Regulatory interactions between APOBEC3B N- and C-terminal domains

Mac Kevin E. Braza,[1] Özlem Demir,[1,+] Surl-Hee Ahn,[2] Clare K. Morris,[1] Carla Calvó-Tusell,[3] Kelly L. McGuire,[1] Bárbara de la Peña Avalos,[4] Michael A. Carpenter,[4,5] Yanjun Chen,[4] Lorenzo Casalino,[3] Hideki Aihara,[6] Mark A. Herzik, Jr.,[1] Reuben S. Harris,[4,5] and Rommie E. Amaro[3*]

[1]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA
[2]Department of Chemical Engineering, University of California, Davis, Davis, CA
[3]Department of Molecular Biology, University of California, San Diego, La Jolla, CA
[4]Department of Biochemistry and Structural Biology, University of Texas Health San Antonio, San Antonio, TX
[5]Howard Hughes Medical Institute, University of Texas Health San Antonio, San Antonio, TX
[6]Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN
[+]Current Address: Novartis Biomedical Research, San Diego, CA

Corresponding Author's E-mail: ramaro@ucsd.edu

KEYWORDS: protein-nucleic acid interactions, full-length APOBEC3B, weighted ensemble method simulations, Gaussian accelerated molecular dynamics simulations, active site opening, protein dynamics, molecular dynamics simulations

Abstract

APOBEC3B (A3B) is implicated in DNA mutations that facilitate tumor evolution. Although structures of its individual N- and C-terminal domains (NTD and CTD) have been resolved through X-ray crystallography, the full-length A3B (fl-A3B) structure remains elusive, limiting understanding of its dynamics and mechanisms. In particular, the APOBEC3B C-terminal domain (A3Bctd) active site is frequently closed in models and structures. In this study, we built several new models of fl-A3B using integrative structural biology methods and selected a top model for further dynamical investigation. We compared dynamics of the truncated (A3Bctd) to the fl-A3B via conventional and Gaussian accelerated molecular dynamics (MD) simulations. Subsequently, we employed weighted ensemble methods to explore the fl-A3B active site opening mechanism, finding that interactions at the NTD-CTD interface enhance the opening frequency of the fl-A3B active site. Our findings shed light on the structural dynamics of fl-A3B, which may offer new avenues for therapeutic intervention in cancer.

Abbreviations
A3A: APOBEC3A
A3B: APOBEC3B
A3Bctd: APOBEC3B C-terminal domain
A3Bntd: APOBEC3B N-terminal domain
A3G: APOBEC3G
APOBEC3: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3
CTD: C-terminal domain
fl-A3B: Full-length APOBEC3B
GaMD: Gaussian accelerated molecular dynamics
MD: molecular dynamics
NTD: N-terminal domain
RADD: Real-time APOBEC-mediated DNA deamination
RFU: Relative fluorescence unit
UDG: uracil-DNA glycosylase
WE: Weighted ensemble method

Introduction

The Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3 (APOBEC3) enzymes are key for human immunity against retroviruses and transposons.[1,2] APOBEC3 enzymes catalyze the deamination of cytosines to uracil (C-to-U), targeting newly synthesized viral single-stranded DNA (ssDNA). The APOBEC3 family consists of seven cytosine deaminases namely APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H. The presence of an APOBEC3-attributed mutational signature (single-base-substitution C-to-T and C-to-G mutations) in the 5'-TC motif in a wide array of cancer types suggests that many of the cancers were associated with A3 activity. This mutagenesis implies their importance in the mutational milieu of human cancer.[3,4] APOBEC3B (A3B), in particular, is predominantly localized in the nucleus and has been linked to a wide range of cancers, including breast, head, neck, cervical, and bladder cancers, where its expression correlates with drug resistance and metastasis.[5–7] Furthermore, A3B expression is associated with poor clinical outcomes in estrogen-positive breast cancer patients.[4,8,9]

A3B is one of the double-domain-A3s that has a noncatalytic N-terminal domain (NTD) (sometimes called the "pseudocatalytic" domain) and a catalytic C-terminal domain. Similar to other A3s, A3B chelates one Zn^{2+} ion as a cofactor, bridging the active site residues to the DNA substrate. In recent years, X-ray crystal structures of A3Bctd mutants have been determined.[10–12] Despite progress in determining the X-ray crystal structures of various A3Bctd mutants, the structure of wild-type fl-A3B remains unresolved, largely due to is tendency to aggregate in solution and poor protein solubility.[13,14] Although A3B has been proposed as a promising drug target, there is limited knowledge of the active site accessibility and its opening mechanism due to the aforementioned experimental challenges.[10,15,16]

Previous studies have demonstrated that the NTD is involved in DNA and RNA binding and nuclear localization.[7,13,17–19] Notably, fl-A3B exhibits up to 40-fold greater activity than A3Bctd.[20] Despite the NTD's known regulatory role in fl-A3B, there is limited knowledge of its functional significance in the opening of active site and ssDNA deamination activity. Moreover, the absence of fl-A3B structural and dynamical studies limits our understanding of its specificity and mechanism. We hypothesize that the absence of the N-terminal domain (NTD), despite being non-catalytic, might hinder or affect the opening of the active site.

One of the most interesting dynamical properties of A3B is the active site opening. Experimental and computational studies have revealed how rarely the A3Bctd active site opens.[15,19,21] The A3Bctd active site favors a closed active site conformation and only opened after heavily mutating the active site residues. Furthermore, it remains elusive how the ssDNA binds to A3B active site especially in its wild-type form. We hypothesized that, (experimentally) the absence of NTD and (computationally) the limitation of classical MD simulations hindered the observation of A3B active site opening.

In this work, we used integrative structural biology methods to construct an all-atom model of the fl-A3B and performed extensive all-atom molecular dynamics (MD) simulations to elucidate the role of the NTD in the active site opening. By means of state-of-the-art enhanced sampling methods for MD simulations, such as Gaussian accelerated MD (GaMD) and weighted ensemble (WE) simulations, we characterize the opening mechanism of the ssDNA binding pocket. We find that opening of the fl-A3B active site depends on interactions between residues at the interface of the NTD and the CTD, highlighting the importance of the NTD for more than just DNA binding. This work provides critical insights into the conformational dynamics of fl-A3B, advancing our understanding of the molecular basis underlying ssDNA binding site opening.
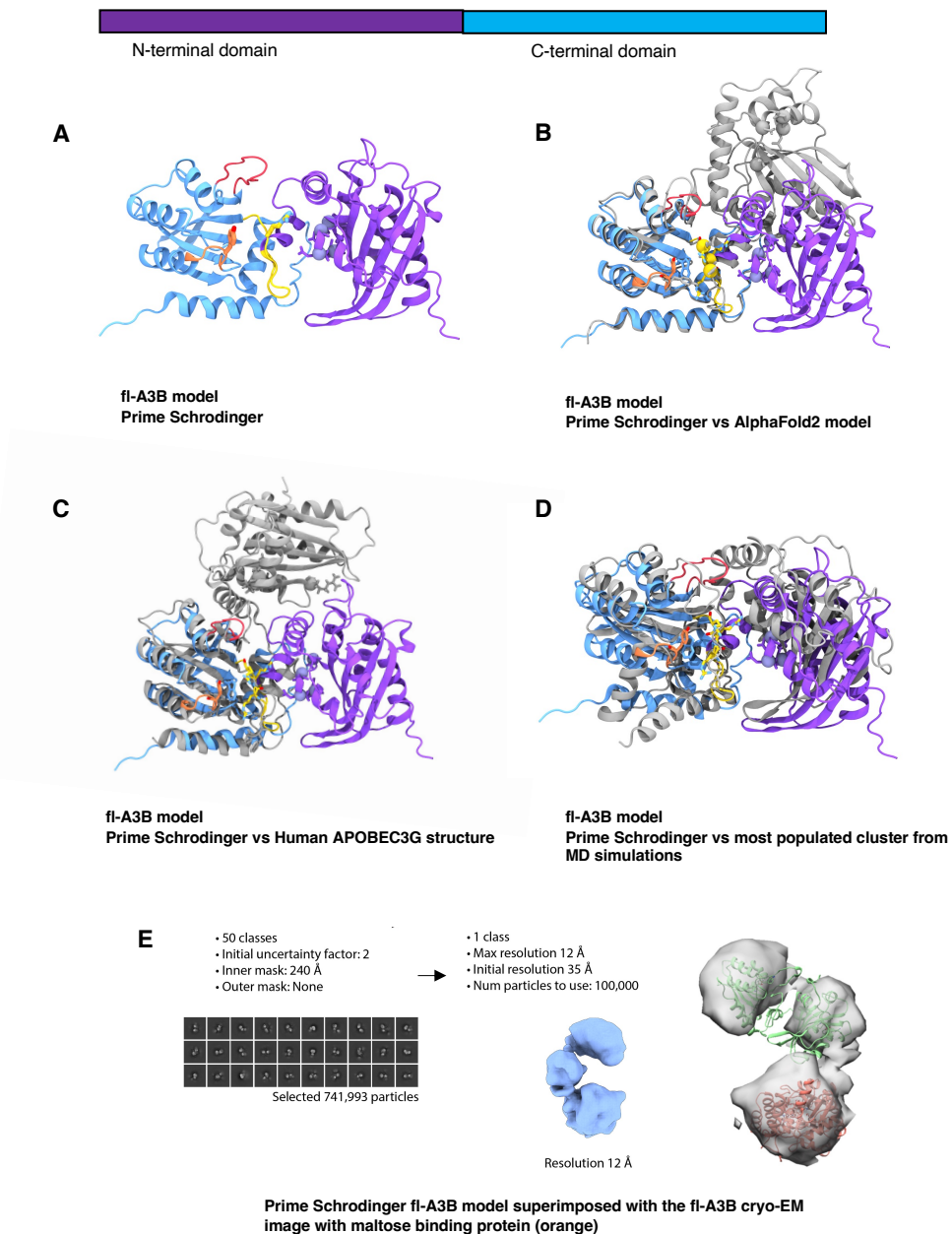
Figure 1. Comparative modeling of fl-A3B. (a) fl-A3B model from Prime - Schrodinger superimposed to (b) AlphaFold2 model; (c) Cryo-EM structure of wild-type human APOBEC3G (PDB ID 8CX0)[22], (d) most populated clustered structure from our MD simulations; and (e) Representative particle from the cryo-EM experiments of the fl-A3B and maltose binding protein complex.

Results and Discussion

In this study, we aimed to elucidate key structural features of full-length APOBEC3B (fl-A3B) that facilitate the opening of the active site. Despite fl-A3B's significant role in immunity and cancer mutagenesis, experimental structural studies are limited due to inherent solubility and aggregation issues.[13,14] Furthermore, the specific mechanisms by which the NTD influences nucleic acid recognition and overall protein dynamics remain poorly understood.[10,14,20,23] Due to extensive evidence that the NTD of other double domain APOBEC3 plays an important role in substrate recognition, we aimed to investigate if the A3B NTD served a similar functional purpose.

Constructing the full-length APOBEC3B model

In the absence of an experimentally-resolved fl-A3B structure, homology modeling tools were used to develop full-length models. Different tools including Schrödinger Prime,[24] I-TASSER,[25] SWISS-MODEL,[26] and AlphaFold 2[27] were used to explore the orientation of the fl-A3B NTD and CTD. The sequence from UniProt corresponding to the human fl-A3B was used and submitted to the software package or web server. As a basis of the NTD and CTD orientation to each other, we used ClusPro web server to generate protein-protein docked poses of two domains.[28] Surveying the results of the various modeling tools, differences in the predicted orientation of the NTD and CTD were apparent. AlphaFold models exhibited poor model confidence for the loop regions, most especially the active site loops 1, 3, and 7 (Figure 1). AlphaFold 2 and AlphaFold 3 exhibited low predicted local distance difference test (pLDDT) scores at critical active site loops, potentially impacting the accuracy of conformational dynamics. The models developed here were superimposed alongside the experimentally-characterized functionally-related double-domain deaminase, A3G.[22]

To attempt to characterize the full-length structure, we performed cryo-EM experiments of fl-A3B bound to maltose binding protein (Figure 1E, SI Figure 1). This enabled the characterization of the fl-A3B structure to maximal resolution of 12 Å. We subsequently aligned all predicted models (Figure 1) to the cryo-EM density. Ultimately, the fl-A3B model from Schrödinger Prime achieved the best (and substantial) alignment with low-resolution cryo-EM particles. We thus selected the Schrödinger Prime model for subsequent MD simulations.

## Full-length APOBEC3B active site opens more frequently than the APOBEC3B C-terminal domain alone

After modeling both the A3Bctd and fl-A3B, we ran a series of conventional MD simulations to characterize their baseline dynamics. Here, we used AMBER pmemd.cuda[29] with ff14SB[30] and TIP3P[31] force fields for protein and water, respectively. Our group and several others have shown that the A3Bctd's active site (WT and mutated) remain closed in the absence of ssDNA substrate.[15,18,19,21,32,33] Experimentally, it has been demonstrated that residues R211 and Y315 are crucial for ssDNA substrate interaction within A3B. R211, located in loop 1, interacts with the -1 thymine nucleotide of ssDNA, while Y315 in loop 7 plays a critical role in the active site opening.[10] Notably, the distances between the N atom of R211 and the OH group of Y315 differ significantly between the open (ssDNA-bound) and closed (apo form) states, with the closed A3Bctd active site characterized by the distance between R211-Y315 as being less than 10 Å (Figure 2A). Additionally, we characterized the conformation of Y315 by its Chi1 dihedral angle, defining a closed conformation at 60° and an open conformation at 180° (Figure 2A).

The fl-A3B model structure used as starting point for MD simulations has a partially open active site, with the R211-Y315 interaction exhibiting a distance >10 Å and the Y315 Chi1 dihedral angle in trans conformation (180 degrees). This is because the template used for model construction was the ssDNA-bound A3Bctd discussed above. The fl-A3B was prepared for simulation (see Methods), with all missing loops restored, mutated residues reverted back to wild type, and explicit solvent and counter ions added. We subsequently performed 4 replicates of 1 µs conventional MD (4 µs aggregate sampling) for the fl-A3B model.

Overall the fl-A3B model exhibited stable dynamics over all 4 µs (SI Figure 3). In contrast with the A3Bctd, the fl-A3B active site opens more frequently (Figure 3), as exhibited by the shift to open-state for the R211-Y315 distance. Notably, in the A3Bctd alone, the R211-Y315 distance peaked near 10 Å, whereas in the fl-A3B, the R211-Y315 distance metric samples a significantly more open state, with distances ranging from 15 - 22 Å. The Y315 Chi1 dihedral angle in fl-A3B sampled both closed and open states, although it predominately sampled angles corresponding to the closed state.
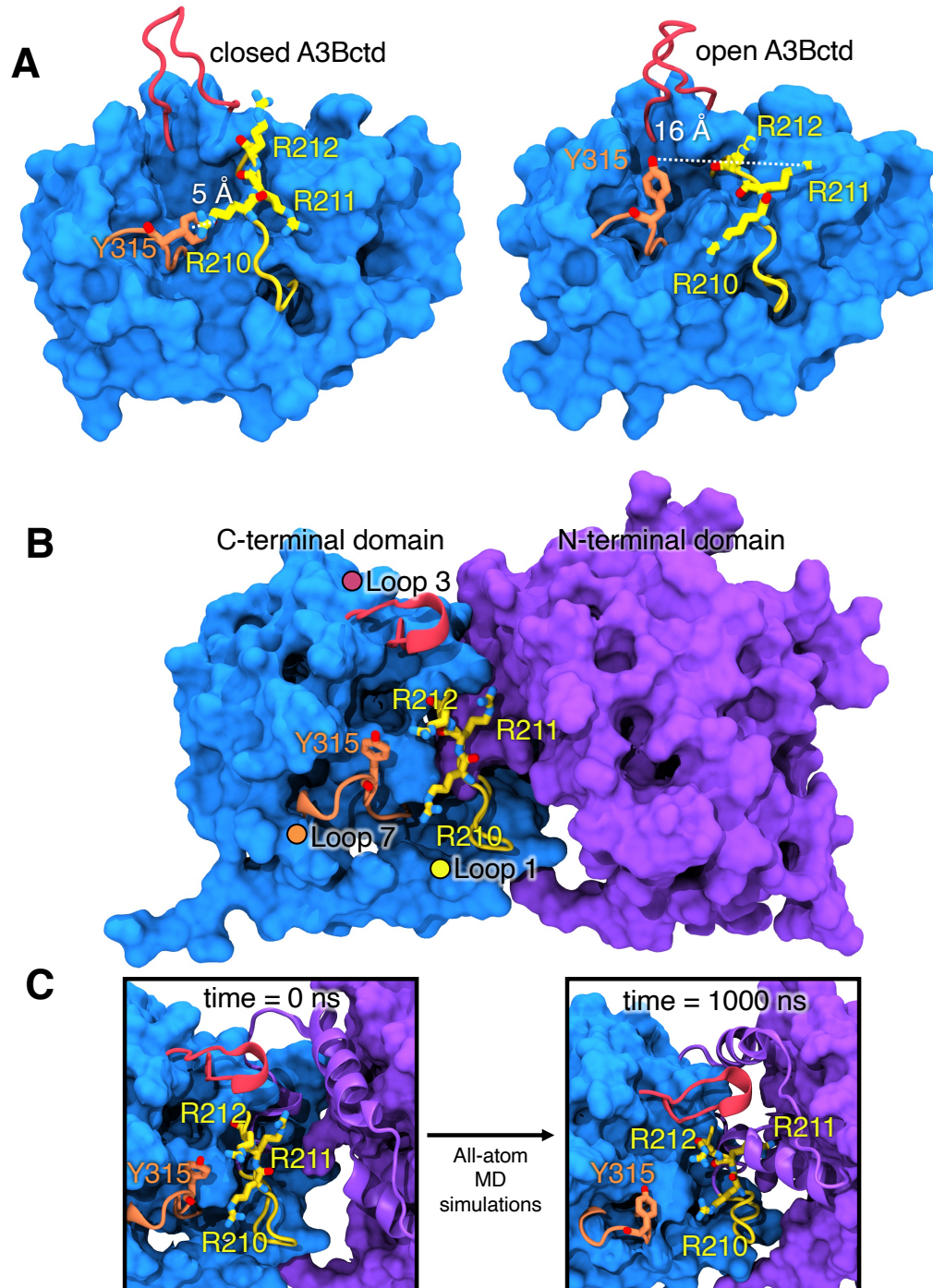
Figure 2. A3Bctd X-ray structures and fl-A3B model. (A) Representative structures of open and closed-state A3Bctd from available PDB X-ray structures (with ssDNA bound, not shown, from PDB ID 5TD5).[3] The closed A3Bctd active site is defined with a 5 Å distance between R211-Y315. An open A3Bctd active site is defined with 16 Å distance between R211-Y315. (B) fl-A3B model. (C) Conformational dynamics of fl-A3B ssDNA active site, in loop 1, residues R210, R211, and R212 interacting with N-terminal domain.
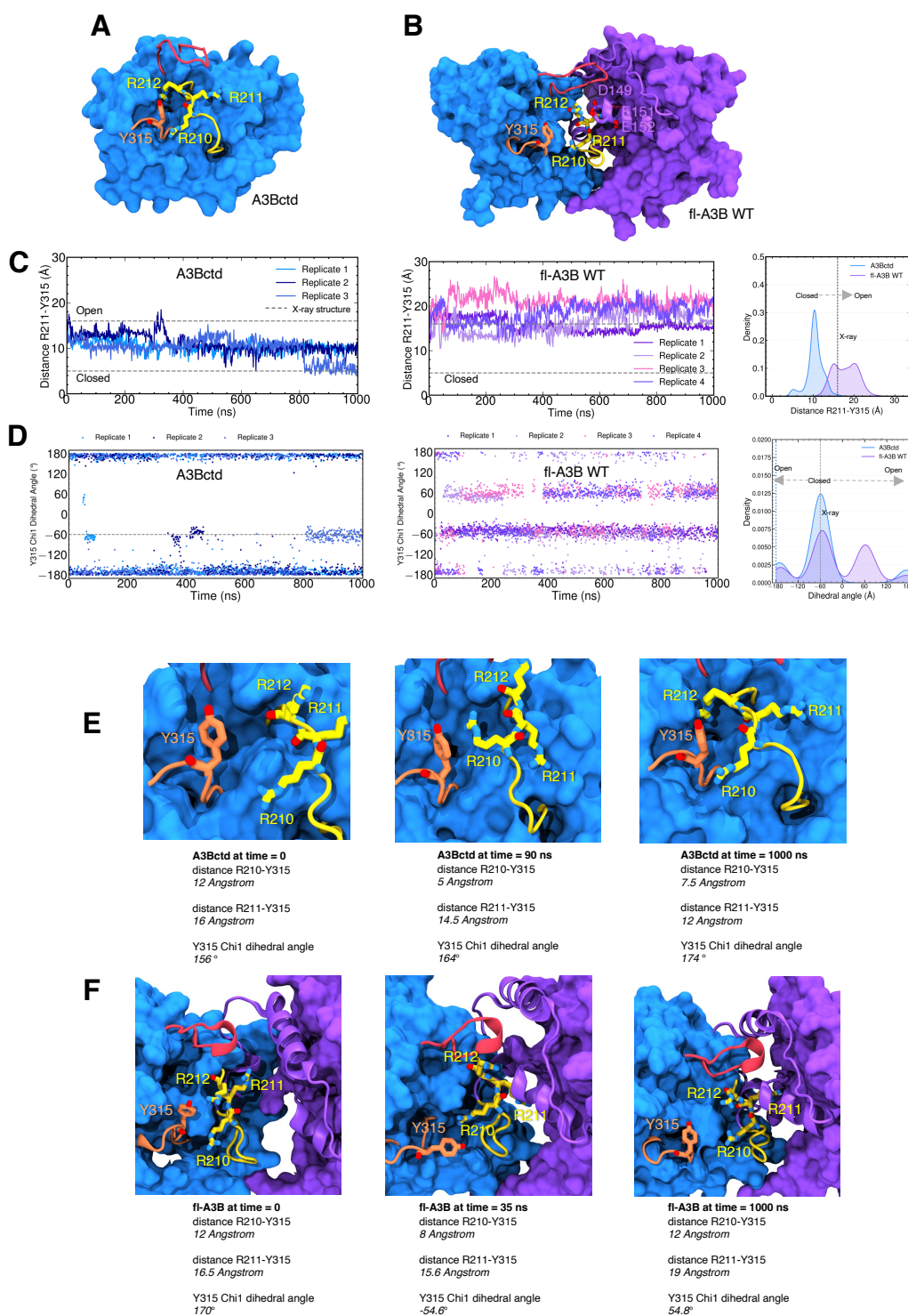
Figure 3. Structures and conformational dynamics of A3Bctd and fl-A3B. (A) A3Bctd; (B) fl-A3B; (C) fl-A3B and A3Bctd time-series distance of R211-Y315 with corresponding kernel density estimate (KDE) from several replicates of our simulations; (D) Time-series Chi1 dihedral angle; (E) A3Bctd representative snapshots highlighting the interactions of loop 1 residues R210, R211, and R212 with loop 7 residue Y315. Trajectories from our previous work were reanalyzed.[15,21] (F) fl-A3B representative snapshots highlighting the interaction of R210, R211, and R212 with NTD residues.

To increase the likelihood of sampling the closed-to-open transition, we employed Gaussian accelerated molecular dynamics (GaMD) simulations in AMBER pmemd.cuda. GaMD, an enhanced sampling technique, uses a harmonic boost potential that smoothens energy barriers on the potential energy surface.[34] During GaMD, the fl-A3B Y315 Chi1 dihedral angle values remained in the closed state (-60 degrees). The R211-Y315 distance is stable in its closed state at 9 Å (SI Figure 5). Thus, GaMD did not help in the opening transition of the fl-A3B active site.

To more rigorously characterize the active site opening mechanism, we subsequently carried out weighted ensemble (WE) simulations. WE is a transition pathway method that increases the chance of sampling biologically 'rare' events (e.g., in this case, sampling the closed to open active site transition).[35,36] Application of this method requires selecting a so-called 'progress coordinate', i.e., a set of principle components that captures the critical dynamics of the system, to accurately describe the structural dynamics of interest in APOBEC3B. The optimal progress coordinate should reflect the slowest motions pertinent to the rare-event processes under investigation.[37] Here, we used the R211-Y315 distance and the Y315 Chi1 dihedral angle as a 2-dimensional (2D) progress coordinate, where the closed state is defined by a R211-Y315 distance < 16 Å and a Y315 Chi1 dihedral of -50 to -70 degrees, and the target open state is defined by R211-Y315 distance > 16.0 Å and a Y315 Chi1 dihedral angle of -180 to 180 degrees.

Using the Python library WE Simulation Toolkit with Parallelization and Analysis (WESTPA) 2.0, we ran WE for both the A3Bctd and fl-A3B.[37] We ran thousands of short MD simulations (0.05 ns) in parallel while iteratively replicating trajectories that transitioned from the closed to open state. Furthermore, a minimum adaptive binning (MAB) scheme was applied.[38] Here, we maximized the successful trajectories by feeding back these trajectories to the initial state (i.e., steady state weighted ensemble).[39] WE increased the sampling of the fl-A3B active site opening without altering the system's free energy. The closed active site conformations derived from the fl-A3B GaMD simulations served as the initial structures for subsequent WE simulations. We simulated three independent replicates of fl-A3B closed-to-open state transitions to rigorously sample this active site opening.

Using the 2D progress coordinate, we observed over 7000 closed-to-open transitions leading to the active site opening of both A3Bctd and fl-A3B (Figure 4). As illustrated in Figure 4, including the NTD in the simulations increased the frequency of the fl-A3B active site opening. WE simulations demonstrated the efficient sampling of active site opening for A3Bctd and fl-A3B. Here, we observed that both models sampled the open states (i.e., -180.0 and 180.0 degrees). With Y315 Chi1 dihedral angle, the lowest energy minima accounted were at closed state, -60.0 degree and at open state -180 and 180 degrees. The wide distribution of R211-Y315 distance for the fl-A3B value was also observed. This is in contrast with the A3Bctd which is mostly concentrated at around 6.0 Å. Furthermore, in fl-A3B closed state Y315 Chi1 dihedral angle, we identified how the phenyl side chain blocked the ssDNA site (Figure 4C). The R211-Y315 distance distribution is much higher in fl-A3B than in A3Bctd. The fl-A3B R211-Y315 distance even reached >19.0 Å (Figure 4D). The low values of the free energy of closed-to-open state transition in WE were due to the lack of convergence. In both all-atom MD and WE method simulations, we observed a more frequent opening of the ssDNA active site in fl-A3B than in the A3Bctd alone.
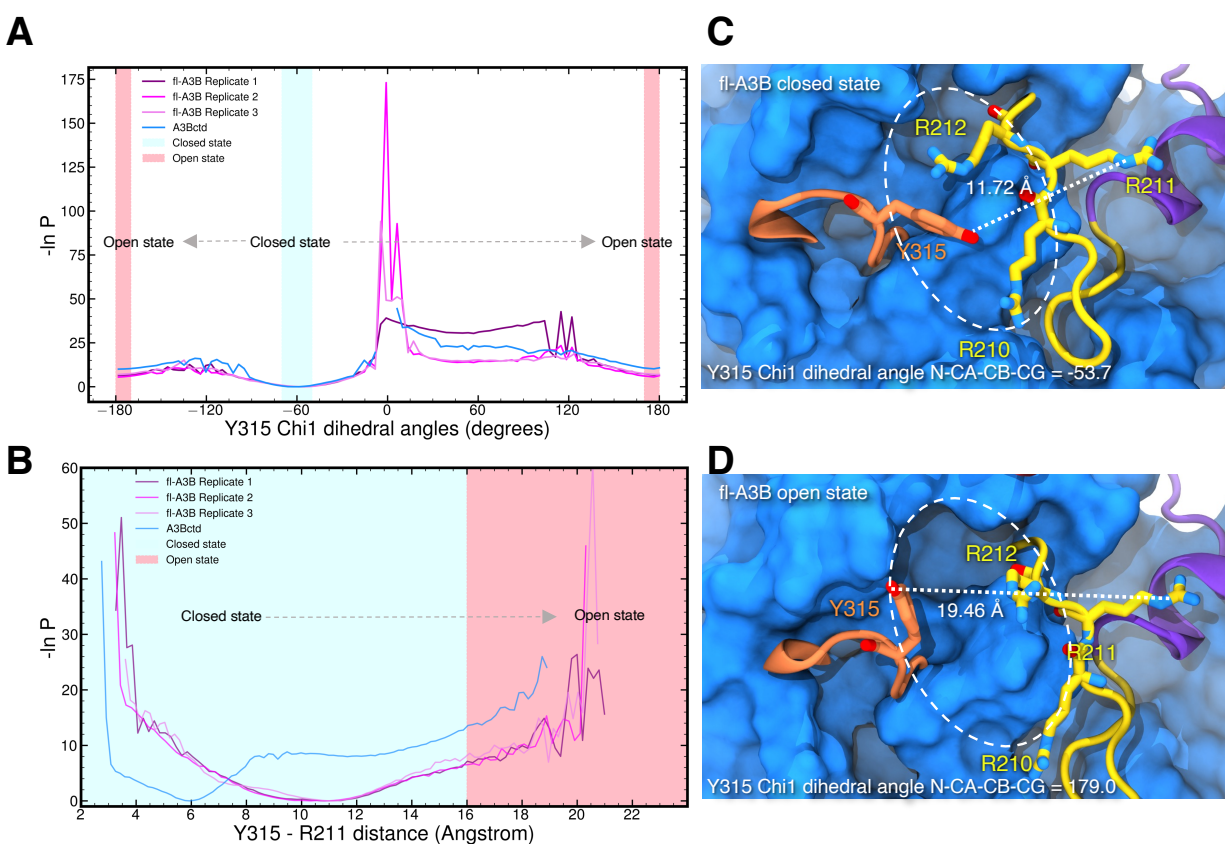
Figure 4. APOBEC3B ssDNA active site opening from steady-state weighted ensemble simulations. One-dimensional progress coordinate (A) Y315 Chi1 dihedral angle; (B) Y315-R211 distance; Probability distribution is plotted in an inverted natural log scale with free energy unit expressed as $1/k_BT$. Transition of the closed state (blue region) to open state (red region) active site was marked with gray arrow. WE simulations of fl-A3B closed-to-open active site transitions were run in 3 replicates while A3Bctd with 1 replicate. (C) fl-A3B closed state; (D) fl-A3B open state. Active site was enclosed in an oval dashed line along with the corresponding Y315 Chi1 dihedral angle and R211-Y315 distance.

The N-terminal domain residues T146, D149, E151, and E152 form salt bridges with R210, R211, and R212

We hypothesized that the NTD of fl-A3B plays a crucial role in molecular recognition and stability, as evidenced by the binding and interactions of longer ssDNA (>10 nucleotides) with the NTD.[40] In the apo form of fl-A3B, the residues T146, D149, E151, and E152 (in the NTD) form electrostatic salt bridges with residues R210, R211, and R212. We illustrated these interactions during 1 μs simulations, showing a closer distance at the domain interface that impacts ssDNA binding (Figure 5A).

We used Adaptive Poisson-Boltzmann Solver (APBS) to visualize the electrostatic of the fl-A3B.[41] Individual APBS calculations from multiple clustered structures from our MD simulations were averaged to consider the structural dynamics of the fl-A3B systems. In APBS calculations of the apo form and ssDNA bound fl-A3B, there is a highly negative patch at the NTD-CTD interface (Figure 5B). Interestingly, the positive charges in the side chains of R210, R211, and R212 can be attracted by the negative charges at D149, E151, and E152 allowing the favorable opening of the ssDNA active site due to a strong electrostatic interaction. This structural feature aligns with observed differences in enzymatic activity between fl-A3B and A3Bctd.
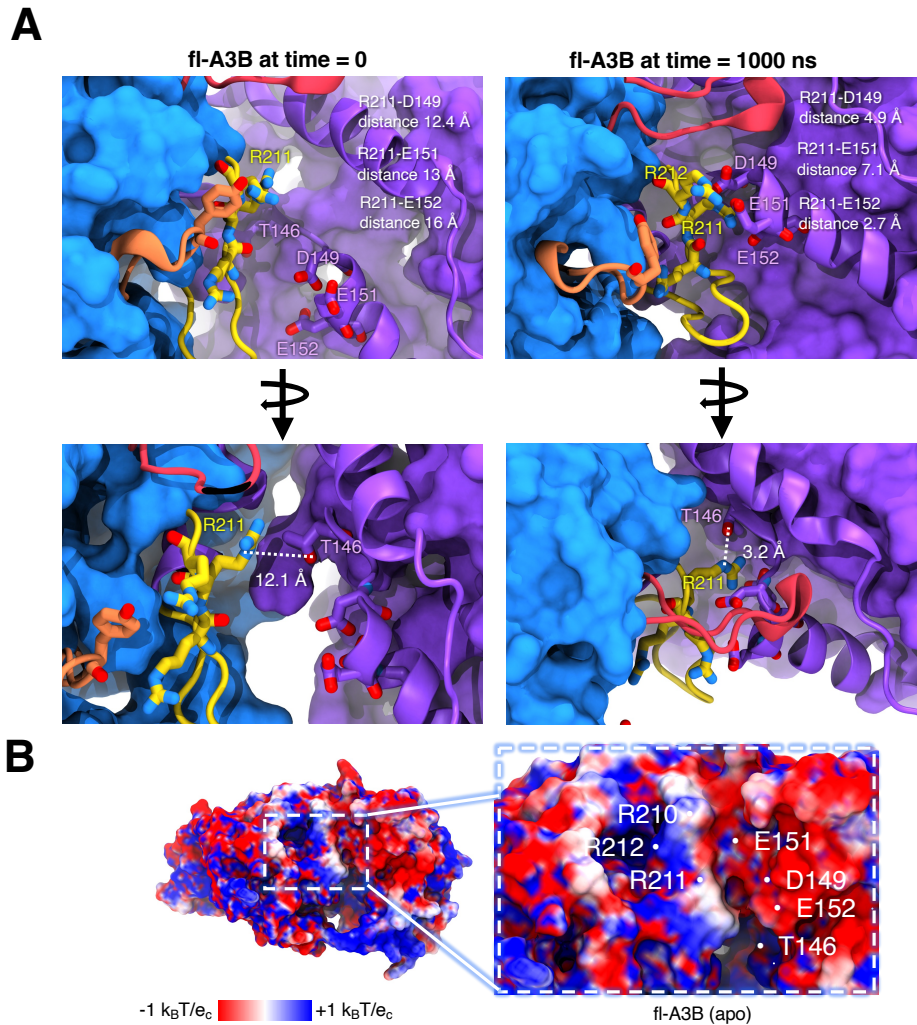
Figure 5. fl-A3B conform... ...residues with the NTD residues. (A) Intera... ...0, R211, R212 with T146, D149, E151, ...lation on fl-A3B rendered in the top most...

Presence of the APOBE...

To assess relative rates ... ...ombinant proteins from human Expi293F cells and quantified activity using a recently developed real-time APOBEC3-mediated DNA deamination (RADD) assay.[42] The RADD assay provides a direct read-out of APOBEC activity by including excess endonuclease Q (EndoQ) in each reaction, which functions downstream of C-to-U deamination to cleave newly uracilated substrate ssDNA oligonucleotide and thereby liberate fluorescence signal. Each enzyme was compared to a baseline oligo control reaction. A3Bctd showed nearly 50% lower deaminase activity compared to fl-A3B , as observed in real-time fluorescence readings (slope: 4.340 and 7.146, respectively) (Figure 6).

fl-A3B exhibits significantly higher activity and stability compared to A3Bctd as from previous experimental work. This aligns with the findings from Belica et al., who noted enhanced activity in both wild-type fl-A3B and A3Bctd,[42] as well as with reports from Papini et al., which highlight the NTD's role in allosteric regulation during the deamination reaction in longer ssDNA sequences.[40] Kinetic analyses reveal that fl-A3B has comparable K_m and higher K_{cat} than A3Bctd, underscoring its improved catalytic efficiency. Furthermore, in *Escherichia coli*, the expression and activity of fl-A3B exceed those of A3Bctd by four- to

eight-fold.[43] Further corroborating these functional enhancements, Butt et al. reported that the NTD not only accelerates substrate search but also contributes to the overall enzyme stability.[43] This is despite fl-A3B and A3Bctd having similar preferences for structured ssDNA targets, such as hairpin loops over linear DNA forms.

The NTD of fl-A3B is crucial for its biological function, influencing both cellular localization and ssDNA interactions.[44,45] Key residues within the NTD, such as Y18 and D19, are essential for nuclear localization, and W127 is important for ssDNA binding and regulation, as mutations in these residues lead to reduced catalytic efficiency.[44] The NTD also enhances the enzyme's functional stability and interaction with ssDNA, supported by its role in forming high molecular weight complexes and its involvement in RNA-dependent attenuation processes.[14] This evidence synergistically provides a more comprehensive view of how the NTD can have multiple roles in ssDNA deamination, activation, allosteric inhibitor recognition, and nucleic acid substrate processivity.

Overall, we surmised how the NTD can increase the activity of fl-A3B and its unique interaction with the CTD, which has implications for several cancers in terms of inhibitor discovery and design. This multifunctional domain thus plays a pivotal role in the dynamical and functional properties of fl-A3B, highlighting its significance beyond nucleating RNA interactions.
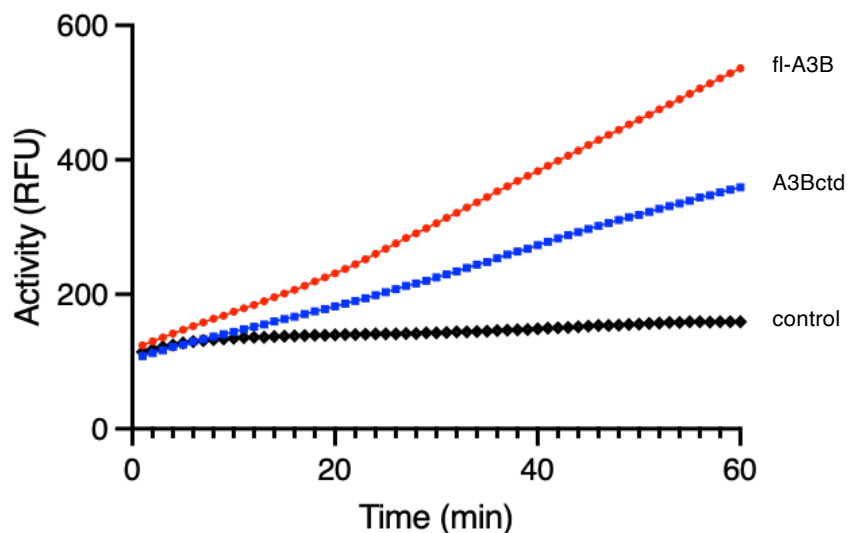
Figure 6. RADD assay readout for fl-A3B and A3Bctd. The fluorescent signal of 100 nM fl-A3B (red) and A3Bctd (blue) for a 1-hour reaction containing 0.8 μM reporter oligo is shown. The negative control (black) for the reaction contains 0.8 μM reporter oligo (mean fluorescent signal shown for n = 3 biological replicates).

Summary and Conclusion

A long-term goal for fl-A3B is to fully characterize the thermodynamics, kinetics, and structural and dynamic properties to better inform drug discovery efforts against various cancer types. This study employed multi-microsecond, all-atom, explicitly solvated MD simulations, including conventional, GaMD, and WE simulations, to explore the conformational dynamics of fl-A3B with a particular focus on the active site opening. Our results demonstrate that the N-terminal domain (NTD) significantly enhances the likelihood of active site opening in fl-A3B through salt bridge interactions between active site residues R210, R211, R212, and NTD residues T146, D149, E151, and E152. Along with the available experimental results, we showed that the interaction of the CTD and NTD in the fl-A3B facilitate the active site opening, thus making the ssDNA deamination more efficient. The insights gained from this mechanistic exploration inform our general

understanding of DNA editing systems and may improve attempts to develop more selective APOBEC inhibitors.

Systems and Methods

In summary, we built a complete full-length A3B model based on the known A3Bntd and A3Bctd X-ray structures. This model was compared with cryo-EM micrographs of the fl-A3B. After validating our model, we ran all-atom MD simulations (i.e., conventional, GaMD, and WE simulations) to sample structural changes in the two domains and the ssDNA active site. Structural determinants of active site opening were measured, analyzed, and complemented by A3B deamination assays (i.e. RADD).

Full-length APOBEC3B model generation

To provide a 3D model of the wild-type A3Bctd, we reverted the crystallized ssDNA-bound A3Bctd construct (PDB ID 5TD5)[10] back to wild-type (WT) using *in silico* modeling tools as described in our previous work.[15,21] SI Figure 1 outlines our fl-A3B model construction. Using the same protocol, the structure of an A3Bntd construct (PDB ID 5TKM)[14] was also reverted back to wild-type. The ClusPro web server was then used to generate 30 protein-protein docked poses of A3Bctd and A3Bntd.[46] Filtering out the poses that had unreasonable linker distance between the N-terminal residue of A3Bctd and the C-terminal residue of A3Bntd left us with three docked poses. We then extended the missing 6 residues at the N-terminus of A3Bntd and the missing 3 residues at the C-terminus of A3Bctd using Chemical Computing Group's MOE molecular modeling tools for all 3 docked poses selected.[47] Finally, we used Schrodinger's Prime module to generate 3 homology models of fl-A3B with the sequence from UniProt (ID Q9UH17) using the 3 docked poses as templates.[24] The closest model to the cryo-EM micrograph data was used in subsequent MD simulations.

Conventional all-atom molecular dynamics (MD) simulations

All-atom MD simulations were run on the Delta supercomputing system at the National Center for Supercomputing Applications (NCSA) using AMBER's GPU MD engine pmemd.cuda.[29] Our MD simulation workflow was consistent with the previous work reported.[15,21] Four replicates for each of the systems were simulated. AMBER ff14SB and TIP3P force fields were used for protein and water, respectively.[30,31] VMD PROPKA was used for assigning protein protonation states at pH 7.[48] The Zn^{2+} ion was modeled with the Cationic Dummy Atom Model.[49] Each system underwent energy minimization, followed by gradual heating, and equilibration with decreasing restraints. In the first 500 steps of minimization, only hydrogen atoms were minimized leaving all other atoms fixed. In the second 500 steps, all hydrogen atoms, water molecules, and ions (except Zn^{2+} ion) were minimized. In the third 500 steps, all atoms except the protein backbone were minimized. The entire systems were then minimized without any restraints for 40,000 steps. After the minimizations, each system was slowly heated up from 0 K to 310 K using a positional restraint of 2.0 kcal/mol/$Å^2$ on all non-hydrogen atoms of protein except the zinc-coordinating residues. Each system was then relaxed further in three consecutive 250 ps restrained MD simulations at 310 K and 1 atm using restraints of 3.0, 2.0, and 1.0 kcal/mol/$Å^2$ on all non-hydrogen atoms of protein except the zinc-coordinating residues. In total, we gathered 4 replicates of one microsecond, totaling 4 µs MD simulation data. We also used the apo form A3Bctd MD simulations datasets we previously reported for comparison.[15,21]

Gaussian accelerated molecular dynamics (GaMD) simulations

After this multi-step system relaxation protocol, GaMD simulations were performed without any restraints.[34] A 2 fs time step was used. Three replicates of GaMD were run based on the three models of fl-A3B generated. Temperature was maintained at 310 K using Langevin dynamics with a collision frequency of 5 ps and pressure was maintained at 1 atm using isotropic position scaling with a pressure relaxation time of 2 ps. Long-range electrostatics was treated by the Particle Mesh Ewald method, and a non-bonded cutoff of 10 Å was used.[29] The SHAKE algorithm was used to constrain the bonds involving hydrogen atoms.[50] Given an average system size for protein, the maximum, minimum, average, and standard deviation values of the system potential have been obtained from an initial 1.2 ns NPT simulation with no boost potential.

Finally, 300 ns of GaMD simulations have been carried out. Each system was simulated in 3 replicates, differing in the initialization seed for the velocities, and each replicate underwent 300 ns of unrestrained MD simulation in an NPT ensemble.

Weighted ensemble (WE) method simulations

WE Simulation Toolkit with Parallelization and Analysis (WESTPA) 2.0, a software package in Python and a tool for constructing and running stochastic simulations with the weighted ensemble method approach, was utilized.[37,51,52] To initialize the WE simulations, we employed 100 closed fl-A3B active site structures, each assigned equal statistical weights of 0.01. Progress coordinates (fl-A3B Y315 Chi1 dihedral angle and the R211-Y315 distance), which define the closed and open active site states, were calculated using CPPTRAJ.[53] A Y315 Chi1 dihedral angle is considered closed or open if the values -60 and 180 degrees are reached, respectively. For tyrosine residue, the Chi1 dihedral angle is defined as the angle between atoms N-CA-CB-CG where the CG *cis* to N is defined as a 0-degree angle. The target open state, defined by a Y315 Chi1 dihedral angle of -180 to 180 degrees and an R211-Y315 distance of at least 16.0 Å, was continually sampled. Trajectories that reached the open state in a non-equilibrium steady state were recycled, initiating new trajectories from the initial state with carried-over trajectory weight to enhance sampling of close-to-open active site transitions.[39]Non-equilibrium steady-state WE simulations were conducted to sample transitions from the closed to open active site states efficiently.

A short simulation time of 50 ps, denoted as τ, was implemented for resampling using the Minimum Adaptive Binning (MAB) scheme.[38] In this scheme, the positions of bins are strategically placed along the defined progress coordinates (i.e., Y315 dihedral angle and R211-Y315 distance) and adjusted according to trailing and leading trajectories using the westpa.core.binning module specified in the west.cfg file.

Our simulations were executed at the National Center for Supercomputing Applications (NCSA) Delta supercomputer utilizing NVIDIA A100 GPUs. Dynamics were propagated using AMBER pmemd.cuda with the temperature maintained at 300 K using a Langevin thermostat and pressure controlled by a Berendsen barostat within an orthogonal box.[29] The fl-A3B closed-to-open state non-equilibrium steady state weighted ensemble simulations were conducted in triplicates, accumulating a total simulation time of 45.58 μs data.

MD simulations data analysis

Distances between the center of mass of residue were measured using pytraj.[53] The average potentials from these clusters were plotted back to the structure of fl-A3B. The root mean square deviation (RMSD) of the protein backbone atoms (C, CA, N, and O) was calculated with pytraj, using the first frame of the production run as a reference. To account for global translational and rotational movements in the RMSD calculations, all frames were fitted to the reference frame. A combined histogram of these RMSD values was also generated. Visualization of all structures and trajectory movies was done using VMD.[54] Clustering based on the RMSD of both R211 and Y315 Cα was done using the Python package scikit-learn's kmeans method.[55] The clustered structures were also subjected to Adaptive Poisson Boltzmann Solver (APBS) electrostatic calculations using the APBS package (version 3.0).[41]

fl-A3B protein expression and purification for cryo-EM data collection

The A3B protein used in cryoEM studies contained human A3B residues 7 to 378 fused to maltose-binding protein (MBP) on the N-terminus and 6xHis-tag on the C-terminus. This A3B construct had solubility-enhancing mutations (Y13D, F61S, Y65H, Y83D, W127S, T146K, Y162D, F200S, W228S, L230K, F308K) and an internal truncation of loop 3.[16] The protein was expressed in *E. coli* strain BL21(DE3) and purified by nickel-affinity and size-exclusion chromatography as a monomer. Purified protein in 20 mM Tris-HCl (pH 7.4), 0.5 M NaCl, and 5 mM 2-mercaptoethanol was flash-frozen and stored at –80°C. The protein concentration was determined based on UV absorption at 280 nm.

CryoEM data collection

Purified fl-A3B, prepared at 1 mg/mL, was deposited (3 μL) onto UltrAuFoil R1.2/1.3 300-mesh grids, which were cleaned for 10 seconds using a Solarus plasma cleaner (Gatan, Inc.) at 15 watts in a 75% nitrogen/25% oxygen atmosphere. The sample was manually blotted for 6-7 seconds with Whatman No. 1 filter paper and immediately plunge-frozen in liquid ethane, cooled by liquid nitrogen, under cold room conditions (^{3} 95% humidity at 4°C). Squares were selected for the thinnest ice.

Micrographs of the A3B complex were acquired using a Talos Arctica electron microscope (Thermo Fisher Scientific) operating at 200 kV, equipped with a K2 Summit direct electron detector (Gatan, Inc.). A total of 1,420 micrographs were recorded. Movies of A3B were captured in counting mode at a pixel size of 1.16 Å/pixel and a magnification of 36,000x, with a defocus range between −1.7 μm and −2.1 μm. The acquisition was performed using an exposure rate of 6.98 e$^-$/pixel/s over 12.5 seconds (250 ms per frame, 50 frames), resulting in a total dose of approximately 65 e$^-$/Å^2 (1.32 e$^-$/Å^2 per frame). Patch-based motion correction and contrast transfer function (CTF) estimation were carried out in cryoSPARC Live.[56] Particle picking was performed using the Blob Picker, with particle diameters ranging from 50 Å to 200 Å. A total of 3,417,871 particles were extracted with a box size of 256 pixels. One round of 2D classification was conducted using 50 classes, an initial uncertainty factor of 2, and an inner mask diameter of 240 Å. An ab initio reconstruction was generated from 100,000 particles, with an initial resolution of 35 Å and a maximum resolution of 12 Å. The final reconstruction achieved a resolution of 12 Å.

fl-A3B and A3Bctd protein purification

Protein expression and purification as previously described with modifications.[57] Briefly, using the standard Expi293 transfection protocol (ThermoFisher), 25 mL cultures of Expi293F cells (ThermoFisher) were transfected with pcDNA3.1-hA3Bi-mycHis or pcDNA3.1-A3Bctd (193-382)-mycHis. Transfection enhancers were added 18-20 hours post-transfection. Cells were harvested 3 days post-transfection by centrifugation at 1,000 rpm for 5 minutes; then the cell pellets were frozen at -80°C. Cells were thawed on ice, then resuspended in 15 mL lysis buffer (25 mM Tris-Cl pH 8.0, 500 mM NaCl, 5 mM MgCl$_2$, 20 mM imidazole, 5% glycerol, and 0.1% IGEPAL CA-630), followed by sonication on ice for 2 minutes, at power 5 and 40% duty. RNase was added to 100 μg/mL, as well as 100 U of Salt Active Nuclease (MilliporeSigma), and incubated at 37°C for 2 hours for nucleases to act. Following centrifugation (16,000 × g, 30 minutes), clarified lysates were collected, and sodium chloride was added to a final concentration of 1 M. 100 μL Ni-NTA Superflow resin (Qiagen) was added to the lysates and incubated for 1 hour at 4°C with agitation. The resin was washed with 1 CV wash buffer (25 mM HEPES pH 7.4, 0.1% Triton X-100, 10% glycerol, 300 mM NaCl and 40 mM imidazole) and then incubated 3 times with 200 μL of an elution buffer (25 mM HEPES pH 7.4, 0.1% Triton X-100, 20% glycerol, 150 mM NaCl, 400 mM imidazole, and 1 mM TCEP) for 10 minutes to elute proteins. Elutions were then pooled and concentrated using an Amicon (MilliporeSigma) device with a 10 kDa cut-off before storage. Proteins were snap-frozen in liquid nitrogen and stored in 5 μL aliquots at -80°C. Once thawed, proteins were kept on ice for no longer than 24 hours and never refrozen.

Pyrococcus furiosis Endonuclease Q purification

BL21(DE3)-Star cells were transformed with pET24a-PfuEndoQ-6xHis.[58] PfuEndoQ was purification was adapted from previous works.[58,59] Briefly, single colonies were picked into an overnight starter culture in Terrific Broth, which was used to inoculate 1 L of Terrific Broth. Cells were grown at 37°C until the OD600 reached ~1. Expression of Pfu EndoQ was induced by the addition of 0.5 mM IPTG followed by shaking overnight at 18°C. Cells were harvested by centrifugation and lysed by addition of 100 mL of 50 mM Tris-Cl pH 8.0, 10% glycerol, 600 mM NaCl, 1 mM MgCl$_2$, two Roche cOmplete EDTA-free Protease Inhibitor Cocktail tablets (MilliporeSigma), and hen egg white lysozyme to 1 mg/mL followed by rocking for 30 minutes. The lysate was then sonicated with a Branson Sonifier 450 for 4 two-minute cycles at 60% duty cycle, power 6, with at least 1 minute between runs. Imidazole was added to 20 mM and TCEP to 0.5 mM and the lysate was heated to 70°C for 30 minutes to denature most of the *E. coli* proteins. The lysate was then centrifuged for 30 minutes at 16,000 × g to pellet insoluble material. The resulting clarified lysate was purified over a two mL Ni-NTA column (Qiagen 30230), washed twice with 5 column volumes (CV) of 50 mM Tris-Cl pH 8.0, 300 mM NaCl, 20 mM imidazole, and eluted in 12 mL of 50 mM Tris-Cl pH 8.0, 150 mM

NaCl, 400 mM imidazole, 0.5 mM TCEP. Following analysis by SDS-PAGE, the protein was dialyzed into storage buffer (25 mM Tris-Cl pH 8, 150 mM NaCl, 50% glycerol) and stored at -20°C with 0.5 mM TCEP.

Real-time APOBEC3-mediated DNA deamination (RADD) assay

Each 20 μL reaction was carried out on a flat-bottom black 384-well plate (Corning Ref. 3821). Prepared enzyme mixture, consisting of 200 nM A3B (A3Bfl, A3Bctd or control) in 1x NEBuffer 2 (New England Biolabs) made in nuclease-free water and oligo mixture, consisting of 4 μM EndoQ and 1.6 μM TC oligo (5'-6-FAM-TAGGT**C**ATTATTGTG-IAB-3') in 1x NEBuffer 2 made in nuclease-free water. Added 10 μL of enzyme mixture to each well, followed by 10 μL of oligo mixture (final concentrations for A3B, EndoQ, and substrate oligonucleotide are 100 nM, 2 μM, and 0.8 μM, respectively). All fluorescence readings were recorded instantly on a TECAN Spark 10M instrument in concert with the SparkControl V3.0 software. Data was collected every minute for 1 hour at 37°C using the scanning fluorescence kinetic mode (excitation: 485 nm; emission: 535 nm) with the gain set to 50%. Slope determination was performed by simple linear regression using Prism 10.

Conflict of Interest

The authors declare no conflict of interest.

Authors Contribution

R.E.A., H.A., and R.S.H. designed and oversaw the research project and secured funding. M.K.E.B. and Ö.D. built the full-length APOBEC3B model and performed MD simulations. M.K.E.B., Ö.D., S.-H.A., and C.K.M. performed weighted ensemble simulations. M.K.E.B., Ö.D., S.-H.A., C.K.M., C.C.T., L.C., and R.E.A. performed simulation analyses and data visualizations. M.A.C., Y.C., B.d.I.P.A., R.S.H., and H.A. run protein biochemistry and purification experiments. K.L.M. and M.A.H.,Jr. processed cryoEM dataset. M.K.E.B., Ö.D., and R.E.A. wrote the initial draft of the paper with contributions from all authors. All authors reviewed and approved the manuscript.

Software and Data Availability

Trajectories, structures, simulation scripts, analysis scripts, and data files are uploaded in our group's website and can be accessed thru this link https://amarolab.ucsd.edu/data.php . All software used is community developed and freely available.

Supplementary Information

SI Figures 1 to 6.

SI Movies 1 to 2.

Author Information

ORCID

Mac Kevin E. Braza: 0000-0002-6929-7397
Özlem Demir: 0000-0002-3317-4731
Surl-Hee Ahn: 0000-0002-3422-805X
Clare K. Morris: 0000-0002-4314-5387
Carla Calvó-Tusell: 0000-0003-2681-8460
Kelly L. McGuire: 0009-0005-7520-3675
Michael A. Carpenter: 0000-0001-9524-5323
Yanjun Chen: 0009-0000-2450-3401
Bárbara de la Peña Avalos: 0000-0001-9796-8015
Lorenzo Casalino: 0000-0003-3581-1148
Hideki Aihara: 0000-0001-7508-6230
Reuben S. Harris: 0000-0002-9034-9112
Mark A. Herzik, Jr.: 0000-0001-6653-6682
Rommie E. Amaro: 0000-0002-9275-9553

References

(1)   Harris, R. S.; Dudley, J. P. APOBECs and Virus Restriction. *Virology* **2015**, *479–480*, 131–145. https://doi.org/10.1016/J.VIROL.2015.03.012.
(2)   Harris, R. S.; Liddament, M. T. Retroviral Restriction by APOBEC Proteins. *Nat Rev Immunol* **2004**, *4* (11), 868–877. https://doi.org/10.1038/NRI1489.
(3)   Alexandrov, L. B.; Kim, J.; Haradhvala, N. J.; Huang, M. N.; Tian Ng, A. W.; Wu, Y.; Boot, A.; Covington, K. R.; Gordenin, D. A.; Bergstrom, E. N.; Ashiqul Islam, S. M.; Lopez-Bigas, N.; Klimczak, L. J.; McPherson, J. R.; Morganella, S.; Sabarinathan, R.; Wheeler, D. A.; Mustonen, V.; Getz, G.; Rozen, S. G.; Stratton, M. R.; Boutros, P.; Chan, K.; Fujimoto, A.; Kazanov, M.; Lawrence, M.; Martincorena, I.; Nakagawa, H.; Polak, P.; Prokopec, S.; Roberts, S. A.; Saini, N.; Shibata, T.; Shiraishi, Y.; Teh, B. T.; Vázquez-García, I.; Yousif, F.; Yu, W. The Repertoire of Mutational Signatures in Human Cancer. *Nature* **2020**, *578* (7793), 94–101. https://doi.org/10.1038/s41586-020-1943-3.
(4)   Durfee, C.; Temiz, N. A.; Levin-Klein, R.; Argyris, P. P.; Alsøe, L.; Carracedo, S.; Alonso de la Vega, A.; Proehl, J.; Holzhauer, A. M.; Seeman, Z. J.; Liu, X.; Lin, Y. H. T.; Vogel, R. I.; Sotillo, R.; Nilsen, H.; Harris, R. S. Human APOBEC3B Promotes Tumor Development in Vivo Including Signature Mutations and Metastases. *Cell Rep Med* **2023**, *4* (10). https://doi.org/10.1016/j.xcrm.2023.101211.
(5)   Lackey, L.; Law, E. K.; Brown, W. L.; Harris, R. S. Subcellular Localization of the APOBEC3 Proteins during Mitosis and Implications for Genomic DNA Deamination. *Cell Cycle* **2013**, *12* (5), 762–772. https://doi.org/10.4161/CC.23713.
(6)   Sieuwerts, A. M.; Willis, S.; Burns, M. B.; Look, M. P.; Gelder, M. E. M. Van; Schlicker, A.; Heideman, M. R.; Jacobs, H.; Wessels, L.; Leyland-Jones, B.; Gray, K. P.; Foekens, J. A.; Harris, R. S.; Martens, J. W. M. Elevated APOBEC3B Correlates with Poor Outcomes for Estrogen-Receptor-Positive Breast Cancers. *Horm Cancer* **2014**, *5* (6), 405–413. https://doi.org/10.1007/S12672-014-0196-8/.
(7)   Law, E. K.; Sieuwerts, A. M.; Lapara, K.; Leonard, B.; Starrett, G. J.; Molan, A. M.; Temiz, N. A.; Vogel, R. I.; Meijer-Van Gelder, M. E.; Sweep, F. C. G. J.; Span, P. N.; Foekens, J. A.; Martens, J. W. M.; Yee, D.; Harris, R. S. The DNA Cytosine Deaminase APOBEC3B Promotes Tamoxifen

Resistance in ER-Positive Breast Cancer. *Sci Adv* **2016**, *2* (10). https://doi.org/10.1126/sciadv.1601737.

(8)   Petljak, M.; Dananberg, A.; Chu, K.; Bergstrom, E. N.; Striepen, J.; von Morgen, P.; Chen, Y.; Shah, H.; Sale, J. E.; Alexandrov, L. B.; Stratton, M. R.; Maciejowski, J. Mechanisms of APOBEC3 Mutagenesis in Human Cancer Cells. *Nature* **2022**, *607* (7920), 799–807. https://doi.org/10.1038/s41586-022-04972-y.

(9)   Petljak, M.; Alexandrov, L. B.; Brammeld, J. S.; Nik-Zainal, S.; Campbell, P. J.; Stratton Correspondence, M. R. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell* **2019**, *176*, 1282-1294.e20. https://doi.org/10.1016/j.cell.2019.02.012.

(10)   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.; Aihara, H. Structural Basis for Targeted DNA Cytosine Deamination and Mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol* **2016**, *24* (2), 131–139. https://doi.org/10.1038/nsmb.3344.

(11)   Fanunza, E.; Cheng, A. Z.; Auerbach, A. A.; Stefanovska, B.; Moraes, S. N.; Lokensgard, J. R.; Biolatti, M.; Dell'Oste, V.; Bierle, C. J.; Bresnahan, W. A.; Harris, R. S. Human Cytomegalovirus Mediates APOBEC3B Relocalization Early during Infection through a Ribonucleotide Reductase-Independent Mechanism. *J Virol* **2023**, *97* (8). https://doi.org/10.1128/jvi.00781-23/.

(12)   Cheng, A. Z.; Moraes, S. N.; Attarian, C.; Yockteng-Melgar, J.; Jarvis, M. C.; Biolatti, M.; Galitska, G.; Dell'Oste, V.; Frappier, L.; Bierle, C. J.; Rice, S. A.; Harris, R. S. A Conserved Mechanism of APOBEC3 Relocalization by Herpesviral Ribonucleotide Reductase Large Subunits. *J Virol* **2019**, *93* (23). https://doi.org/10.1128/jvi.01539-19/.

(13)   Maiti, A.; Hou, S.; Schiffer, C. A.; Matsuo, H. Interactions of APOBEC3s with DNA and RNA. *Curr Opin Struct Biol* **2021**, *67*, 195–204. https://doi.org/10.1016/J.SBI.2020.12.004.

(14)   Xiao, X.; Yang, H.; Arutiunian, V.; Fang, Y.; Besse, G.; Morimoto, C.; Zirkle, B.; Chen, X. S. Structural Determinants of APOBEC3B Non-Catalytic Domain for Molecular Assembly and Catalytic Regulation. *Nucleic Acids Res* **2017**, *45* (12), 7494–7506. https://doi.org/10.1093/nar/gkx362.

(15)   Shi, K.; Demir, Ö.; 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), 1–12. https://doi.org/10.1038/s41598-017-17694-3.

(16)   Shi, K.; Carpenter, M. A.; Kurahashi, K.; Harris, R. S.; Aihara, H. Crystal Structure of the DNA Deaminase APOBEC3B Catalytic Domain. *Journal of Biological Chemistry* **2015**, *290* (47), 28120–28130. https://doi.org/10.1074/JBC.M115.679951.

(17)   Carpenter, M. A.; Temiz, N. A.; Ibrahim, M. A.; Jarvis, M. C.; Brown, M. R.; Argyris, P. P.; Brown, W. L.; Starrett, G. J.; Yee, D.; Harris, R. S. Mutational Impact of APOBEC3A and APOBEC3B in a Human Cell Line and Comparisons to Breast Cancer. *PLoS Genet* **2023**, *19* (11), e1011043. https://doi.org/10.1371/journal.pgen.1011043.

(18)   Tang, H.; Demir, Ö.; Kurniawan, F.; Brown, W. L.; Shi, K.; Moeller, N. H.; Carpenter, M. A.; Belica, C.; Orellana, K.; Du, G.; Lebeau, A. M.; Amaro, R. E.; Harris, R. S.; Aihara, H. Structural Characterization of a Minimal Antibody against Human APOBEC3B. *Viruses* **2021**, *13* (4), 663. https://doi.org/10.3390/v13040663.

(19)   Shi, K.; Demir, Ö.; Carpenter, M. A.; Banerjee, S.; Harki, D. A.; Amaro, R. E.; Harris, R. S.; Aihara, H. Active Site Plasticity and Possible Modes of Chemical Inhibition of the Human DNA Deaminase APOBEC3B. *FASEB Bioadv* **2020**, *2* (1), 49–58. https://doi.org/10.1096/fba.2019-00068.

(20)   Adolph, M. B.; Love, R. P.; Feng, Y.; Chelico, L. Enzyme Cycling Contributes to Efficient Induction of Genome Mutagenesis by the Cytidine Deaminase APOBEC3B. *Nucleic Acids Res* **2017**, *45* (20), 11925–11940. https://doi.org/10.1093/NAR/GKX832.

(21)   Wagner, J. R.; Demir, Ö.; Carpenter, M. A.; Aihara, H.; Harki, D. A.; Harris, R. S.; Amaro, R. E. Determinants of Oligonucleotide Selectivity of APOBEC3B. *J Chem Inf Model* **2018**, *59* (5), 2264–2273. https://doi.org/10.1021/acs.jcim.8B00427.

(22)   Li, Y. L.; Langley, C. A.; Azumaya, C. M.; Echeverria, I.; Chesarino, N. M.; Emerman, M.; Cheng, Y.; Gross, J. D. The Structural Basis for HIV-1 Vif Antagonism of Human APOBEC3G. *Nature* **2023**, *615* (7953), 728–733. https://doi.org/10.1038/s41586-023-05779-1.

(23)   Sanchez, A.; Ortega, P.; Sakhtemani, R.; Manjunath, L.; Oh, S.; Bournique, E.; Becker, A.; Kim, K.; Durfee, C.; Temiz, N. A.; Chen, X. S.; Harris, R. S.; Lawrence, M. S.; Buisson, R. Mesoscale DNA

Features Impact APOBEC3A and APOBEC3B Deaminase Activity and Shape Tumor Mutational Landscapes. *Nat Commun* **2024**, *15* (1), 1–16. https://doi.org/10.1038/s41467-024-45909-5.

(24) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins: Structure, Function, and Bioinformatics* **2004**, *55* (2), 351–367. https://doi.org/10.1002/PROT.10613.

(25) Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics* **2008**, *9* (1), 1–8. https://doi.org/10.1186/1471-2105-9-40/.

(26) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; De Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res* **2018**, *46* (W1), W296–W303. https://doi.org/10.1093/NAR/GKY427.

(27) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(28) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro Web Server for Protein–Protein Docking. *Nature Protocols 2017 12:2* **2017**, *12* (2), 255–278. https://doi.org/10.1038/nprot.2016.169.

(29) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **2013**, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314Y.

(30) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J Chem Theory Comput* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/ACS.JCTC.5B00255.

(31) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *Journal of Physical Chemistry A* **2001**, *105* (43), 9954–9960. https://doi.org/10.1021/JP003020W/.

(32) Hou, S.; Lee, J. M.; Myint, W.; Matsuo, H.; Yilmaz, N. K.; Schiffer, C. A. Structural Basis of Substrate Specificity in Human Cytidine Deaminase Family APOBEC3s. *Journal of Biological Chemistry* **2021**, *297* (2), 100909. https://doi.org/10.1016/J.JBC.2021.100909.

(33) Hou, S.; Silvas, T. V.; Leidner, F.; Nalivaika, E. A.; Matsuo, H.; Yilmaz, N. K.; Schiffer, C. A. Structural Analysis of the Active Site and DNA Binding of Human Cytidine Deaminase APOBEC3B. *J Chem Theory Comput* **2018**, *15* (1), 637–647. https://doi.org/10.1021/ACS.JCTC.8B00545.

(34) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J Chem Theory Comput* **2015**, *11* (8), 3584–3595. https://doi.org/10.1021/acs.jctc.5B00436.

(35) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu Rev Biophys* **2017**, *46*, 43–57. https://doi.org/10.1146/annurev-biophys-070816-033834.

(36) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophys J* **1996**, *70* (1), 97–110. https://doi.org/10.1016/S0006-3495(96)79552-8.

(37) Bogetti, A. T.; Leung, J. M. G.; Russo, J. D.; Zhang, S.; Thompson, J. P.; Saglam, A. S.; Ray, D.; Mostofian, B.; Pratt, A. J.; Abraham, R. C.; Harrison, P. O.; Dudek, M.; Torrillo, P. A.; DeGrave, A. J.; Adhikari, U.; Faeder, J. R.; Andricioaei, I.; Adelman, J. L.; Zwier, M. C.; LeBard, D. N.; Zuckerman, D. M.; Chong, L. T. A Suite of Tutorials for the WESTPA 2.0 Rare-Events Sampling Software [Article v2.0]. *Living J Comput Mol Sci* **2023**, *5* (1), 1655–1655. https://doi.org/10.33011/LIVECOMS.5.1.1655.

(38) Torrillo, P. A.; Bogetti, A. T.; Chong, L. T. A Minimal, Adaptive Binning Scheme for Weighted Ensemble Simulations. *Journal of Physical Chemistry A* **2021**, *125* (7), 1642–1649. https://doi.org/10.1021/ACS.JPCA.0C10724.

(39) Bhatt, D.; Zhang, B. W.; Zuckerman, D. M. Steady-State Simulations Using Weighted Ensemble Path Sampling. *J Chem Phys* **2010**, *133* (1), 014110. https://doi.org/10.1063/1.3456985.

(40) Papini, C.; Wang, Z.; Kudalkar, S. N.; Schrank, T. P.; Tang, S.; Sasaki, T.; Wu, C.; Tejada, B.; Ziegler, S. J.; Xiong, Y.; Issaeva, N.; Yarbrough, W. G.; Anderson, K. S. Exploring APOBEC3A and APOBEC3B Substrate Specificity and Their Role in HPV Positive Head and Neck Cancer. *iScience* **2022**, *25* (10), 105077. https://doi.org/10.1016/J.ISCI.2022.105077.

(41) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; Mccammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. *Wiley Online Library* **2018**, *27* (1), 112–128. https://doi.org/10.1002/pro.3280.

(42) Belica, C. A.; Carpenter, M. A.; Chen, Y.; Brown, W. L.; Moeller, N. H.; Boylan, I. T.; Harris, R. S.; Aihara, H. A Real-Time Biochemical Assay for Quantitative Analyses of APOBEC-Catalyzed DNA Deamination. *Journal of Biological Chemistry* **2024**, 107410. https://doi.org/10.1016/J.JBC.2024.107410.

(43) Butt, Y.; Sakhtemani, R.; Mohamad-Ramshan, R.; Lawrence, M. S.; Bhagwat, A. S. Distinguishing Preferences of Human APOBEC3A and APOBEC3B for Cytosines in Hairpin Loops, and Reflection of These Preferences in APOBEC-Signature Cancer Genome Mutations. *Nature Communications 2024 15:1* **2024**, *15* (1), 1–13. https://doi.org/10.1038/s41467-024-46231-w.

(44) Salamango, D. J.; McCann, J. L.; Demir, Ö.; Brown, W. L.; Amaro, R. E.; Harris, R. S. APOBEC3B Nuclear Localization Requires Two Distinct N-Terminal Domain Surfaces. *J Mol Biol* **2018**, *430* (17), 2695–2708. https://doi.org/10.1016/j.jmb.2018.04.044.

(45) Pak, V.; Heidecker, G.; Pathak, V. K.; Derse, D. The Role of Amino-Terminal Sequences in Cellular Localization and Antiviral Activity of APOBEC3B. *J Virol* **2011**, *85* (17), 8538–8547. https://doi.org/10.1128/JVI.02645-10.

(46) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro Web Server for Protein–Protein Docking. *Nature Protocols 2017 12:2* **2017**, *12* (2), 255–278. https://doi.org/10.1038/nprot.2016.169.

(47) Vilar, S.; Cozza, G.; Moro, S. Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. *Curr Top Med Chem* **2008**, *8* (18), 1555–1572. https://doi.org/10.2174/156802608786786624.

(48) Rostkowski, M.; Olsson, M. H.; Søndergaard, C. R.; Jensen, J. H. Graphical Analysis of PH-Dependent Properties of Proteins Predicted Using PROPKA. *BMC Struct Biol* **2011**, *11* (1), 1–6. https://doi.org/10.1186/1472-6807-11-6.

(49) Pang, Y. P. Successful Molecular Dynamics Simulation of Two Zinc Complexes Bridged by a Hydroxide in Phosphotriesterase Using the Cationic Dummy Atom Method. *Proteins: Structure, Function, and Bioinformatics* **2001**, *45* (3), 183–189. https://doi.org/10.1002/PROT.1138.

(50) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J Comput Phys* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(51) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suárez, E.; Lettieri, S.; Wang, D. W.; Grabe, M.; Zuckerman, D. M.; Chong, L. T. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. *J Chem Theory Comput* **2015**, *11* (2), 800–809. https://doi.org/10.1021/ct5010615.

(52) Russo, J. D.; Zhang, S.; Leung, J. M. G.; Bogetti, A. T.; Thompson, J. P.; Degrave, A. J.; Torrillo, P. A.; Pratt, A. J.; Wong, K. F.; Xia, J.; Copperman, J.; Adelman, J. L.; Zwier, M. C.; Lebard, D. N.; Zuckerman, D. M.; Chong, L. T. WESTPA 2.0: High-Performance Upgrades for Weighted Ensemble Simulations and Analysis of Longer-Timescale Applications. *J Chem Theory Comput* **2022**, *18* (2), 638–649. https://doi.org/10.1021/acs.jctc.1C01154.

(53) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/CT400341P/.

(54) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(55) Pedregosa, F.; Michel, V.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Vanderplas, J.; Cournapeau, D.; Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Thirion, B.; Grisel, O.; Dubourg, V.; Passos, A.; Brucher, M. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(56)   Punjani, A.; Rubinstein, J. L.; Fleet, D. J.; Brubaker, M. A. CryoSPARC: Algorithms for Rapid
        Unsupervised Cryo-EM Structure Determination. *Nature Methods 2017 14:3* **2017**, *14* (3), 290–
        296. https://doi.org/10.1038/nmeth.4169.
(57)   McCann, J. L.; Cristini, A.; Law, E. K.; Lee, S. Y.; Tellier, M.; Carpenter, M. A.; Beghè, C.; Kim, J. J.;
        Sanchez, A.; Jarvis, M. C.; Stefanovska, B.; Temiz, N. A.; Bergstrom, E. N.; Salamango, D. J.;
        Brown, M. R.; Murphy, S.; Alexandrov, L. B.; Miller, K. M.; Gromak, N.; Harris, R. S. APOBEC3B
        Regulates R-Loops and Promotes Transcription-Associated Mutagenesis in Cancer. *Nat Genet*
        2023, *55* (10), 1721–1734. https://doi.org/10.1038/s41588-023-01504-w.
(58)   Shi, K.; Moeller, N. H.; Banerjee, S.; McCann, J. L.; Carpenter, M. A.; Yin, L.; Moorthy, R.;
        Orellana, K.; Harki, D. A.; Harris, R. S.; Aihara, H. Structural Basis for Recognition of Distinct
        Deaminated DNA Lesions by Endonuclease Q. *Proc Natl Acad Sci U S A* **2021**, *118* (10),
        e2021120118. https://doi.org/10.1073/pnas.2021120118.
(59)   Shiraishi, M.; Ishino, S.; Yamagami, T.; Egashira, Y.; Kiyonari, S.; Ishino, Y. A Novel Endonuclease
        That May Be Responsible for Damaged DNA Base Repair in Pyrococcus Furiosus. *Nucleic Acids
        Res* **2015**, *43* (5), 2853–2863. https://doi.org/10.1093/NAR/GKV121.