Determinants of Substrate Specificity in APOBEC3B

Jeffrey Wagner[1], Özlem Demir[1], Michael Carpenter[2 3] , Hideki Aihara[2 3], Daniel A. Harki[4] , Reuben S. Harris[2 3 5], Rommie E. Amaro[1]

1 Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California, USA.

2 Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, Minnesota, USA.

3 Masonic Cancer Center, Institute for Molecular Virology, University of Minnesota, Minneapolis, Minnesota, USA 55455

4 Department of Medicinal Chemistry, University of Minnesota, Minneapolis, Minnesota, USA.

5 Howard Hughes Medical Institute, University of Minnesota, Minneapolis, Minnesota, USA 55455

Abstract

APOBEC3B (A3B) is a newly discovered driver of mutation in many cancers. We use computational tools to revert a recent crystal structure of an A3B construct to its native sequence, and run molecular dynamics simulations to study its underlying dynamics and substrate recognition mechanisms. The A3B-oligonucleotide substrate simulations show a series of dynamic substrate-protein contacts that correlate with previous work on A3B substrate selectivity. A second series of simulations in which the target cytosine nucleotide was computationally mutated from a deoxyribose to a ribose showed a change in sugar ring pucker, leading to a rearrangement of the binding site and revealing a potential intermediate in the binding pathway. Finally, apo simulations of A3B beginning in the open state experience a rapid and consistent closure of the binding site, reaching a conformation incompatible with substrate binding. These simulations agree with previous experimental studies, and we report the atomistic details of these events to further therapeutic studies on A3B.

Introduction

The APOBEC3 (A3) family of cytidine deaminases is a recently-discovered endogenous source of mutation in cancer.[1,2] Previously, A3 proteins were studied in the context of their interactions with viruses and efforts were undertaken to discover small molecules that modulate the mutational activities of the virally restrictive APOBEC3G enzyme.[3,4] Recent studies have linked cancer progression and recurrence to A3 activity.[5–9] A3 proteins have specific substrate sequence preferences, and analyses of some cancer genomes have shown an enrichment of A3 mutation signatures.[10–15] Recent evidence suggests that APOBEC3B (A3B) is an important driver of tumor progression in the A3 family.[16,17]

A3B is a dual-domain A3 that prefers to deaminate cytosines at TC motifs in DNA, with weak preference placed on further upstream and downstream bases.[18] Each A3 protein consists of either one or two deaminase domains. In A3s with two deaminase domains like A3B, only the second, C-terminal domain (ctd), shows significant catalytic activity. In these dual-domain A3s, the role of the N-terminal domain (ntd) is unknown, but studies have found the full-length construct to be more active than the ctd alone.[19,20] All A3 domains share a minimum sequence identity of 30%, which corresponds to a high degree of overall structure similarity.[21,22] Only single A3 domain structures have been elucidated through X-ray crystallography. Among the A3 members, A3Bctd proved to be one of the most challenging proteins to crystallize, unlike its closest homolog A3A (85% sequence identity).[23] The crystal structure of A3Bctd in apo form could only be elucidated recently after loop 3 truncation and a few stabilizing mutations and showed a tightly closed active site.[24] Further attempts to capture the A3Bctd active site in an open conformation using different crystal forms resulted in alternative, yet still closed conformations of the active site in apo form.[25] Simulations of A3Bctd also indicated the intrinsic bias of its active site toward being in a closed conformation via distinct modes of interaction between loop 1 and loop 7.[25] Finally, a DNA-bound crystal structure of A3Bctd was achieved but only after an additional A3A loop 1 swap in addition to the previously used loop 3 truncation and solubilizing mutations.[26] Despite its high value, the information in the A3Bctd crystal structure still leaves some open questions about the native interactions, especially of loop 1, of wild-type A3Bctd with its DNA substrate.

Protein-DNA recognition is often mediated by hydrogen bonds and shape complementarity.[27–30] Interestingly, the substrate preferences of A3 proteins can be exchanged through the transfer of certain sequence regions.[31] In conjunction with other studies, this work established the role of loops 1, 3 and 7 in the process of DNA substrate recognition.[24,31–35] Initial crystal structures of A3 deaminase domains show these loops being adjacent to the binding site.[22] Subsequent crystallization of A3 proteins in complex with substrates showed residues on loops 1 and 7 binding to oligonucleotide substrates.[24,26,36–38] Previous experimental studies have explored the binding of A3s and other cytidine deaminases to chemically modified oligonucleotides, such as those with a ribose-cytidine (rC) base at the target site, as well as other modifications to the oligonucleotide backbone and base.[20,34,39–43] These studies have concluded that A3B prefers DNA substrates but can bind and catalyze other substrates with significantly lower activity.[44] DNA and RNA differ in structure by only one hydroxyl group (at the C2' position), which is

enough to change the backbone sugar ring pucker from C2' endo (for DNA) to C3' endo (for RNA). It is suspected that this difference in conformational preference is a contributing factor for A3B's selectivity for DNA over RNA, but the exact mechanism is not known.

One powerful technique to understand the biophysics of proteins and biological interactions is molecular dynamics (MD) computer simulations.[45] These simulations model dynamics of all the atoms of a chemical system, starting with an initial geometry and undergoing motion according to the laws of physics at physiologic temperature. MD simulation is now capable of handling not just proteins, but also solvents, ions, and nucleic acids accurately.[46–49] With increases in computing power, more questions have fallen into the scope of computability, and MD simulations have begun to find valuable synergies with traditional biochemistry, often explaining mechanisms underlying biochemical observations in atomic detail or proposing new routes for further experiments.[50–53]

In this work, we use explicitly solvated molecular dynamics simulations to explore long-awaited native interactions of wild-type A3Bctd with oligonucleotide substrates. These simulations reveal the importance of base-specific hydrogen bonds, pocket shape, and backbone sugar conformation in A3Bctd substrate binding. We also simulate an A3Bctd-substrate complex with a ribose-C as the target nucleotide instead of deoxyribose-C, and observe a ring pucker change to the RNA-preferred C3' endo conformation. This pucker change leads to a shift away from the crystallized binding pose, revealing a potential intermediate conformation in A3B-oligonucleotide binding that may provide structural insights for further biochemical or therapeutic work. Correlations between these simulations and experimental findings suggest that these models can accelerate studies of A3B and may generalize to other A3-ssDNA complexes.

Methods

Simulations of A3B were parameterized using the AMBER FF14SB force field for protein atoms, and FF99BSC0 and FF99BSC0_chi0L3 force field for DNA and RNA atoms, respectively.[54–60] The starting coordinates for oligonucleotide-bound simulations were based on PDB entry 5TD5. In each system, mutations were reverted to wild-type and missing residues were modeled using the Schrodinger PRIME software suite.[61,62] Simulations were embedded in a TIP3P water box generated by LEaP from the AmberTools suite with a buffer distance of 10 Å, with Na and Cl ions added to neutralize charge and attain a concentration of 0.2M.[57] Solvent in the crystal structure other than waters were removed. Crystal waters were left in place, and protonation states and hydrogen coordinates were assigned by VMD PropKa.[63,64] The catalytic zinc ion and the zinc-coordinating residues in the active site were modeled according to the Cationic Dummy Atom Model.[65] The catalytic zinc was also modeled bound to a OH^- ion, in order to model the pre-catalysis substrate recognition dynamics of A3B.

Three A3Bctd systems were simulated: A DNA-bound system based on coordinates from 5TD5 with nucleotide sequence 5'-TTCATG-3', a hybrid oligonucleotide-bound system based on coordinates from 5TD5 with nucleotide sequence 5'-TTrCATG-3' (where rC indicates a

ribonucleotide cytidine), and an apo simulation based on coordinates from 5TD5 but with all DNA atoms removed. Each system underwent energy minimization in its force field, followed by gradual heating and equilibration with decreasing restraints. AMBER input scripts for each step are provided in the Supplemental Materials. Each system was simulated in triplicate, differing in temperature initialization seed, and each replicate underwent 1 μs of unrestrained MD simulation in an NPT ensemble at 310 K.

Analysis of hydrogen bonds was performed using the MDTraj Python package[66] and visualized using Matplotlib.[67] The existence of hydrogen bonds was defined by Baker-Hubbard criteria.[68] The hydrogen bond analysis was performed on snapshots taken at increments of 5 ns in the trajectories. Only hydrogen bonds that involve a base atom or the 2' ribose oxygen and that appear in at least 15% of any simulation's snapshots are shown in Table 1. To monitor the sugar pucker of target nucleotide, we used cpptraj.[69–71]

Pocket volumes were studied using POVME3.0[72–74] and visualized using Visual Molecular Dynamics.[75] The pocket region was defined by a set of inclusion spheres which cover the observed DNA-binding region (Figure S1). This region is defined as running between loops 1 and 7, down into the zinc-containing active site pocket, and out between loops 1 and 3. Because quantitative comparison of the pockets was performed, the POVME convex hull exclusion option was not used, per suggested POVME3.0 best practices.[72] All trajectories were aligned by their backbone atoms to the starting structure of the DNA-bound A3B MD simulation (after equilibration).

Results and Discussion

Perturbation of the Target Nucleotide Sugar from C2' Endo to C3' Endo Conformation Can Trigger Substrate Unbinding from A3Bctd Active Site

The simulations containing the ribose-C (rC) nucleotide at the target position displayed major differences from those containing the deoxyribose-C (dC). In the dC simulations, both the protein residues of the binding site and the -1, +1 and target cytidine nucleotides remained in the same position (RMSD < 2.2 Å) throughout the simulations. However, in the second rC simulation, the RMSD of the target nucleotide was much higher, as shown in Figure 1E. This change indicates a shift in DNA-rC binding pose, characterized by a binding site rearrangement. The simulations show that one of the the driving events in this "shift" was a change in the sugar pucker of the rC nucleotide.

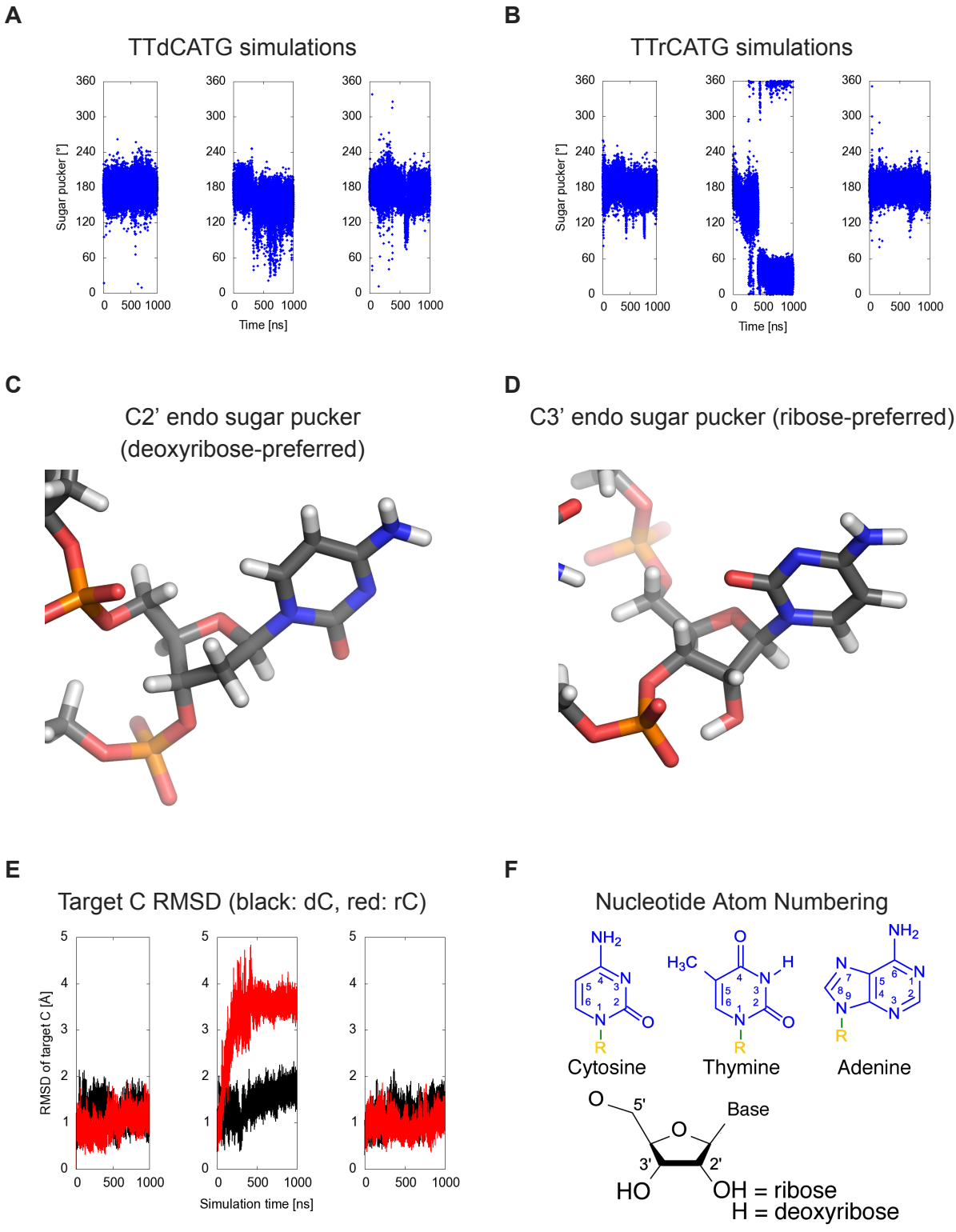

Figure 1: (A) *Sugar pucker of the target C, measured in both A3B-dC and* ***(B)*** *A3B-rC simulations. In rC simulation replicate 2, the RNA transitions from a C2' endo (DNA-preferred) to*

a C3’ endo (RNA-preferred) sugar pucker. ***(C)*** *Geometry of C2’ endo sugar pucker taken from the starting configuration of the dC simulation, and* ***(D)*** *C3’ endo sugar pucker taken from the second rC simulation (middle right).* ***(E)*** *RMSD of target C compared to its initial pose in A3B-DNA (black) and A3B-DNA-rC (red) simulations.* ***(F)*** *Numbering system used to identify atoms in nucleotides.*

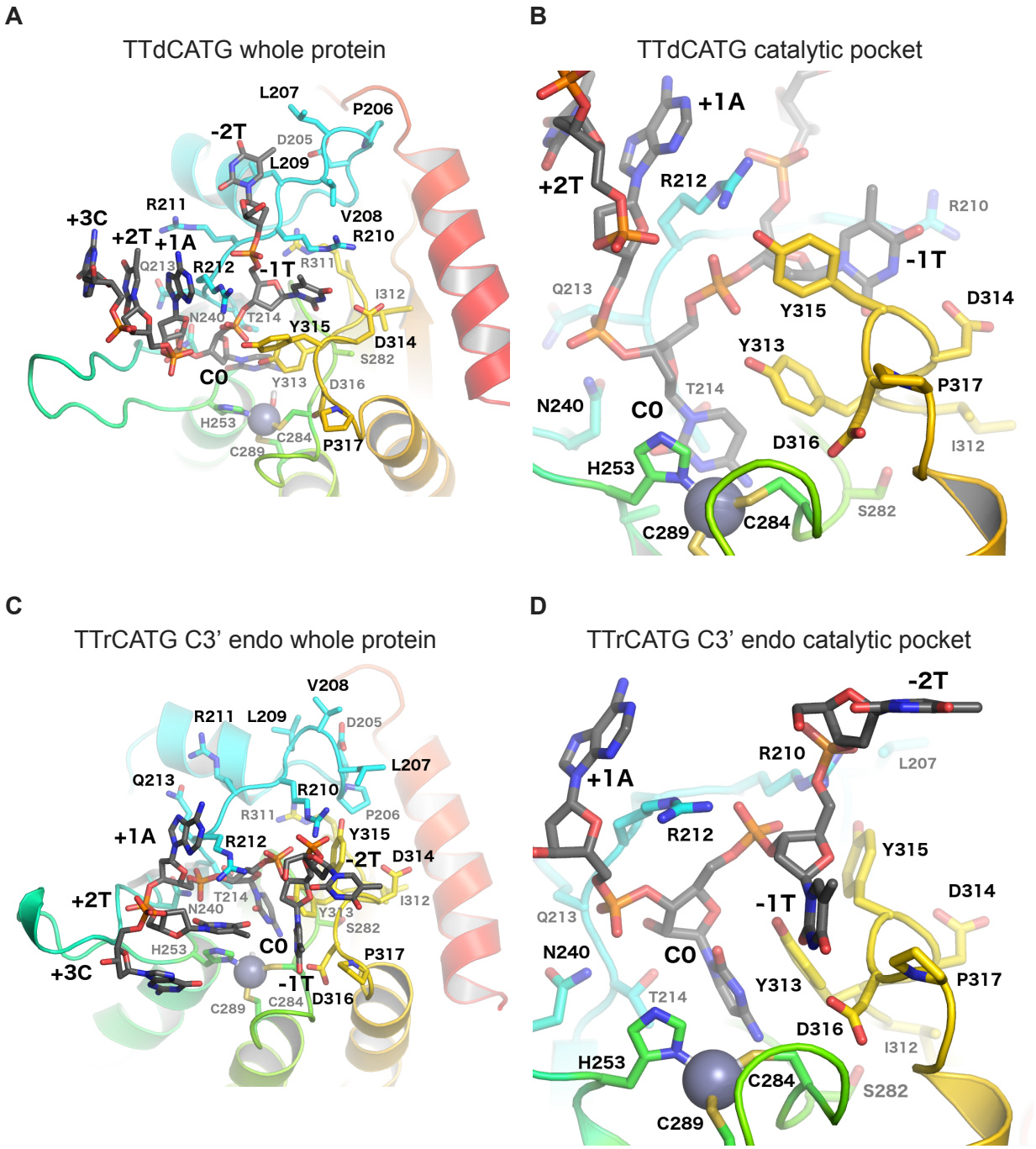

Figure 2: *Binding site-adjacent residues.* ***(A)*** *Whole protein view and* ***(B)*** *catalytic pocket focused view of the simulation with target dC, showing the starting (crystal structure-based) conformation of the substrate oligonucleotide after minimization.* ***(C)*** *Whole protein view and* ***(D)*** *catalytic pocket focused view of the average structure of the C3' endo portion of the rC simulation, as determined by POVME pocket shape analysis. Loop 1 is shown in light blue, loop 3 in dark green, loop 5 in light green, and loop 7 in yellow.*

The rC pucker change correlates with a change in binding site shape. Notably, the change in sugar pucker shifts the DNA-rC substrate toward loop 1 and away from the catalytic glutamate residue and loop 3 (Figure S6). A comparison of average binding pocket shape between the A3B-DNA simulation and C3' endo snapshots of the A3B-DNA-rC is shown in Figure 3.

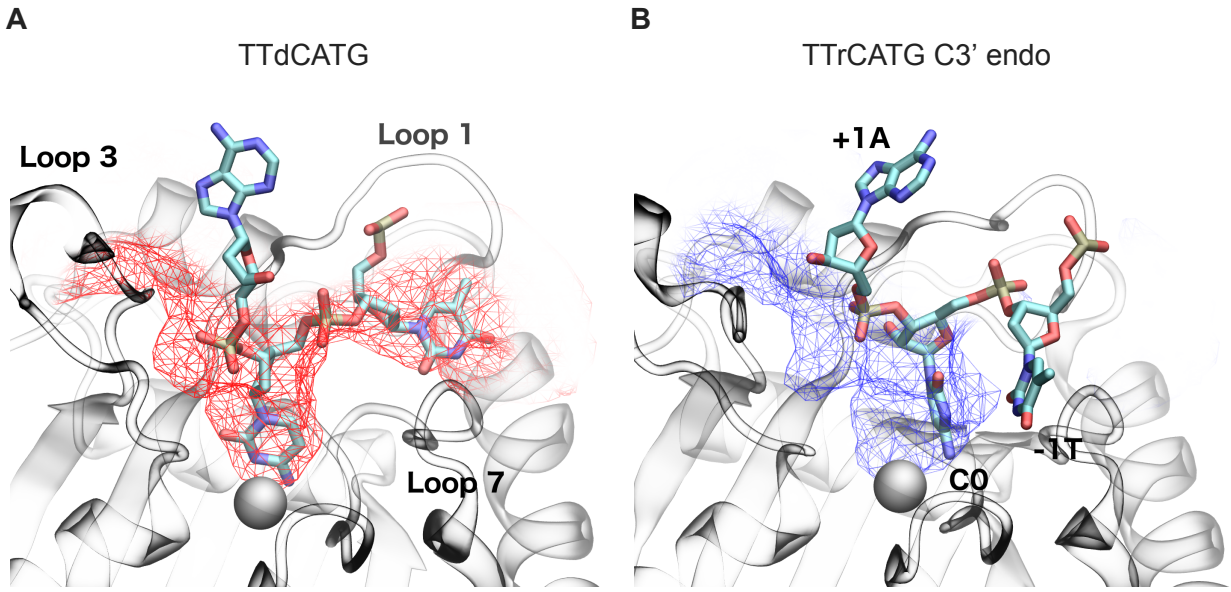

Figure 3: *Average binding cleft shapes differ between* ***(A)*** *dC simulations (C2' endo/deoxyribose, red) and* ***(B)*** *the portion of the rC simulation after the sugar pucker change (C3' endo/ribose, blue). Mesh shows average pocket shape along the DNA binding cleft observed in crystal structures. The atoms in the dC figure structure are the initial coordinates of the simulation, while C3' endo atomic coordinates are taken from the POVME cluster centroid snapshot. The -1 T in the TTrCATG C3' endo snapshots leaves the original substrate binding cleft and is no longer in the pocket defined in the 5TD5 crystal. Images are taken from the same angle and structures are RMSD-aligned. Catalytic zinc is shown as a gray sphere.*

<u>Interaction Footprint of A3Bctd in Complex with Oligonucleotides.</u>

The shifted conformation in the second DNA-rC simulation can also be characterized by movements in the binding pocket residues, and changes in the network of hydrogen bonds made between the protein and target rC. A table of all common hydrogen bonds between base and protein atoms is shown in **Table 1**, and a more detailed time series plot is provided in Figure S2. As the dC simulations do not have the 2' O atom, hydrogen bonds involving it are labeled N/A.

After the ring pucker change, the 2' O atom of the target rC forms a hydrogen bond with Thr214's sidechain. In non-shifted DNA-rC simulations, the Thr214 sidechain maintains a hydrogen bond to the 4' O of the target rC. This shift correlates with a breaking of all major protein-base hydrogen bonds in the other nucleotides of the chain, except that between Ser282

and the amine to be hydrolyzed at the 4-position on the rC. In all rC simulations, the initial hydrogen bond between the target C nucleotide's amine and the catalytic Glu255 residue, is broken partway through each simulation.

Experiments on A3A have shown that Asp131, homologous to A3B's Asp314, confers the preference for T at the -1 position of the oligonucleotide.[26] Both the crystal structures of A3B and our simulations show hydrogen bonds consistently formed between the 3-position NH motif of the T -1 base and the Asp314 sidechain. Interestingly, the rC simulation post-pucker change has broken this hydrogen bond and replaced it with one to the sidechain of Asp316. Asp316 was shown to be essential for A3B antiviral function and is therefore likely involved in DNA binding.[76] Given that this shift only appears in the DNA-rC simulation, it is possible that the perturbation caused by the target C's DNA-to-RNA mutation aided the system in leaving its initial energy well and exploring intermediate binding poses. It is also possible that Asp316 contributes by an indirect electrostatic mechanism when it is not directly forming hydrogen bonds for substrate recognition.

Another less frequent base hydrogen bond with the -1 T intermittently forms with the backbone NH of Tyr315. This hydrogen bond, however, is not specific to the Tyr sidechain, as the contact is to the protein backbone. It is also not specific to the thymine ring, as the 2-position on the nucleobase is also an oxygen atom in the other pyrimidine, cytosine. A3G is the only A3 with a strong -1 preference other than T, and it prefers C. The corresponding position to A3B Tyr315 is A3G Asp317, and A3G is the only A3 with a residue other than Tyr or Phe at that position.

All six nucleotides in the simulation were analyzed, but only hydrogen bonds involving the -1 T, C, and +1 A nucleotides showed greater than 15% frequency. The only hydrogen bond involving the +1 A nucleotide was infrequent, which correlates with the weak preference found toward the +1 position in substrate specificity experiments.[18] Thus, while hydrogen bonding can offer an explanation for the target C and -1 T base specificity, the simulations do not reveal specific sidechain-base hydrogen bonding for other nucleotides. The data implies, however, that electrostatic and/or shape-based recognition may take place. In our simulations, the positively charged sidechains of loop 1 residues contact the negatively charged phosphate backbone of the oligonucleotide. These positively charged loop 1 residues are known to be key for activity in A3A and A3Gctd, as A3A H29 and A3G H216 (homologous to A3B R212) could be mutated to an arginine while maintaining residual activity.[77,78] However, when A3G H216 is mutated to Ala, it loses activity.[78,79] While these backbone contacts appear to be charge-driven and are not specific to one nucleotide sequence, the base atoms of the nucleotide make consistent hydrogen bonds with the loop1 backbone, in what may be a shape-driven recognition process.

Hydrogen Bond	TTCATG	TTrCATG C2' endo	TTrCATG C3' endo
-1 T 3-position NH - Asp314 sidechain O	94%	86%	0%
-1 T 3-position NH - Asp316 sidechain O	0%	0%	96%
-1 T 2-position O - Tyr 315 backbone NH	51%	59%	0%
Target C sugar O2' - Thr214 OH	N/A	3%	94%
Target C sugar O2'H - His253 NE2	N/A	18%	0%
Target C sugar O3' - Asn240 HD21	22%	42%	0%
Target C sugar O4' - Thr214 OH	89%	74%	0%
Target C 2-position O - Ala254 backbone NH	95%	76%	0%
Target C 3-position amine - Ser282 backbone C=O	96%	85%	93%
Target C 3-position amine - Glu255 sidechain O	85%	35%	0%
+1 A 3-position N - Arg212 sidechain guanidinium	21%	17%	42%

Table 1: *Frequent protein-oligonucleotide hydrogen bonds made in either the DNA or DNA-rC simulations. Colors indicate frequency of hydrogen bonds (% of frames in which bonds are made). Dark red: 0%, light red: 1-33%, yellow: 34-66%, green: 67+%.*

After the change of the rC sugar to the C3' endo conformation, the target cytosine is too far from the catalytic glutamate 255 to perform deamination. This new binding mode may be an intermediate conformation in normal DNA binding and/or dissociation, however it was only observed in one of the DNA-rC simulations. Further simulation of all systems might eventually show the same shift, or complete oligonucleotide dissociation.

Notably, the DNA simulations and the C2' endo portion of the DNA-rC simulations maintain a relatively similar DNA interaction interface (Figure 4). While the C3' endo portion of the DNA-rC simulations still has the target C buried in the catalytic pocket, the neighboring nucleotides experience an outward shift and define a new interaction surface. This shift moves the -1 T down loop 7, and partially dissociates the -2 T from the protein. The shift of the target rC away from loop 3 brings the +2 T and +3 G in contact with the surface of loop 3 that faces the catalytic pocket, though our hydrogen bond analysis does not indicate specific interactions driving this association.

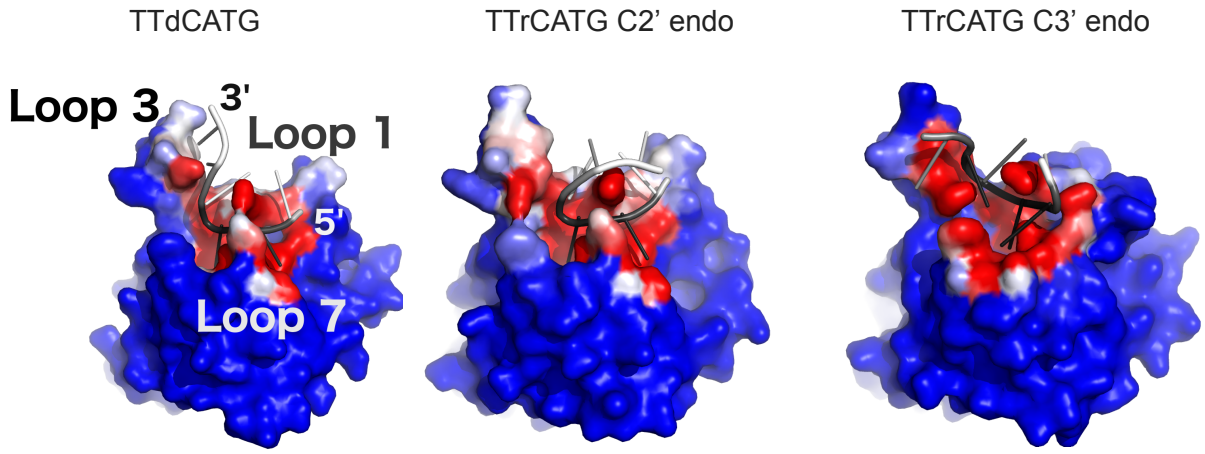

***Figure 4**: Oligonucleotide interaction surface differences between DNA **(left)** and DNA-rC **(center and right)** simulations. Structures from the end of each simulation are shown. Each protein atom is colored by its frequency of contact (<5 A distance) with the oligonucleotide, on a color scale from blue (no contact) to white (50% contact) to red (100% contact). The oligonucleotides are colored by RMSF, with black corresponding to 0 Angstroms and white to 5 Angstroms. The C3' endo portion of the DNA-rC simulation shows a shift in position of the 0 and -1 nucleotides of the substrate, but low RMSD after the shift. Structures are RMSD-aligned.*

Apo A3Bctd simulations show closure of substrate binding cleft

Loops 1, 3, and 7 have been identified as being primarily responsible for substrate recognition, and our data indicate that their interaction patterns are directly affected by the presence of the oligonucleotide (Figure 5). Both the DNA-bound and DNA-rC-bound simulations show significantly fewer loop-loop contacts compared to the apo A3Bctd simulations, commensurate with their high number of loop-oligonucleotide interactions.

The apo simulations show extensive loop 1 - loop 3 interactions, specifically Arg212 and Gln213 to Asn240, Glu241, Ala242, and Lys243. These contacts are made less frequently in the DNA-bound and DNA-rC-bound simulations. This is to be expected, as the substrate is bound between loops 1 and 3. The almost complete loss of loop 1 - loop 3 contacts after the change to the C3' endo sugar pucker is due to the substrate adopting a series of non-specific contacts between the +1, +2, and +3 nucleotides and loop 3, thereby preventing loop 3 residues from contacting loop 1.

The apo simulations also show the most loop 1 - loop 7 contacts. This is to be expected, as the substrate oligonucleotide passes directly between these loops. Arg311 in loop 7 makes contact with most residues in the first half of loop 1, from Asn203 to Arg210. The apo simulation is the

only simulation in which Tyr313 contacts loop 1, primarily via Arg211, but also less frequently through the flanking Arg210 and Arg212. Both the apo and rC C3’ endo simulations show frequent contacts between Tyr315 on loop 7 and Leu209, Arg210, and Arg211 on loop 1. Generally, the large number and frequency of contacts in the apo and rC C3’ endo snapshots indicate a more closed binding site, again implying that the DNA-rC simulation may have captured an intermediate-bound state of the complex.

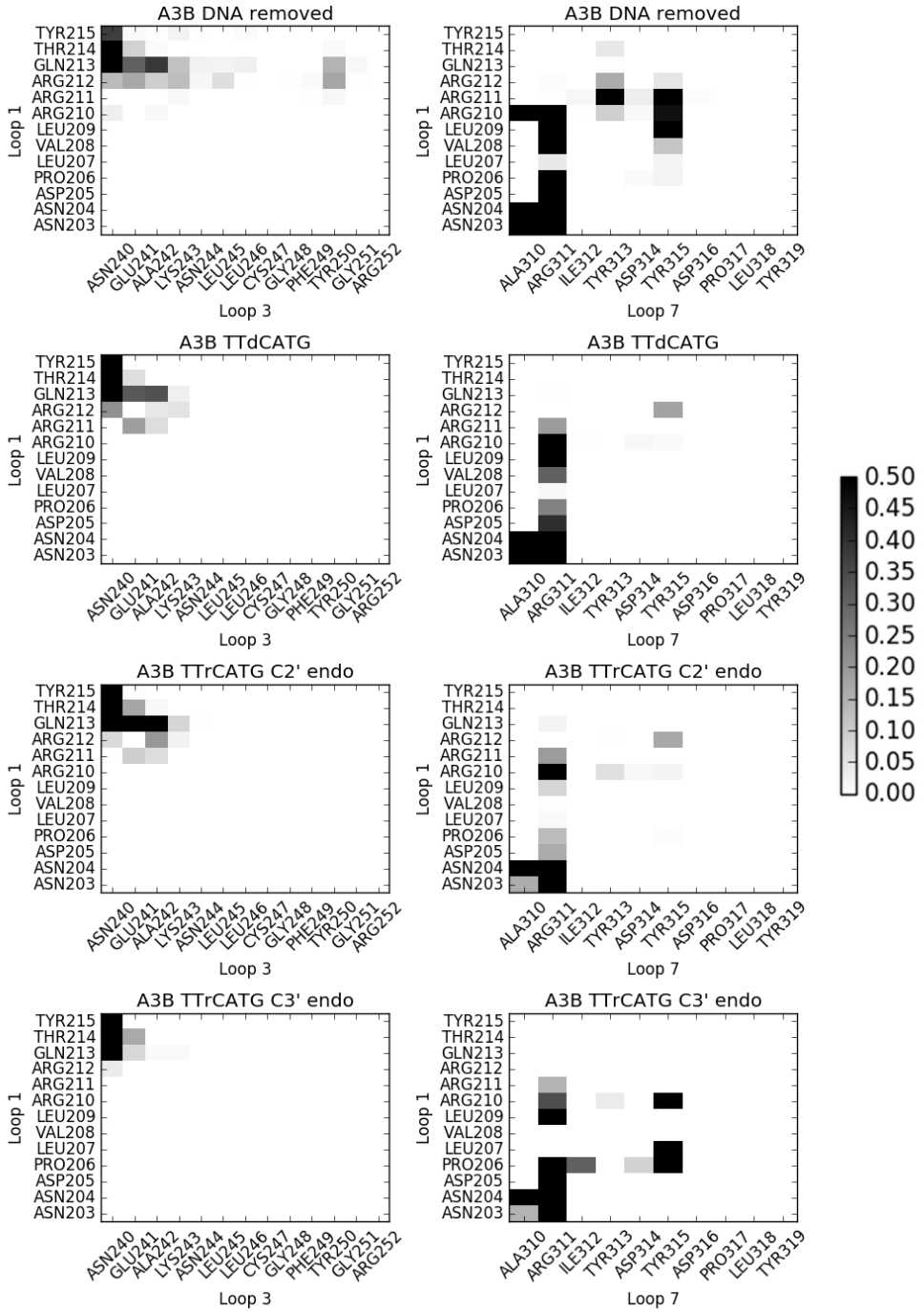

Figure 5: *Frequency of loop-loop contacts in A3B simulations. White indicates infrequent contacts, and black indicates contacts 50% of the time or more. Contacts are defined as a closest heavy atom distance of < 4 Angstroms*

Figure 6 shows an analysis of binding cleft volume of different A3B simulations. The simulation in which DNA was removed shows a rapid closure of the cleft, leading to a protein conformation incompatible with substrate binding. This trend was consistent in all three simulation replicates.

The DNA and DNA-rC simulations maintain an open binding cleft, except for the DNA-rC replicate which experiences a change in sugar pucker and partial unbinding of the substrate. Further, the apo simulation of A3B converges to a much lower pocket volume than was observed in our prior work on A3A (average A3A apo volume shown in magenta).[25] This difference is largely due to A3A's shorter loop 1, and the propensity of the A3A binding cleft to remain open could explain its higher level of deaminase activity relative to A3B.

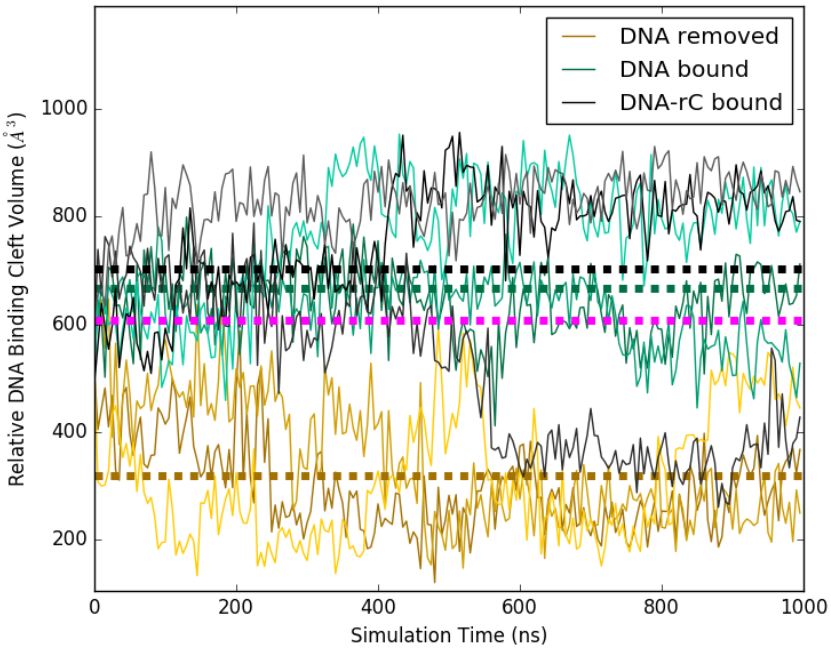

Figure 6: *Relative pocket volumes of APOBEC3B substrate binding cleft during MD simulations. Each simulation was run in triplicate, with replicates shown as lighter and darker shades of the same color. Average pocket volumes are shown as thick dotted lines. Average volume from the same analysis performed on A3A apo simulation used in prior work shown in magenta.*[25]

Conclusions

APOBEC enzymes are of high therapeutic interest, and efforts to use them as targeted biochemical tools, for example as CRISPR/Cas9 fusions, will benefit from a thorough understanding of A3 substrate recognition. Due to the highly dynamic nature of the A3Bctd active site, it has not been possible to capture its native DNA interaction footprint via x-ray crystallography. In this work, we generate a DNA-bound wild-type A3Bctd model from the recent A3Bctd variant crystal structure in complex with DNA, and investigate the long-sought native A3Bctd-DNA interactions and their dynamics using MD simulations. Our simulations offer insight into the mechanisms of substrate recognition and binding. Further, they show significant differences arising from the presence of a non-preferred oligonucleotide.

The analysis of MD simulations show a set of dynamic hydrogen bonds important for substrate recognition at the -1, 0, and +1 positions of the oligonucleotide. Many of these contacts agree with previous experimental findings, while several others provide novel routes for further experiments. The interactions of the substrate-recognition loops with each other are also monitored. The loop-loop interactions of the apo protein provide insight into the thermodynamics and kinetics of substrate binding, as many of these contacts must be broken to accommodate an oligonucleotide.

The perturbation of the DNA substrate to a ribose sugar at only the target nucleotide caused one simulation replicate to undergo a ring pucker change and explore a less tightly-bound conformation of the oligonucleotide substrate. In this novel binding pose, the -1T H-bond with Asp314 is replaced with one to Asp316, the shape of the catalytic pocket is widened, and the target C is shifted away from the catalytic glutamate. This partially unbound pose ismay represent a binding/dissociation intermediate and could be useful in discovering alternate approaches to A3B inhibition.

Acknowledgements
This work was funded in part by the Director's New Innovator Award Program NIH DP2 OD007237 to REA; R01-GM110129 to DAH, RSH, and REA; R01-GM118000 to RSH, HA, and DAH. Funding and support from the National Biomedical Computation Resource (NBCR) is provided through NIH P41 GM103426. JRW was supported by the NIH Molecular Biophysics Training Grant T32 GM008326.

RSH is the Margaret Harvey Schering Land Grant Chair for Cancer Research, a Distinguished McKnight University Professor, and an Investigator of the Howard Hughes Medical Institute.

Conflict of Interest Statement:
DAH & RSH are co-founders, consultants, and shareholders of ApoGen Biotechnologies, Inc.
REA & HA are consultants of ApoGen Biotechnologies, Inc.

REA is a cofounder of, has equity interest in, and is on the scientific advisory board of Actavalon, Inc.

Supplemental Information

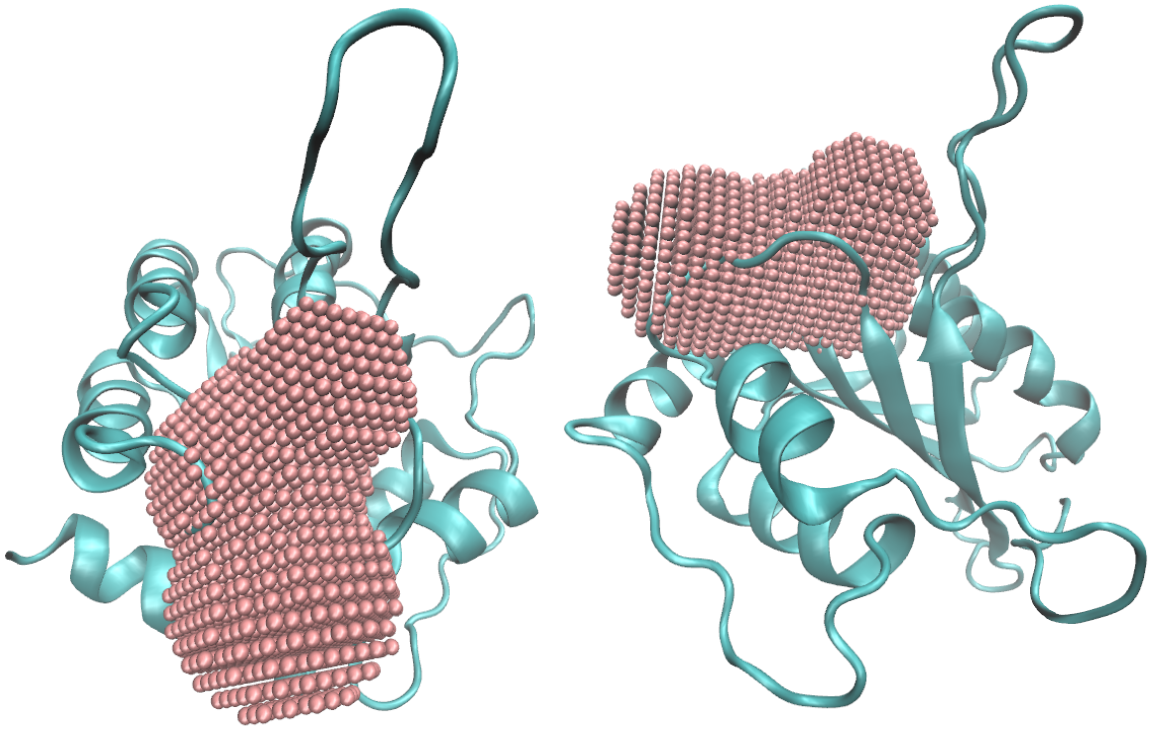

Figure S1: *Inclusion Regions used for POVME analysis*

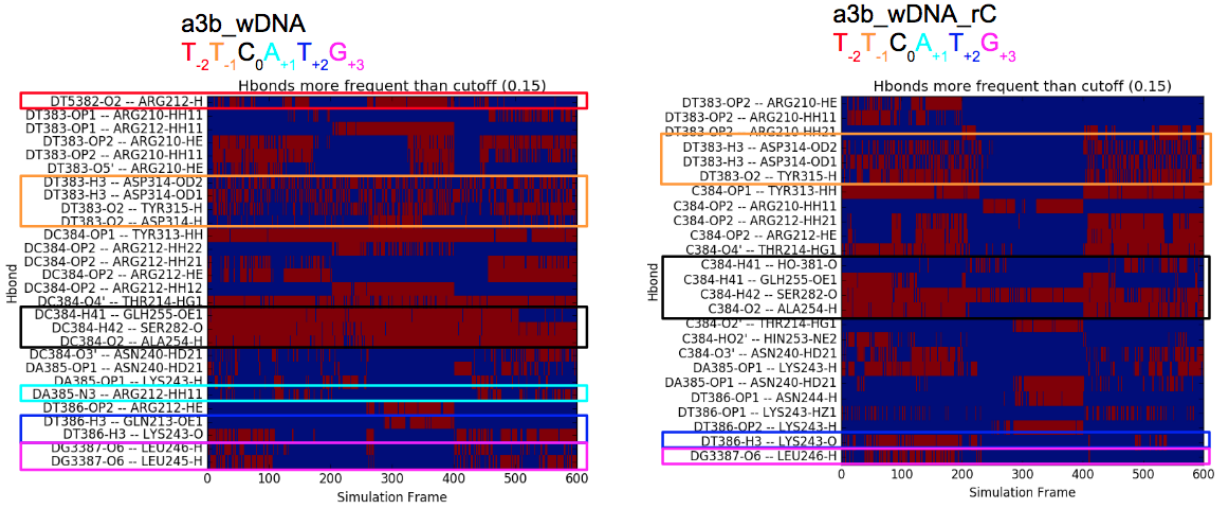

***Figure S2**: Full mapping of protein-oligonucleotide hydrogen bonds more frequent than 15% of simulation snapshots.*

Analysis of crystal structures to discover oligomerization interfaces of A3 domains

Previous work has shown that interaction between A3 domains is an important phenomenon. Evidence for this interaction has been seen both in the activity differences of dual-domain APOBEC3s when expressed as full-length versus as the catalytic domain alone, and in the in-vitro oligomerization of wild type A3 domains which also frequently leads to activity differences. While no full-length A3 crystal structures have yet been solved, the packing of single-domain structures may offer hints to the basis of these observed oligomerizations. Figure 7 shows the frequently-observed crystal packing arrangements of A3 domains, which gives rise to two clusters of interfaces.

This work was performed using open-source PyMOL.[80] The structures considered in this analysis were 5CQD[81], 5CQK[81], 5CQI[81], 5CQH[81], 4XXO[37], 2M65[82], 3VM8[83], 3VOW[83], 3WUS[84], 4J4J[85], 4IOU[86], 5HX4[87], 5HX5[87], 2MZZ[88], 5K81[36], 5K82[36], 5K83[36], 2JYW[79], 3E1U[89], 2KBO[90], 2KEM[21], 3IQS[89], 3IR2[91], 3V4K[3], 3V4J[3], 4ROV[92], and 4ROW.[92] Interface surface area was calculated using the EPPIC webserver.[93] Interfaces were filtered to only show those with more than 500 A^2 of surface area.[94] Structures which have a greater number of monomers in their asymmetric unit can show the same interface multiple times, and this is counteracted by only labeling each interface once.

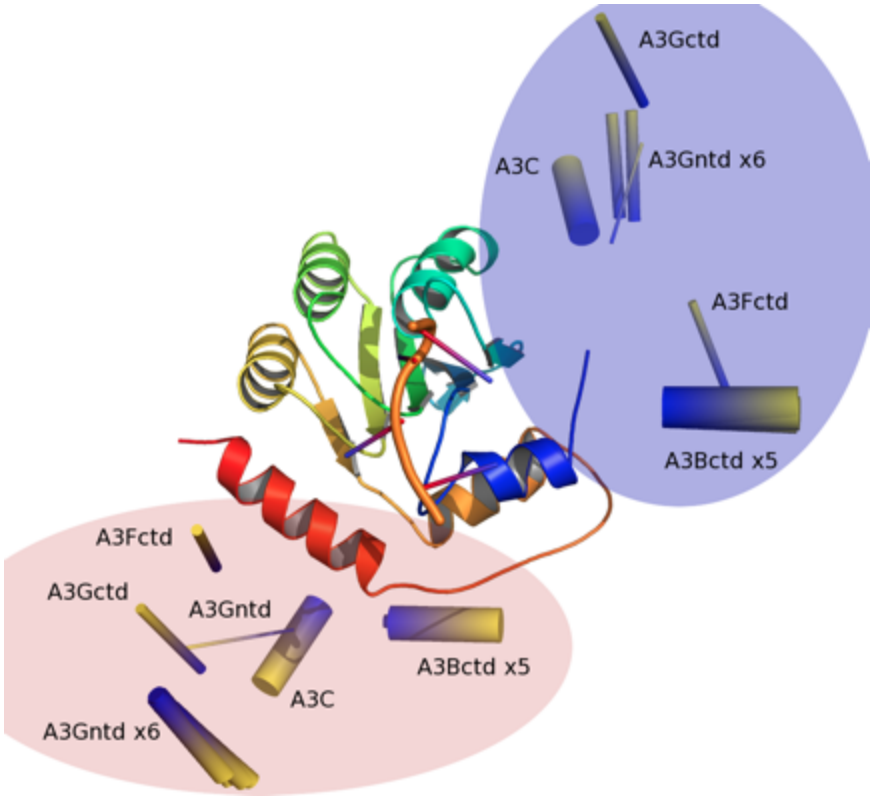

Figure S3: *Analysis of crystal structures to discover A3 domain oligomerization interface. A3 crystal structures were analyzed to find crystal interfaces with at least 500 A^2 surface area. Each*

interface is shown according to the position of the other domain as a cylinder, with the yellow end at the center of mass of the domain and the blue end indicating the position of the catalytic Zn. The analysis shows clusters of interfaces at the N and C terminals of the reference domain (shown as blue and red cartoon/highlighting, respectively).

Comparative structural biology of APOBEC enzymes

Python package available at https://github.com/j-wags/DepictStructBioinf

Due to the high degree of homology between APOBEC3 enzymes, we expect that structural and functional insights from one family member may have implications for others. To this end, we have developed a Python package to explore structural and biochemical data within the APOBEC3 family. This package contains all APOBEC3 domains, pre-aligned to each other. It also enables annotation of structures, which can be visualized in Pymol as an overlay on crystal structures or homology models. This package not only shows annotations on the structure that they belong to, but can also identify the homologous residues that the annotations apply to on other APOBEC3 domains Figure S4 shows an example of two such aligned APOBEC3 domains and with an annotation of Asp314 in A3B and the homologous residue in A3G.

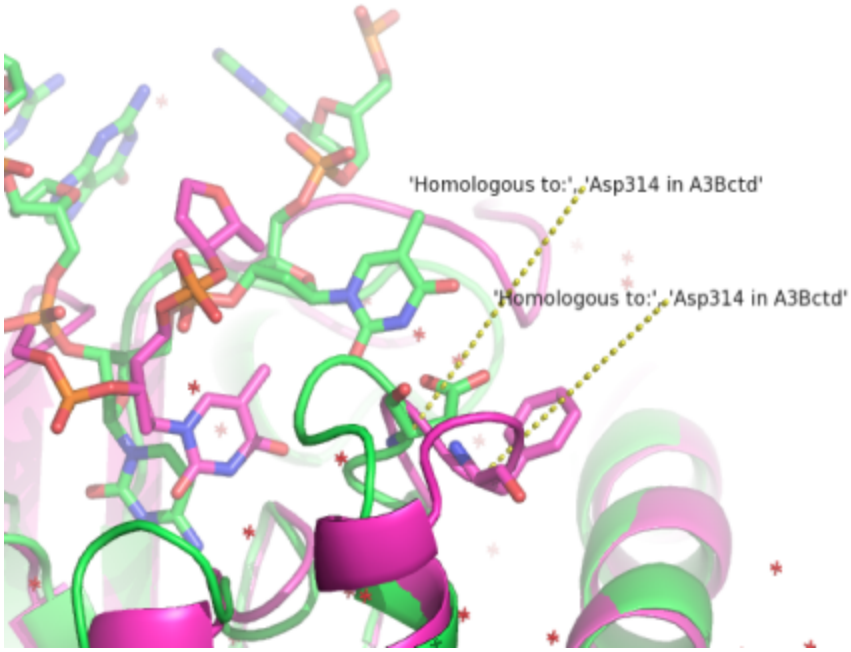

Figure S4: *Structural and functional annotation of APOBEC3B (green) and APOBEC3G (purple), with the important A3B substrate recognition residue Asp314 and its homologue in A3G annotated.*

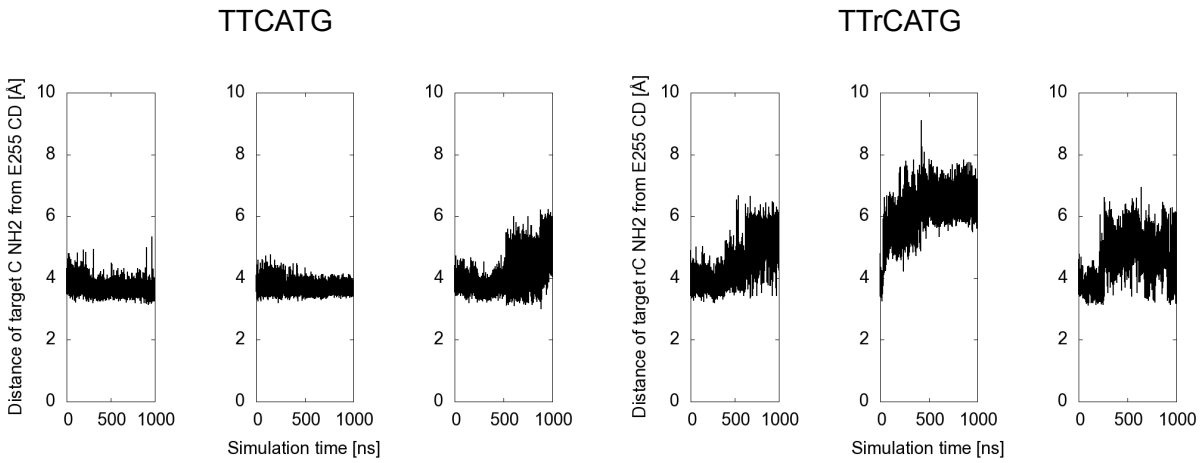

Figure S5: *Distance between catalytic Glu255 and target cytosine. In the second rC simulation, the target cytosine shifts away from Glu255 in the binding pocket.*

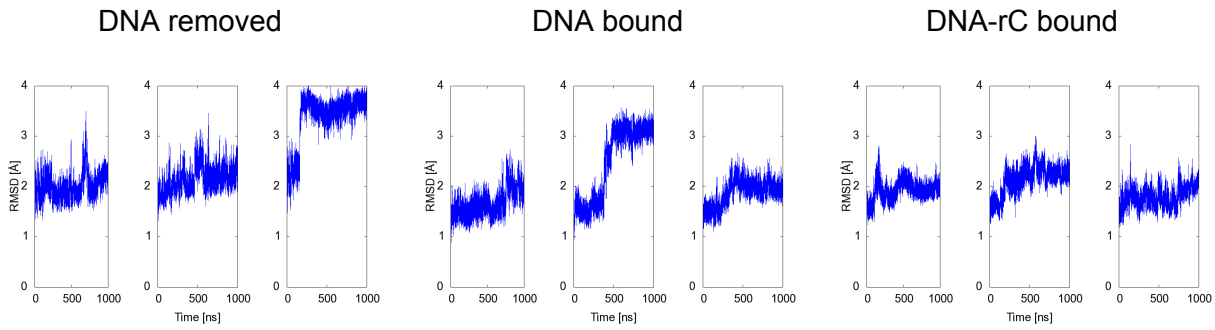

Figure S6: *C-alpha RMSD of A3B simulations, measured from the initial configuration of each simulation.*

```
$AMBERHOME/bin/pmemd.cuda -O -i $DIR/min1.in -o $DIR/min1.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/a3b_xtal_dna_new_rC.crd -ref
$DIR/a3b_xtal_dna_new_rC.crd -r $DIR/min1.rst -x $DIR/min1.nc

$AMBERHOME/bin/pmemd.cuda -O -i $DIR/min2.in -o $DIR/min2.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/min1.rst -ref $DIR/min1.rst -r $DIR/min2.rst -x
$DIR/min2.nc

$AMBERHOME/bin/pmemd.cuda -O -i $DIR/min3.in -o $DIR/min3.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/min2.rst -ref $DIR/min2.rst -r $DIR/min3.rst -x
$DIR/min3.nc

$AMBERHOME/bin/pmemd.cuda -O -i $DIR/min4.in -o $DIR/min4.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/min3.rst -ref $DIR/min3.rst -r $DIR/min4.rst -x
$DIR/min4.nc

mpirun -np 4 $AMBERHOME/bin/pmemd.cuda.MPI -O -i $DIR/md1.in -o $DIR/md1.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/min4.rst -ref $DIR/min4.rst -r $DIR/md1.rst -x
$DIR/md1.nc

mpirun -np 4 $AMBERHOME/bin/pmemd.cuda.MPI -O -i $DIR/md2.in -o $DIR/md2.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/md1.rst -ref $DIR/md1.rst -r $DIR/md2.rst -x
$DIR/md2.nc

mpirun -np 4 $AMBERHOME/bin/pmemd.cuda.MPI -O -i $DIR/md3.in -o $DIR/md3.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/md2.rst -ref $DIR/md2.rst -r $DIR/md3.rst -x
$DIR/md3.nc

mpirun -np 4 $AMBERHOME/bin/pmemd.cuda.MPI -O -i $DIR/md4.in -o $DIR/md4.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/md3.rst -ref $DIR/md3.rst -r $DIR/md4.rst -x
$DIR/md4.nc

mpirun -np 4 $AMBERHOME/bin/pmemd.cuda.MPI -O -i $DIR/md5.in -o $DIR/md5.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/md4.rst -ref $DIR/md4.rst -r $DIR/md5.rst -x
$DIR/md5.nc

mpirun -np 4 $AMBERHOME/bin/pmemd.cuda.MPI -O -i $DIR/md6.in -o $DIR/md6.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/md5.rst -ref $DIR/md5.rst -r $DIR/md6.rst -x
$DIR/md6.nc

mpirun -np 4 $AMBERHOME/bin/pmemd.cuda.MPI -O -i $DIR/md7.in -o $DIR/md7.out -p
$DIR/a3b_xtal_dna_new_rC.top -c $DIR/md6.rst -ref $DIR/md6.rst -r $DIR/md7.rst -x
$DIR/md7.nc
```

min1.in

```
-------------
Minimization run
cat <mdin
   minimize structure
   &cntrl
      imin=1,maxcyc=500,
      ncyc=500,ntb=1,cut=10,
      ntr=1, restraintmask="!@H=",restraint_wt=100.0
/
END
```

min2.in

```
-------------
Minimization run
cat <mdin
   minimize structure
   &cntrl
      imin=1,maxcyc=500,
      ncyc=500,ntb=1,cut=10,
      ntr=1, restraintmask=":1-200 & !@H=",restraint_wt=50.0
/
END
```

min3.in

```
-------------
Minimization run
cat <mdin
   minimize structure
   &cntrl
      imin=1,maxcyc=500,
      ncyc=500,ntb=1,cut=10,
      ntr=1, restraintmask=":1-192@CA,N,C,O",restraint_wt=10.0
/
END
```

min4.in

```
-------------
```

```
Minimization run
cat <mdin
   minimize structure
   &cntrl
      imin=1,maxcyc=40000,
      ncyc=40000,ntb=1,cut=10,
/
END
```

```
md1.in
-------------
250 ps MD with res on protein
 &cntrl
  imin  = 0,
  irest = 0,
  ntx   = 1,
  ntb   = 1,
  cut   = 10,
  ntr   = 1, restraintmask=":1-63,65-94,96-99,101-192 & !@H=", restraint_wt=2.0,
  ntc   = 2,
  ntf   = 2,
  tempi = 0.0,
  temp0 = 310.0,
  ntt   = 3,
  gamma_ln = 5.0,
  nstlim = 125000, dt = 0.002,
  ntpr = 5000, ntwx = 5000, ntwr = 5000,
  iwrap=1,
 /

md2.in
-------------
250 ps MD
 &cntrl
  imin=0, irest=1,ntx=7,
  ntb=2, pres0=1.0, ntp=1,
  taup=2.0,
  cut=10, ntr=0,
  ntc=2, ntf=2,
  tempi = 310.0, temp0 = 310.0,
  ntt = 3, gamma_ln = 5.0,
  ntr   = 1, restraintmask=":1-192 & !:97,102,66  & !@H=", restraint_wt=3.0
  nstlim = 125000, dt = 0.002,
  ntpr = 5000, ntwx = 5000, ntwr = 5000,
  iwrap=1
 /
```

md3.in
----
250 ps MD
 &cntrl
  imin=0, irest=1,ntx=7,
  ntb=2, pres0=1.0, ntp=1,
  taup=2.0,
  cut=10, ntr=0,
  ntc=2, ntf=2,
  tempi = 310.0, temp0 = 310.0,
  ntt = 3, gamma_ln = 5.0,
  ntr   = 1, restraintmask=":1-192 & !:97,102,66 & !@H=", restraint_wt=2.0
  nstlim = 125000, dt = 0.002,
  ntpr = 5000, ntwx = 5000, ntwr = 5000,
  iwrap=1
 /

md4.in
----
250 ps MD
 &cntrl
  imin=0, irest=1,ntx=7,
  ntb=2, pres0=1.0, ntp=1,
  taup=2.0,
  cut=10, ntr=0,
  ntc=2, ntf=2,
  tempi = 310.0, temp0 = 310.0,
  ntt = 3, gamma_ln = 5.0,
  ntr   = 1, restraintmask=":1-192 & !:97,102,66 & !@H=", restraint_wt=1.0
  nstlim = 125000, dt = 0.002,
  ntpr = 5000, ntwx = 5000, ntwr = 5000,
  iwrap=1
 /

md5.in
----
1 us MD
 &cntrl
  imin=0, irest=0,ntx=1,
  ntb=2, pres0=1.0, ntp=1,
  taup=2.0,
  cut=9, ntr=0,
  ntc=2, ntf=2,
  tempi = 310.0, temp0 = 310.0,
  ntt = 3, gamma_ln = 5.0,
  nstlim = 500000000, dt = 0.002,
  ntpr = 25000, ntwx = 25000, ntwr = 25000,
  iwrap=1,ioutfm=1

/

Scheme S1: Example commands and AMBER input files for MD simulation

(1) Reuben S. Harris, J. P. D. (2015) APOBECs and Virus Restriction. *Virology 0*, 131.
(2) Olson, M. E., Harris, R. S., and Harki, D. A. (2018) APOBEC Enzymes as Targets for Virus and Cancer Therapy. *Cell Chem Biol 25*, 36–49.
(3) Li, M., Shandilya, S. M. D., 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., and Harris, R. S. (2012) First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *ACS Chem. Biol. 7*, 506–517.
(4) Olson, M. E., Li, M., Harris, R. S., and Harki, D. A. (2013) Small-molecule APOBEC3G DNA cytosine deaminase inhibitors based on a 4-amino-1,2,4-triazole-3-thiol scaffold. *ChemMedChem 8*, 112–117.
(5) Middlebrooks, C. D., Banday, A. R., Matsuda, K., Udquim, K.-I., Onabajo, O. O., Paquin, A., Figueroa, J. D., Zhu, B., Koutros, S., Kubo, M., Shuin, T., Freedman, N. D., Kogevinas, M., Malats, N., Chanock, S. J., Garcia-Closas, M., Silverman, D. T., Rothman, N., and Prokunina-Olsson, L. (2016) Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *Nat. Genet. 48*, 1330–1338.
(6) Henderson, S., and Fenton, T. (2015) APOBEC3 genes: retroviral restriction factors to cancer drivers. *Trends Mol. Med. 21*, 274–284.
(7) Swanton, C., McGranahan, N., Starrett, G. J., and Harris, R. S. (2015) APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov. 5*, 704–712.
(8) Periyasamy, M., Patel, H., Lai, C.-F., Nguyen, V. T. M., Nevedomskaya, E., Harrod, A., Russell, R., Remenyi, J., Ochocka, A. M., Thomas, R. S., Fuller-Pace, F., Győrffy, B., Caldas, C., Navaratnam, N., Carroll, J. S., Zwart, W., Coombes, R. C., Magnani, L., Buluwela, L., and Ali, S. (2015) APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer. *Cell Rep. 13*, 108–121.
(9) Knisbacher, B. A., Gerber, D., and Levanon, E. Y. (2016) DNA Editing by APOBECs: A Genomic Preserver and Transformer. *Trends Genet. 32*, 16–28.
(10) Henderson, S., and Fenton, T. (2015) APOBEC3 genes: retroviral restriction factors to cancer drivers. *Trends Mol. Med. 21*, 274–284.
(11) Rebhandl, S., Huemer, M., Gassner, F. J., Zaborsky, N., Hebenstreit, D., Catakovic, K., Grössinger, E. M., Greil, R., and Geisberger, R. (2014) APOBEC3 signature mutations in chronic lymphocytic leukemia. *Leukemia 28*, 1929–1932.
(12) Helleday, T., Eshtad, S., and Nik-Zainal, S. (2014) Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet. 15*, 585–598.
(13) 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., and Gordenin, D. A. (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet. 45*, 970–976.
(14) Kazanov, M. D., Roberts, S. A., Polak, P., Stamatoyannopoulos, J., Klimczak, L. J., Gordenin, D. A., and Sunyaev, S. R. (2015) APOBEC-Induced Cancer Mutations Are Uniquely Enriched in Early-Replicating, Gene-Dense, and Active Chromatin Regions. *Cell Rep. 13*, 1103–1109.
(15) Zou, J., Wang, C., Ma, X., Wang, E., and Peng, G. (2017) APOBEC3B, a molecular driver of mutagenesis in human cancers. *Cell Biosci. 7*, 29.
(16) Nowarski, R., and Kotler, M. (2013) APOBEC3 cytidine deaminases in double-strand DNA

break repair and cancer promotion. *Cancer Res. 73*, 3494–3498.
(17) 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., and Harris, R. S. (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature 494*, 366–370.
(18) Hoopes, J. I., Cortez, L. M., Mertz, T. M., Malc, E. P., Mieczkowski, P. A., and Roberts, S. A. (2016) APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. *Cell Rep. 14*, 1273–1282.
(19) Siriwardena, S. U., Guruge, T. A., and Bhagwat, A. S. (2015) Characterization of the Catalytic Domain of Human APOBEC3B and the Critical Structural Role for a Conserved Methionine. *J. Mol. Biol. 427*, 3042–3055.
(20) Fu, Y., Ito, F., Zhang, G., Fernandez, B., Yang, H., and Chen, X. S. (2015) DNA cytosine and methylcytosine deamination by APOBEC3B: enhancing methylcytosine deamination by engineering APOBEC3B. *Biochem. J 471*, 25–35.
(21) Harjes, E., Gross, P. J., Chen, K.-M., Lu, Y., Shindo, K., Nowarski, R., Gross, J. D., Kotler, M., Harris, R. S., and Matsuo, H. (2009) An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J. Mol. Biol. 389*, 819–832.
(22) Aydin, H., Taylor, M. W., and Lee, J. E. (2014) Structure-guided analysis of the human APOBEC3-HIV restrictome. *Structure 22*, 668–684.
(23) Shandilya, S. M. D., Bohn, M.-F., and Schiffer, C. A. (2014) A computational analysis of the structural determinants of APOBEC3's catalytic activity and vulnerability to HIV-1 Vif. *Virology 471-473*, 105–116.
(24) Shi, K., Carpenter, M. A., Kurahashi, K., Harris, R. S., and Aihara, H. (2015) Crystal Structure of the DNA Deaminase APOBEC3B Catalytic Domain. *J. Biol. Chem. 290*, 28120–28130.
(25) Shi, K., Demir, Ö., Carpenter, M. A., Wagner, J., Kurahashi, K., Harris, R. S., Amaro, R. E., and Aihara, H. (2017) Conformational Switch Regulates the DNA Cytosine Deaminase Activity of Human APOBEC3B. *Sci. Rep. 7*, 17415.
(26) 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, Ö., Amaro, R. E., Harki, D. A., Harris, R. S., and Aihara, H. (2016) Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol. 24*, 131–139.
(27) Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res. 29*, 2860–2874.
(28) Siggers, T., and Gordân, R. (2014) Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Res. 42*, 2099–2111.
(29) Harteis, S., and Schneider, S. (2014) Making the bend: DNA tertiary structure and protein-DNA interactions. *Int. J. Mol. Sci. 15*, 12335–12363.
(30) Nadassy, K., Wodak, S. J., and Janin, J. (1999) Structural features of protein-nucleic acid recognition sites. *Biochemistry 38*, 1999–2017.
(31) Kohli, R. M., Abrams, S. R., Gajula, K. S., Maul, R. W., Gearhart, P. J., and Stivers, J. T. (2009) A Portable Hot Spot Recognition Loop Transfers Sequence Preferences from APOBEC Family Members to Activation-induced Cytidine Deaminase. *J. Biol. Chem. 284*, 22898.
(32) Rathore, A., Carpenter, M. A., Demir, Ö., Ikeda, T., Li, M., Shaban, N. M., Law, E. K., Anokhin, D., Brown, W. L., Amaro, R. E., and Harris, R. S. (2013) The Local Dinucleotide Preference of APOBEC3G Can Be Altered from 5′-CC to 5′-TC by a Single Amino Acid Substitution. *J. Mol. Biol. 425*, 4442–4454.

(33) Kohli, R. M., Maul, R. W., Guminski, A. F., McClure, R. L., Gajula, K. S., Saribasak, H., McMahon, M. A., Siliciano, R. F., Gearhart, P. J., and Stivers, J. T. (2010) Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *J. Biol. Chem. 285*, 40956–40964.
(34) Carpenter, M. A., Rajagurubandara, E., Wijesinghe, P., and Bhagwat, A. S. (2010) Determinants of sequence-specificity within human AID and APOBEC3G. *DNA Repair 9*, 579–587.
(35) Wang, M., Rada, C., and Neuberger, M. S. (2010) Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *J. Exp. Med. 207*, 141–153.
(36) Xiao, X., Li, S.-X., Yang, H., and Chen, X. S. (2016) Crystal structures of APOBEC3G N-domain alone and its complex with DNA. *Nat. Commun. 7*, 12193.
(37) Shi, K., Banerjee, S., Kurahashi, K., and Aihara, H. (2016) Crystal Structure of Human APOBEC3A complexed with ssDNA.
(38) Kouno, T., Hilbert, B. J., Silvas, T., Royer, W. E., Matsuo, H., and Schiffer, C. A. (2017) Crystal structure of APOBEC3A in complex with a single-stranded DNA.
(39) Nabel, C. S., Lee, J. W., Wang, L. C., and Kohli, R. M. (2013) Nucleic acid determinants for selective deamination of DNA over RNA by activation-induced deaminase. *Proc. Natl. Acad. Sci. U. S. A. 110*, 14225.
(40) Carpenter, M. A., Li, M., Rathore, A., Lackey, L., Law, E. K., Land, A. M., Leonard, B., Shandilya, S. M. D., Bohn, M.-F., Schiffer, C. A., Brown, W. L., and Harris, R. S. (2012) Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *J. Biol. Chem. 287*, 34801–34808.
(41) Wijesinghe, P., and Bhagwat, A. S. (2012) Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res. 40*, 9206–9217.
(42) Suspène, R., Aynaud, M.-M., Vartanian, J.-P., and Wain-Hobson, S. (2013) Efficient Deamination of 5-Methylcytidine and 5-Substituted Cytidine Residues in DNA by Human APOBEC3A Cytidine Deaminase. *PLoS One 8*, e63461.
(43) Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T., Zhang, Y., and Kohli, R. M. (2012) AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol. 8*, 751–758.
(44) Salter, J. D., Bennett, R. P., and Smith, H. C. (2016) The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem. Sci. 41*, 578.
(45) Paquet, E., and Viktor, H. L. (2015) Molecular dynamics, monte carlo simulations, and langevin dynamics: a computational review. *Biomed Res. Int. 2015*, 183918.
(46) Vangaveti, S., Ranganathan, S. V., and Chen, A. A. (2017) Advances in RNA molecular dynamics: a simulator's guide to RNA force fields. *Wiley Interdiscip. Rev. RNA 8*.
(47) Galindo-Murillo, R., Bergonzo, C., and Cheatham, T. E., 3rd. (2013) Molecular modeling of nucleic acid structure. *Curr. Protoc. Nucleic Acid Chem. 54*, Unit 7.5.
(48) Gong, Z., Xiao, Y., and Xiao, Y. (2010) RNA stability under different combinations of amber force fields and solvation models. *J. Biomol. Struct. Dyn. 28*, 431–441.
(49) Krepl, M., Havrila, M., Stadlbauer, P., Banas, P., Otyepka, M., Pasulka, J., Stefl, R., and Sponer, J. (2015) Can We Execute Stable Microsecond-Scale Atomistic Simulations of Protein-RNA Complexes? *J. Chem. Theory Comput. 11*, 1220–1243.
(50) Piana, S., Klepeis, J. L., and Shaw, D. E. (2014) Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol. 24*, 98–105.
(51) Vlachakis, D., Bencurova, E., Papangelopoulos, N., and Kossida, S. (2014) Current state-of-the-art molecular dynamics methods and applications. *Adv. Protein Chem. Struct. Biol.*

94, 269–313.
(52) Kalyaanamoorthy, S., and Chen, Y.-P. P. (2014) Modelling and enhanced molecular dynamics to steer structure-based drug discovery. *Prog. Biophys. Mol. Biol. 114*, 123–136.
(53) Spyrakis, F., and Cavasotto, C. N. (2015) Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. *Arch. Biochem. Biophys. 583*, 105–119.
(54) Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J. 92*, 3817–3829.
(55) Pérez, A., Luque, F. J., and Orozco, M. (2007) Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc. 129*, 14739–14745.
(56) Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput. 11*, 3696–3713.
(57) Case, D. A., Cerutti, D. S., Cheatham, T. E., III, Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Greene, D., Homeyer, N., Izadi, S., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo., R., Mermelstein, D., Merz, K. M., Monard, G., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Simmerling, C. L., Botello-Smith, W. M., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., Xiao, L., York, D. M., and Kollman, P. A. (2017) AMBER 16 and AMBERTools 17. *University of California, San Francisco*.
(58) Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2012) An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci. 3*, 198–210.
(59) Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem. 26*, 1668–1688.
(60) Banáš, P., Hollas, D., Zgarbová, M., Jurečka, P., Orozco, M., Cheatham, T. E., Šponer, J., and Otyepka, M. (2010) Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput. 6*, 3836–3849.
(61) Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., and Friesner, R. A. (2004) A hierarchical approach to all-atom protein loop prediction. *Proteins 55*, 351–367.
(62) Jacobson, M. P., Friesner, R. A., Xiang, Z., and Honig, B. (2002) On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J. Mol. Biol. 320*, 597–608.
(63) Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M., and Jensen, J. H. (2011) Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput. 7*, 2284–2295.
(64) Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput. 7*, 525–537.
(65) Pang, Y.-P. (1999) Novel Zinc Protein Molecular Dynamics Simulations: Steps Toward Antiangiogenesis for Cancer Treatment. *J. Mol. Model. 5*, 196–202.
(66) McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., Schwantes, C. R., Wang, L.-P., Lane, T. J., and Pande, V. S. (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J. 109*, 1528–1532.
(67) Hunter, J. D. (2007) Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng. 9*, 90–95.
(68) Baker, E. N., and Hubbard, R. E. (1984) Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol. 44*, 97–179.

(69) Altona, C., and Sundaralingam, M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc. 94*, 8205–8212.
(70) Harvey, S. C., and Prabhakaran, M. (1986) Ribose puckering: structure, dynamics, energetics, and the pseudorotation cycle. *J. Am. Chem. Soc. 108*, 6128–6136.
(71) Roe, D. R., and Cheatham, T. E., 3rd. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput. 9*, 3084–3095.
(72) Wagner, J. R., Sørensen, J., Hensley, N., Wong, C., Zhu, C., Perison, T., and Amaro, R. E. (2017) POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J. Chem. Theory Comput. 13*, 4584–4592.
(73) Durrant, J. D., de Oliveira, C. A. F., and McCammon, J. A. (2011) POVME: an algorithm for measuring binding-pocket volumes. *J. Mol. Graph. Model. 29*, 773–776.
(74) Durrant, J. D., Votapka, L., Sørensen, J., and Amaro, R. E. (2014) POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput. 10*, 5047–5056.
(75) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph. 14*, 33–38.
(76) McDougle, R. M., Hultquist, J. F., Stabell, A. C., Sawyer, S. L., and Harris, R. S. (2013) D316 is critical for the enzymatic activity and HIV-1 restriction potential of human and rhesus APOBEC3B. *Virology 441*, 31–39.
(77) Byeon, I.-J. L., Byeon, C.-H., Wu, T., Mitra, M., Singer, D., Levin, J. G., and Gronenborn, A. M. (2016) Nuclear Magnetic Resonance Structure of the APOBEC3B Catalytic Domain: Structural Basis for Substrate Binding and DNA Deaminase Activity. *Biochemistry 55*, 2944–2959.
(78) Harjes, S., Solomon, W. C., Li, M., Chen, K.-M., Harjes, E., Harris, R. S., and Matsuo, H. (2013) Impact of H216 on the DNA binding and catalytic activities of the HIV restriction factor APOBEC3G. *J. Virol. 87*, 7008–7014.
(79) Chen, K.-M., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y., Shindo, K., Harris, R. S., and Matsuo, H. (2008) Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature 452*, 116–119.
(80) The PyMOL Molecular Graphics System. *Schrödinger, LLC.*
(81) Shi, K., Kurahashi, K., and Aihara, H. (2015) Crystal Structure of the Cancer Genomic DNA Mutator APOBEC3B.
(82) Byeon, I.-J. L., Ahn, J., Mitra, M., Byeon, C.-H., Hercík, K., Hritz, J., Charlton, L. M., Levin, J. G., and Gronenborn, A. M. (2013) NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nat. Commun. 4*, 1890.
(83) Kitamura, S., Ode, H., Nakashima, M., Imahashi, M., Naganawa, Y., Kurosawa, T., Yokomaku, Y., Yamane, T., Watanabe, N., Suzuki, A., Sugiura, W., and Iwatani, Y. (2012) The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat. Struct. Mol. Biol. 19*, 1005–1010.
(84) Nakashima, M., Ode, H., Kawamura, T., Kitamura, S., Naganawa, Y., Awazu, H., Tsuzuki, S., Matsuoka, K., Nemoto, M., Hachiya, A., Sugiura, W., Yokomaku, Y., Watanabe, N., and Iwatani, Y. (2015) Structural Insights into HIV-1 Vif-APOBEC3F Interaction. *J. Virol. 90*, 1034–1047.
(85) Siu, K. K., Sultana, A., Azimi, F. C., and Lee, J. E. (2013) Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. *Nat. Commun. 4*, 2593.
(86) Bohn, M.-F., Shandilya, S. M. D., 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., and Schiffer, C. A. (2013) Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. *Structure 21*, 1042–1050.
(87) Shaban, N. M., Shi, K., Li, M., Aihara, H., and Harris, R. S. (2016) 1.92 Angstrom Zinc-Free APOBEC3F Catalytic Domain Crystal Structure. *J. Mol. Biol. 428*, 2307–2316.
(88) Kouno, T., Luengas, E. M., Shigematsu, M., Shandilya, S. M. D., Zhang, J., Chen, L., Hara, M., Schiffer, C. A., Harris, R. S., and Matsuo, H. (2015) Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G. *Nat. Struct. Mol. Biol. 22*, 485–491.
(89) Holden, L. G., Prochnow, C., Chang, Y. P., Bransteitter, R., Chelico, L., Sen, U., Stevens, R. C., Goodman, M. F., and Chen, X. S. (2008) Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature 456*, 121–124.
(90) Furukawa, A., Nagata, T., Matsugami, A., Habu, Y., Sugiyama, R., Hayashi, F., Kobayashi, N., Yokoyama, S., Takaku, H., and Katahira, M. (2009) Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J. 28*, 440–451.
(91) Shandilya, S. M. D., Nalam, M. N. L., 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., and Schiffer, C. A. (2010) Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure 18*, 28–38.
(92) Lu, X., Zhang, T., Xu, Z., Liu, S., Zhao, B., Lan, W., Wang, C., Ding, J., and Cao, C. (2015) 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. 290*, 4010–4021.
(93) Duarte, J. M., Srebniak, A., Schärer, M. A., and Capitani, G. (2012) Protein interface classification by evolutionary analysis. *BMC Bioinformatics 13*, 334.
(94) Chen, J., Sawyer, N., and Regan, L. (2013) Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci. 22*, 510.