence anisotropy-based binding assay as A3B-CTD shows highest

deamination activity against this sequence.[54] The DNA molecule of the bound models and A3A–

DNA co-crystal structures were modified through program Coot to the oligo sequence

(AATCGAA) that was used in the fluorescence anisotropy-based binding assay. The phosphate

groups of 5' A base were removed to prevent strong electronegative environment. AACCGAA was modeled similarly to test the molecular mechanism of substrate preference at -1 position. All DNA-bound structure models were then energy minimized through Protein Preparation Wizard from Schrodinger using default settings.

Molecular dynamics simulations

All molecular dynamics simulations were performed using Desmond[55] from Schrodinger. The models were first optimized using Protein Preparation Wizard. The simulation systems were then built through Desmond System Setup using OPLS3 force field [56]. We used SPC solvation model and cubic boundary conditions with 12 Å buffer box size. The final system was neutral and had 0.15 M sodium chloride. A multi-stage MD simulation protocol was used, which was previously described[57]. Briefly, the system was initially relaxed for 100 ps/stage using Brownian Dynamics NVT (10 K) with gradually reduced restraints (500, 250, 50 force constant) on backbone heavy atoms to solute heavy atoms. This step was followed by simulations using NPT ensemble with gradually increased simulation time (24, 50 and 500 ps) and decreased restraints on the solute heavy atoms to no restraints. The final production stage was performed at 300 K and 1 bar with no restraints using NPT ensemble. 1 μs MD simulations were performed for all the apo structures to ensure final system convergence. The DNA-bound models (A3A/A3B-CTD R211/A3B-CTD R212) were simulated as triplicates to ensure reproducibility for 100 ns each. One round of 100 ns MD simulation was performed for other A3-DNA structures (A3B-CTD with CCG motif and A3B-CTD R212H with TCG motif) to compare with the final A3B-CTD DNA-bound model (**Table S1**).

Analysis of molecular dynamics simulations

The analysis of MD simulations was performed separately for each trajectory as well as the MD simulation triplicates, which help ensure reproducibility and conservation of the results among separate trajectories. The RMSD and RMSF of protein and DNA molecule as well as the protein-ligand contacts diagram were calculated using Simulation Interactions Diagram from Schrodinger. Hydrogen bonds occupancies over the trajectories were calculated using in-house modified Schrodinger trajectory analysis python scripts. Hydrogen bonds were determined for pairs of eligible donor/acceptor atoms using criteria set by Schrodinger: For a pair of heavy atoms to form a hydrogen bond, the distance between donor-hydrogen and acceptor had to be less than 2.8 Angstrom, the angle between donor, hydrogen and acceptor had to be at most 120 degrees and the angle between hydrogen, acceptor and the next atom had to be at least 90 degrees. The residue vdW potential between A3B and DNA during the MD simulations was extracted from the simulation energies using Desmond. For both hydrogen bonds and vdW potential, errors were calculated using block averaging [58]. The distance histograms display the distance between the CZ atom of Arg211 (28 in A3A) and the benzene ring center of the side chain of Tyr315 (132 in A3A). CG, CA, CB and C of Tyr315 or Phe315 were used to determine the side chain dihedral angle. The time series representation of side chain conformations of Arg211 (28 in A3A) and Tyr315 (132 in A3A) in Figure 4C were generated with program VMD using 50 frames as time step (total 2000 frames).

Cloning and mutagenesis of inactive A3B constructs

Human A3B E255A gene was codon-optimized and synthesized by GenScript. This gene was then cloned into pGEX-6p-1 vector using BamHI and EcoRI restriction sites. The pGEX-6p-1 A3B E255A catalytically inactive overexpression construct was used for all experiments in this

study. All the mutations were introduced using the Q5 site-directed mutagenesis kit (NEB), and

the plasmids were sequenced to verify the mutation by Genewiz.

Protein expressions and purification

The pGEX-6p-1 A3B inactive mutant constructs were transformed into BL21 DE3 STAR

E. coli strain for overexpression. Expression of GST-tagged A3B-CTD recombinant protein was

performed at 17 °C for 22 hours in LB medium containing 0.5 mM IPTG and 100 μg/mL

ampicillin. Cells were then pelleted, re-suspended in purification buffer (50 mM Tris HCl pH 7.4;

250 mM NaCl; 0.01% Tween 20 and 1 mM DTT) and lysed with the cell disruptor. The lysate

was collected and the recombinant protein was separated using a GST column. The GST tag was

cleaved by the PreSecission Protease on the column at room temperature overnight. The

flowthrough was then collected and further purified through size-exclusion chromatography

using a HiLoad 16/60 Superdex 75 column (GE Healthcare).

Fluorescence anisotropy-based DNA binding assay

Fluorescence anisotropy-based DNA binding assays were performed as described (28)

with minor modifications. We used 5'-TAMRA labeled oligonucleotides as the binding substrate.

The linear oligonuclotide sequences used were 5'-AAA-AAA-AAA-AAA-AAA-3' (polyA) and

5'-AAA-AAA-AAT-CGA-AAA-3' (polyA TCG). The hairpin sequences used were 5'-GCC-

ATC-ATT-CGA-TGG-G-3' (DNA hairpin) and 5'-rGrCrC-rArUrC-rUrArU-rCrGrA-rUrGrG-3'

(RNA hairpin). The reaction buffer was 50 mM Tris buffer (pH 7.4), 100 mM NaCl, 0.5 mM

TCEP. The concentration of APOBEC3 was varied from 0 to 20 μM in triplicate wells

containing constant amount (10 nM) of substrate. Plates were incubated for an hour on ice before

reading the plates. For all experiments, fluorescence anisotropy was measured using an EnVision

plate reader (PerkinElmer), with excitation at 531 nm and detecting polarized emission at 579

nm wavelength.

Data analysis was performed using Prism 7 with least-square fitting of the measured

fluorescence anisotropy values (Y) at different protein concentrations (X) with a single-site

binding curve with Hill slope and constant background using the equation $Y = (B_{max} \times X^h)/ (K_d{}^h$

$+ X^h) + Background$, where K_d is the equilibrium dissociation constant, h is the Hill coefficient,

and B_{max} is the extrapolated maximum anisotropy at complete binding. The standard deviation

was calculated for each measurement point from three independent repeats.

ASSOCIATED CONTENT

Supporting Information. Table listing the MD simulations performed (Table S1); sample

curves for fluorescence anisotropy detected binding assay (Figure S1); hydrogen bond network

for R210 (Figure S2); comparison of TC versus CC binding by A3B-CTD (Figure S3); hydrogen

bond interactions between A3B-CTD and -1 thymidine in R212 model (Figure S4); amino acid

sequence and structural differences between A3B-CTD and catalytically active A3 domains

(Figure S5); PLV hydrogen bond network (Figure S6).

The following files are available free of charge.

Supporting Information containing the table and figures (PDF)

Movie S1 comparing A3B-DNA R211 model versus R212 model (mp4)

Movie S2 comparing A3B-DNA TCG model versus CCG model (mp4)

AUTHOR INFORMATION

Corresponding Author

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

* To whom correspondence should be addressed.

Email: Celia.Schiffer@umassmed.edu

Author Contributions

SH and CAS conceived and designed the experiments; SH, TVS and EAN performed the wet-lab

experiments; SH and FL performed the computational analysis; SH, NKY, TVS, HM and CAS

analyzed the data; and SH, NKY and CAS wrote the manuscript with contribution from all co-

authors. All authors have given approval to the final version of the manuscript.

REFERENCES

1.   Betts, L.; Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, C. W., Jr., Cytidine deaminase.
The 2.3 A crystal structure of an enzyme: transition-state analog complex. *J. Mol. Biol.* **1994,**
235 (2), 635-56.
2.   Jarmuz, A.; Chester, A.; Bayliss, J.; Gisbourne, J.; Dunham, I.; Scott, J.; Navaratnam, N.,
An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22.
Genomics **2002,** *79* (3), 285-96.
3.   Wedekind, J. E.; Dance, G. S.; Sowden, M. P.; Smith, H. C., Messenger RNA editing in
mammals: new members of the APOBEC family seeking roles in the family business. *Trends
Genet.* **2003,** *19* (4), 207-16.
4.   Conticello, S. G.; Thomas, C. J.; Petersen-Mahrt, S. K.; Neuberger, M. S., Evolution of
the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol.* **2005,**
22 (2), 367-77.

5.   LaRue, R. S.; Andresdottir, V.; Blanchard, Y.; Conticello, S. G.; Derse, D.; Emerman, M.; Greene, W. C.; Jonsson, S. R.; Landau, N. R.; Lochelt, M.; Malik, H. S.; Malim, M. H.; Munk, C.; O'Brien, S. J.; Pathak, V. K.; Strebel, K.; Wain-Hobson, S.; Yu, X. F.; Yuhki, N.; Harris, R. S., Guidelines for naming nonprimate APOBEC3 genes and proteins. *J. Virol.* **2009,** *83* (2), 494-7.

6.   Sheehy, A. M.; Gaddis, N. C.; Choi, J. D.; Malim, M. H., Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **2002,** *418* (6898), 646-50.

7.   Zheng, Y. H.; Irwin, D.; Kurosu, T.; Tokunaga, K.; Sata, T.; Peterlin, B. M., Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. *J. Virol.* **2004,** *78* (11), 6073-6.

8.   Dang, Y.; Siew, L. M.; Wang, X.; Han, Y.; Lampen, R.; Zheng, Y. H., Human cytidine deaminase APOBEC3H restricts HIV-1 replication. *J. Biol. Chem.* **2008,** *283* (17), 11606-14.

9.   Dang, Y.; Wang, X.; Esselman, W. J.; Zheng, Y. H., Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *J. Virol.* **2006,** *80* (21), 10522-33.

10.   Bogerd, H. P.; Wiegand, H. L.; Doehle, B. P.; Lueders, K. K.; Cullen, B. R., APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res.* **2006,** *34* (1), 89-95.

11.   Muckenfuss, H.; Hamdorf, M.; Held, U.; Perkovic, M.; Lower, J.; Cichutek, K.; Flory, E.; Schumann, G. G.; Munk, C., APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem.* **2006,** *281* (31), 22161-72.

12.   Burns, M. B.; Lackey, L.; Carpenter, M. A.; Rathore, A.; Land, A. M.; Leonard, B.; Refsland, E. W.; Kotandeniya, D.; Tretyakova, N.; Nikas, J. B.; Yee, D.; Temiz, N. A.; Donohue, D. E.; McDougle, R. M.; Brown, W. L.; Law, E. K.; Harris, R. S., APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **2013,** *494* (7437), 366-70.

13.   Taylor, B. J.; Nik-Zainal, S.; Wu, Y. L.; Stebbings, L. A.; Raine, K.; Campbell, P. J.; Rada, C.; Stratton, M. R.; Neuberger, M. S., DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *Elife* **2013,** *2*, e00534.

14.   Starrett, G. J.; Luengas, E. M.; McCann, J. L.; Ebrahimi, D.; Temiz, N. A.; Love, R. P.; Feng, Y.; Adolph, M. B.; Chelico, L.; Law, E. K.; Carpenter, M. A.; Harris, R. S., The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat Commun* **2016,** *7*, 12918.

15.   Burns, M. B.; Temiz, N. A.; Harris, R. S., Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* **2013,** *45* (9), 977-83.

16.   Wissing, S.; Montano, M.; Garcia-Perez, J. L.; Moran, J. V.; Greene, W. C., Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J. Biol. Chem.* **2011,** *286* (42), 36427-37.

17.   Xu, R.; Zhang, X.; Zhang, W.; Fang, Y.; Zheng, S.; Yu, X. F., Association of human APOBEC3 cytidine deaminases with the generation of hepatitis virus B x antigen mutants and hepatocellular carcinoma. *Hepatology* **2007,** *46* (6), 1810-20.

18.   Bonvin, M.; Greeve, J., Effects of point mutations in the cytidine deaminase domains of APOBEC3B on replication and hypermutation of hepatitis B virus in vitro. *J. Gen. Virol.* **2007,** *88* (Pt 12), 3270-4.

19.   Leonard, B.; Hart, S. N.; Burns, M. B.; Carpenter, M. A.; Temiz, N. A.; Rathore, A.; Vogel, R. I.; Nikas, J. B.; Law, E. K.; Brown, W. L.; Li, Y.; Zhang, Y.; Maurer, M. J.; Oberg, A.

L.; Cunningham, J. M.; Shridhar, V.; Bell, D. A.; April, C.; Bentley, D.; Bibikova, M.; Cheetham, R. K.; Fan, J. B.; Grocock, R.; Humphray, S.; Kingsbury, Z.; Peden, J.; Chien, J.; Swisher, E. M.; Hartmann, L. C.; Kalli, K. R.; Goode, E. L.; Sicotte, H.; Kaufmann, S. H.; Harris, R. S., APOBEC3B Upregulation and Genomic Mutation Patterns in Serous Ovarian Carcinoma. *Cancer Res.* **2013,** *73* (24), 7222-31.

20.   Sieuwerts, A. M.; Willis, S.; Burns, M. B.; Look, M. P.; Meijer-Van Gelder, M. E.; Schlicker, A.; Heideman, M. R.; Jacobs, H.; Wessels, L.; Leyland-Jones, B.; Gray, K. P.; Foekens, J. A.; Harris, R. S.; Martens, J. W., Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Horm. Cancer* **2014,** *5* (6), 405-13.

21.   Roberts, S. A.; Lawrence, M. S.; Klimczak, L. J.; Grimm, S. A.; Fargo, D.; Stojanov, P.; Kiezun, A.; Kryukov, G. V.; Carter, S. L.; Saksena, G.; Harris, S.; Shah, R. R.; Resnick, M. A.; Getz, G.; Gordenin, D. A., An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **2013,** *45* (9), 970-6.

22.   Periyasamy, M.; Singh, A. K.; Gemma, C.; Kranjec, C.; Farzan, R.; Leach, D. A.; Navaratnam, N.; Palinkas, H. L.; Vertessy, B. G.; Fenton, T. R.; Doorbar, J.; Fuller-Pace, F.; Meek, D. W.; Coombes, R. C.; Buluwela, L.; Ali, S., p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells. *Nucleic Acids Res.* **2017,** *45* (19), 11056-11069.

23.   Glaser, A. P.; Fantini, D.; Wang, Y.; Yu, Y.; Rimar, K. J.; Podojil, J. R.; Miller, S. D.; Meeks, J. J., APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response. *Oncotarget* **2018,** *9* (4), 4537-4548.

24.   Harris, R. S., Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Res.* **2015,** *17*, 8.

25.   Kidd, J. M.; Newman, T. L.; Tuzun, E.; Kaul, R.; Eichler, E. E., Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet.* **2007,** *3* (4), e63.

26.   Chen, K. M.; Harjes, E.; Gross, P. J.; Fahmy, A.; Lu, Y.; Shindo, K.; Harris, R. S.; Matsuo, H., Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **2008,** *452* (7183), 116-9.

27.   Harjes, E.; Gross, P. J.; Chen, K. M.; Lu, Y.; Shindo, K.; Nowarski, R.; Gross, J. D.; Kotler, M.; Harris, R. S.; Matsuo, H., An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J. Mol. Biol.* **2009,** *389* (5), 819-32.

28.   Shandilya, S. M.; Nalam, M. N.; Nalivaika, E. A.; Gross, P. J.; Valesano, J. C.; Shindo, K.; Li, M.; Munson, M.; Royer, W. E.; Harjes, E.; Kono, T.; Matsuo, H.; Harris, R. S.; Somasundaran, M.; Schiffer, C. A., Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure* **2010,** *18* (1), 28-38.

29.   Li, M.; Shandilya, S. M.; Carpenter, M. A.; Rathore, A.; Brown, W. L.; Perkins, A. L.; Harki, D. A.; Solberg, J.; Hook, D. J.; Pandey, K. K.; Parniak, M. A.; Johnson, J. R.; Krogan, N. J.; Somasundaran, M.; Ali, A.; Schiffer, C. A.; Harris, R. S., First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *ACS Chem. Biol.* **2012,** *7* (3), 506-17.

30.   Bohn, M. F.; Shandilya, S. M.; Albin, J. S.; Kouno, T.; Anderson, B. D.; McDougle, R. M.; Carpenter, M. A.; Rathore, A.; Evans, L.; Davis, A. N.; Zhang, J.; Lu, Y.; Somasundaran, M.; Matsuo, H.; Harris, R. S.; Schiffer, C. A., Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. *Structure* **2013,** *21* (6), 1042-50.

31.   Bohn, M. F.; Shandilya, S. M.; Silvas, T. V.; Nalivaika, E. A.; Kouno, T.; Kelch, B. A.; Ryder, S. P.; Kurt-Yilmaz, N.; Somasundaran, M.; Schiffer, C. A., The ssDNA Mutator APOBEC3A Is Regulated by Cooperative Dimerization. *Structure* **2015,** *23* (5), 903-11.

32.   Kouno, T.; Luengas, E. M.; Shigematsu, M.; Shandilya, S. M.; Zhang, J.; Chen, L.; Hara, M.; Schiffer, C. A.; Harris, R. S.; Matsuo, H., Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G. *Nat. Struct. Mol. Biol.* **2015,** *22* (6), 485-91.

33.   Chelico, L.; Pham, P.; Calabrese, P.; Goodman, M. F., APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA. *Nat. Struct. Mol. Biol.* **2006,** *13* (5), 392-9.

34.   Holden, L. G.; Prochnow, C.; Chang, Y. P.; Bransteitter, R.; Chelico, L.; Sen, U.; Stevens, R. C.; Goodman, M. F.; Chen, X. S., Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **2008,** *456* (7218), 121-4.

35.   Chelico, L.; Sacho, E. J.; Erie, D. A.; Goodman, M. F., A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV. *J. Biol. Chem.* **2008,** *283* (20), 13780-91.

36.   Furukawa, A.; Nagata, T.; Matsugami, A.; Habu, Y.; Sugiyama, R.; Hayashi, F.; Kobayashi, N.; Yokoyama, S.; Takaku, H.; Katahira, M., Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J.* **2009,** *28* (4), 440-51.

37.   Chelico, L.; Prochnow, C.; Erie, D. A.; Chen, X. S.; Goodman, M. F., Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. *J. Biol. Chem.* **2010,** *285* (21), 16195-205.

38.   Kitamura, S.; Ode, H.; Nakashima, M.; Imahashi, M.; Naganawa, Y.; Kurosawa, T.; Yokomaku, Y.; Yamane, T.; Watanabe, N.; Suzuki, A.; Sugiura, W.; Iwatani, Y., The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat. Struct. Mol. Biol.* **2012,** *19* (10), 1005-10.

39.   Siu, K. K.; Sultana, A.; Azimi, F. C.; Lee, J. E., Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. *Nat Commun* **2013,** *4*, 2593.

40.   Byeon, I. J.; Ahn, J.; Mitra, M.; Byeon, C. H.; Hercik, K.; Hritz, J.; Charlton, L. M.; Levin, J. G.; Gronenborn, A. M., NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nat Commun* **2013,** *4*, 1890.

41.   Mitra, M.; Hercik, K.; Byeon, I. J.; Ahn, J.; Hill, S.; Hinchee-Rodriguez, K.; Singer, D.; Byeon, C. H.; Charlton, L. M.; Nam, G.; Heidecker, G.; Gronenborn, A. M.; Levin, J. G., Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties. *Nucleic Acids Res.* **2014,** *42* (2), 1095-110.

42.   Lu, X.; Zhang, T.; Xu, Z.; Liu, S.; Zhao, B.; Lan, W.; Wang, C.; Ding, J.; Cao, C., Crystal structure of DNA cytidine deaminase ABOBEC3G catalytic deamination domain suggests a binding mode of full-length enzyme to single-stranded DNA. *J. Biol. Chem.* **2015,** *290* (7), 4010-21.

43.   Shi, K.; Carpenter, M. A.; Kurahashi, K.; Harris, R. S.; Aihara, H., Crystal Structure of the DNA Deaminase APOBEC3B Catalytic Domain. *J. Biol. Chem.* **2015,** *290* (47), 28120-30.

44.   Shaban, N. M.; Shi, K.; Li, M.; Aihara, H.; Harris, R. S., 1.92 Angstrom Zinc-Free APOBEC3F Catalytic Domain Crystal Structure. *J. Mol. Biol.* **2016,** *428* (11), 2307-16.

45.   Byeon, I. J.; Byeon, C. H.; Wu, T.; Mitra, M.; Singer, D.; Levin, J. G.; Gronenborn, A. M., Nuclear Magnetic Resonance Structure of the APOBEC3B Catalytic Domain: Structural Basis for Substrate Binding and DNA Deaminase Activity. *Biochemistry* **2016,** *55* (21), 2944-59.

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

47.   Fu, Y.; Ito, F.; Zhang, G.; Fernandez, B.; Yang, H.; Chen, X. S., DNA cytosine and methylcytosine deamination by APOBEC3B: enhancing methylcytosine deamination by engineering APOBEC3B. *Biochem. J* **2015,** *471* (1), 25-35.

48.   Caval, V.; Bouzidi, M. S.; Suspene, R.; Laude, H.; Dumargne, M. C.; Bashamboo, A.; Krey, T.; Vartanian, J. P.; Wain-Hobson, S., Molecular basis of the attenuated phenotype of human APOBEC3B DNA mutator enzyme. *Nucleic Acids Res.* **2015,** *43* (19), 9340-9.

49.   Kouno, T.; Silvas, T. V.; Hilbert, B. J.; Shandilya, S. M. D.; Bohn, M. F.; Kelch, B. A.; Royer, W. E.; Somasundaran, M.; Kurt Yilmaz, N.; Matsuo, H.; Schiffer, C. A., Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nat Commun* **2017,** *8*, 15024.

50.   Shi, K.; Carpenter, M. A.; Banerjee, S.; Shaban, N. M.; Kurahashi, K.; Salamango, D. J.; McCann, J. L.; Starrett, G. J.; Duffy, J. V.; Demir, O.; Amaro, R. E.; Harki, D. A.; Harris, R. S.; Aihara, H., Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.* **2017,** *24* (2), 131-139.

51.   Silvas, T. V.; Hou, S.; Myint, W.; Nalivaika, E.; Somasundaran, M.; Kelch, B. A.; Matsuo, H.; Kurt Yilmaz, N.; Schiffer, C. A., Substrate sequence selectivity of APOBEC3A implicates intra-DNA interactions. *Sci. Rep.* **2018,** *8* (1), 7511.

52.   Sharma, S.; Patnaik, S. K.; Taggart, R. T.; Kannisto, E. D.; Enriquez, S. M.; Gollnick, P.; Baysal, B. E., APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat Commun* **2015,** *6*, 6881.

53.   Shi, K.; Demir, O.; Carpenter, M. A.; Wagner, J.; Kurahashi, K.; Harris, R. S.; Amaro, R. E.; Aihara, H., Conformational Switch Regulates the DNA Cytosine Deaminase Activity of Human APOBEC3B. *Sci. Rep.* **2017,** *7* (1), 17415.

54.   Liu, M.; Mallinger, A.; Tortorici, M.; Newbatt, Y.; Richards, M.; Mirza, A.; van Montfort, R. L. M.; Burke, R.; Blagg, J.; Kaserer, T., Evaluation of APOBEC3B recognition motifs by NMR reveals preferred substrates. *ACS Chem. Biol.* **2018**.

55.   Bowers, K. J., Chow, E., Xu, H., Dror, R.O., Eastwood, M.P., Gregersen, B.A., Klepeis, J.L., Kolossvary, I., Moraes, M.A., Sacerdoti, F.D., Salmon, J.K., Shan, Y. and Shaw, D.E. In *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*, Proceedings of the 2006 ACM/IEEE conference on Supercomputing, 2006; p 84.

56.   Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016,** *12* (1), 281-96.

57.   Leidner, F.; Kurt Yilmaz, N.; Paulsen, J.; Muller, Y. A.; Schiffer, C. A., Hydration Structure and Dynamics of Inhibitor-Bound HIV-1 Protease. *J. Chem. Theory Comput.* **2018,** *14* (5), 2784-2796.

58.   Flyvbjerg, H.; Petersen, H. G., Error estimates on averages of correlated data. *J. Chem. Phys.* **1989,** *91* (1), 461-466.

Table 1. DNA binding affinity of A3B-CTD inactive (E255A) variants. The binding affinities, represented as Kd, of linear DNA and hairpin DNA with TCG motif in the loop to A3B and variants, determined by fluorescence anisotropy-based assays. All A3B variants contain the E255A mutation to catalytically inactivate the enzyme and prevent substrate deamination. Kd >10 μM indicates weak binding but binding is detectable. NB indicates no detectable binding over the range of concentrations tested (up to 20 μM of A3B-CTD). Representative binding curves for binding (WT), weak binding (R210K) and no binding (P206G) are shown in Figure S2. Activity (fold) indicates catalytic activity relative to wild type A3B-CTD.

A3B-CTD variant	poly A_TCG (μM)	DNA hairpin (μM)	Activity (fold)
WT	5.4 ± 2.6	2.0 ± 0.5	1
Y315F	7.3 ± 4.6	1.8 ± 0.4	1[a]
R212H	NB	1.2 ± 0.3	2.5[b]
R212A	NB	1.4 ± 0.4	-
R211A	NB	NB	0.05[a]
R211H	NB	NB	-
R210A	> 10	> 10	-
R210K	> 10	> 10	-
P206G	NB	NB	-

Footnote:

[a] Shi K et al. (2015). J. Biol. Chem.

[b] Shi K et al. (2017). Sci. Rep. (fold change was estimated from gel band intensities using ImageJ on Figure 6B)
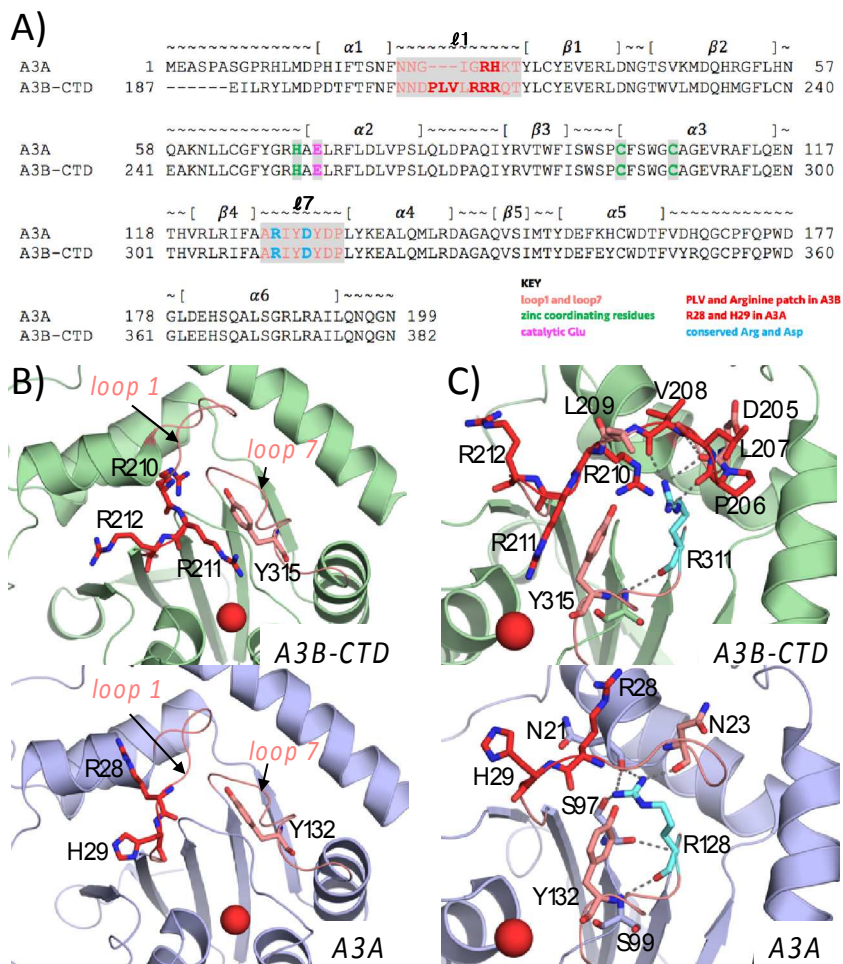
Figure 1. Protein sequence alignment and structure comparison between A3B-CTD and A3A. A)

Amino acid sequence alignment of A3A and the catalytic domain of A3B (A3B-CTD). B) A3B

has a closed active site conformation while A3A has an open active site in crystal structures. C)

Extra PLV residues alter the conformation of the conserved Arg311 in A3B through extensive

hydrogen bond interactions. A3B-CTD (pdb: 5CQH) and A3A (pdb: 4XXO) are shown in

cartoon representation (A3B in green; A3A in slate blue). The catalytic zinc is shown as red

sphere. Loop 1 and loop 7 are colored salmon. $_{210}RRR_{212}$ and $_{206}PLV_{208}$ in A3B and 28RH29 in

A3A are colored red. The conserved arginine is colored cyan. All the labelled residues are shown
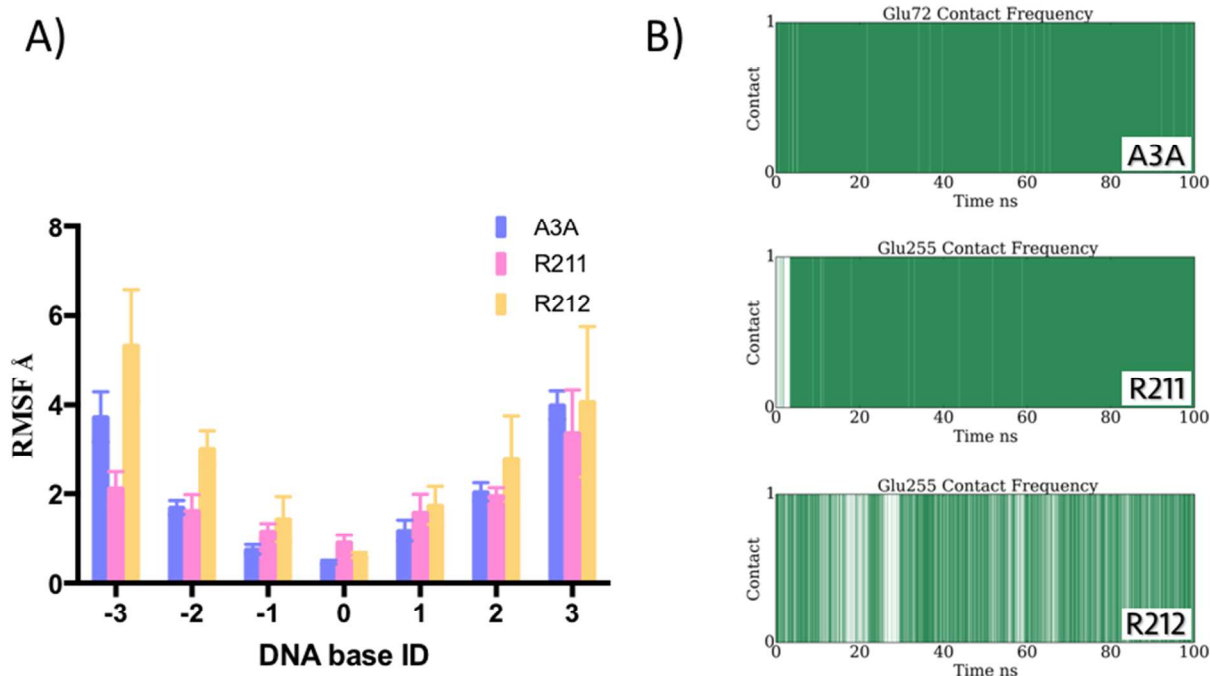
in stick representation.

27

Figure 2. Comparison of A3B-DNA model structures with either R211 or R212 latching the DNA in the active site. A) The root-mean-squared-fluctuations (RMSF) of individual bases of DNA molecule during MD simulations in A3A, A3B-CTD R211 and R212 models. B) The contact frequency of the intermolecular interactions between catalytic residue Glu255 and substrate cytidine base over the simulation time in A3A, A3B-CTD R211 and R212 models.
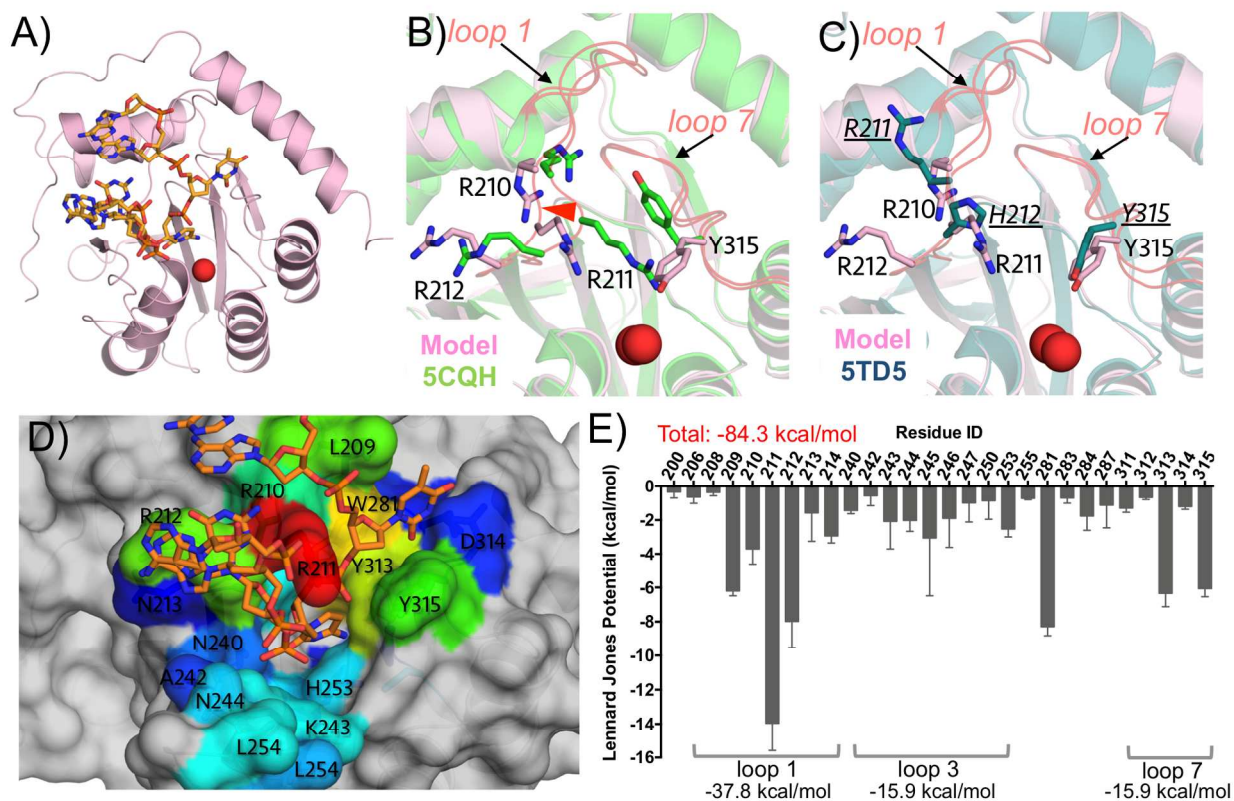
Figure 3. Structural model of A3B-CTD in complex with ssDNA. A) Overall structure of DNA bound A3B-CTD model. B) Conformational changes of residues R210, R211, R212 and Y315 upon DNA binding, with side chains displayed in stick representation. Loop 1 and loop 7 are colored salmon and the conformational changes upon DNA binding are indicated by arrows. C) Structural comparison between the modeled fully native A3B-CTD and the chimeric A3B-CTD DNA crystal structure (PDB: 5TD5). D,E) Mean vdW contacts between protein and ssDNA calculated from triplicate MD simulations. The residues are colored on a rainbow scale from blue to red for increasing contacts; hence warmer colors indicate residues with the most contribution to the intermolecular contacts. The cut-off for the scale is -0.5 kcal/mol.
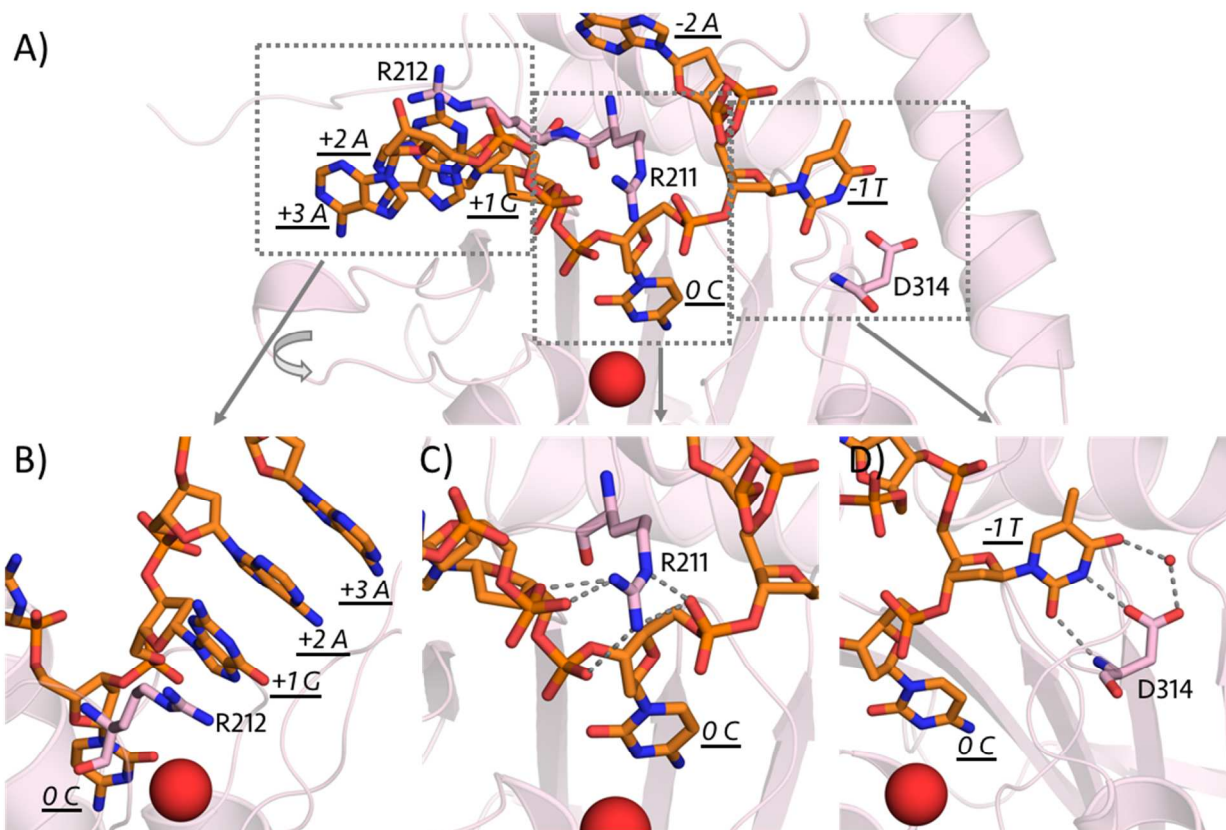
Figure 4. Intermolecular interactions between A3B-CTD and ssDNA. A) Overview of key residues in binding to DNA. B) R212 can stack with downstream DNA bases. C) R211 forms extensive hydrogen bond interactions with DNA backbone. D) D314 makes extensive hydrogen bonds with -1 base that defines substrate specificity. The final frame of representative MD simulation is displayed. The protein is displayed as a pink-colored ribbon diagram and the bound DNA is in orange stick representation. The zinc ion at the active site is depicted as a red-colored sphere. The side chains of R210, R211, R212, D314 and Y315 are shown as sticks. The hydrogen bond interactions are indicated with grey dashed lines.
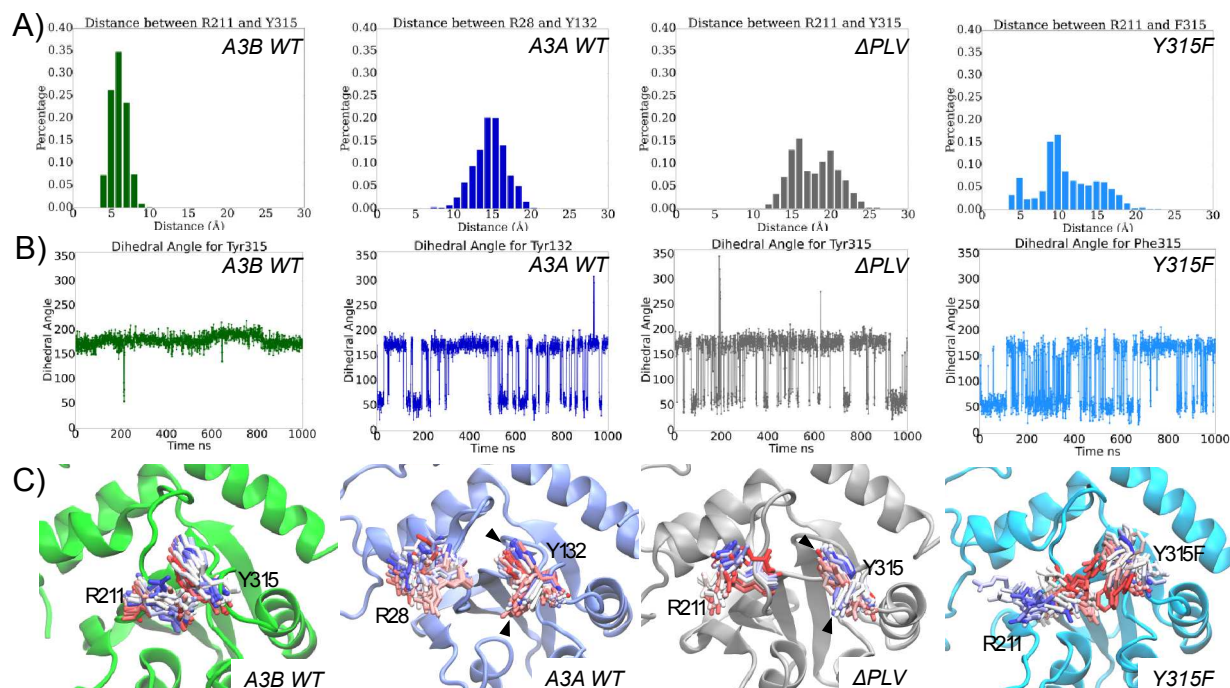
Figure 5. Dynamics of the active site in A3A, A3B-CTD and A3B-CTD mutants. A) The histogram of the distance between CZ atom of Arg311 (Arg28 in A3A) and center of benzene ring of Tyr315 (Tyr132 in A3A) in wild type A3B-CTD, A3A, A3B-CTD-ΔPLV and A3B-CTD Y315F variants during 1 μs MD simulations. B) The dihedral angle of the side chain (C, CA, CB and CG) of Tyr315 (Tyr132 in A3A) over 1 μs MD simulations in wild type A3B, A3A, A3B ΔPLV and Y315F mutants. C) The time series of the side chain conformations of Arg211 (Arg28 in A3A) and Tyr315 (Tyr132 in A3A) (shown as stick; colored based on the simulation time, start as red and end with blue) during 1 μs MD trajectory of wild type A3B-CTD (green), A3A (slate blue), A3B-CTD-ΔPLV (grey) and A3B-CTD Y315F (cyan) variants. Different conformations of Tyr315 are indicated with arrows.
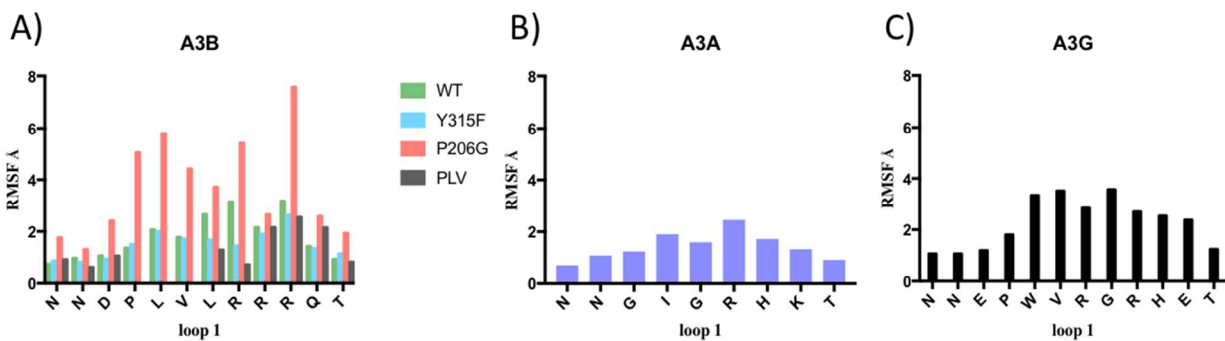
Figure 6. The dynamics of loop 1 during 1 µs MD simulations. A) The root-mean-squared-fluctuations (RMSF) of individual residues of loop 1 in wild type A3B-CTD and A3B-CTD variants. B) The RMSF of all residues in loop 1 of wild type A3A. C). The RMSF of all residues in loop 1 of wild type A3G-CTD.
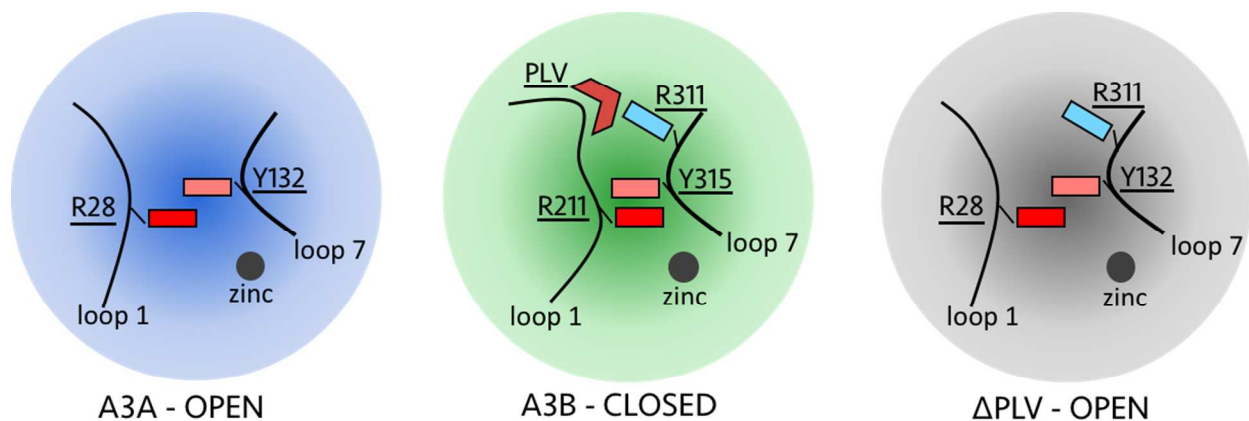
Figure 7. A schematic representation of the mechanism by which A3B-CTD regulates activity. The structural features at the active site of A3A, A3B-CTD and PLV deletion variant that regulate catalytic activity. Loop 1 and loop 7 are shown as lines, and the catalytic zinc is represented as a grey sphere. The side chain of R28, Y132 in A3A and R211, Y315, R311 in A3B-CTD are shown as rectangles. PLV in A3B-CTD is represented as a wedge.

Table of Contents Graphic