

Computational investigation of APOBEC3H substrate orientation and selectivity

Mark A. Hix and G. Andrés Cisneros^{**}

Department of Chemistry, University of North Texas, Denton, Texas 76203, United States

E-mail: andres@unt.edu

APOBEC3 enzymes (A3s) are cytidine deaminases responsible for dC→dU mutations. Most A3s act selectively on a 5'-dTdC-3' motif, with the exception of A3G which prefers a 5'-dCdC-3' motif.¹⁻³ A3H is known for its role in innate human immunity to retroviruses including HIV.⁴ A3H acts upon a single-stranded DNA or RNA substrate and mutates cytidine to uracil, preferring a 5'-dTdCdA-3' sequence.⁵ There are several available crystal structures of A3H, however, none of the reported structures have been crystalized in complex with a substrate or substrate analog. Here, we present a computational investigation on the orientation of the substrate on A3H and possible structural determinants for the selectivity of the preferred 5'-dTdCdA-3' substrate motif.⁵

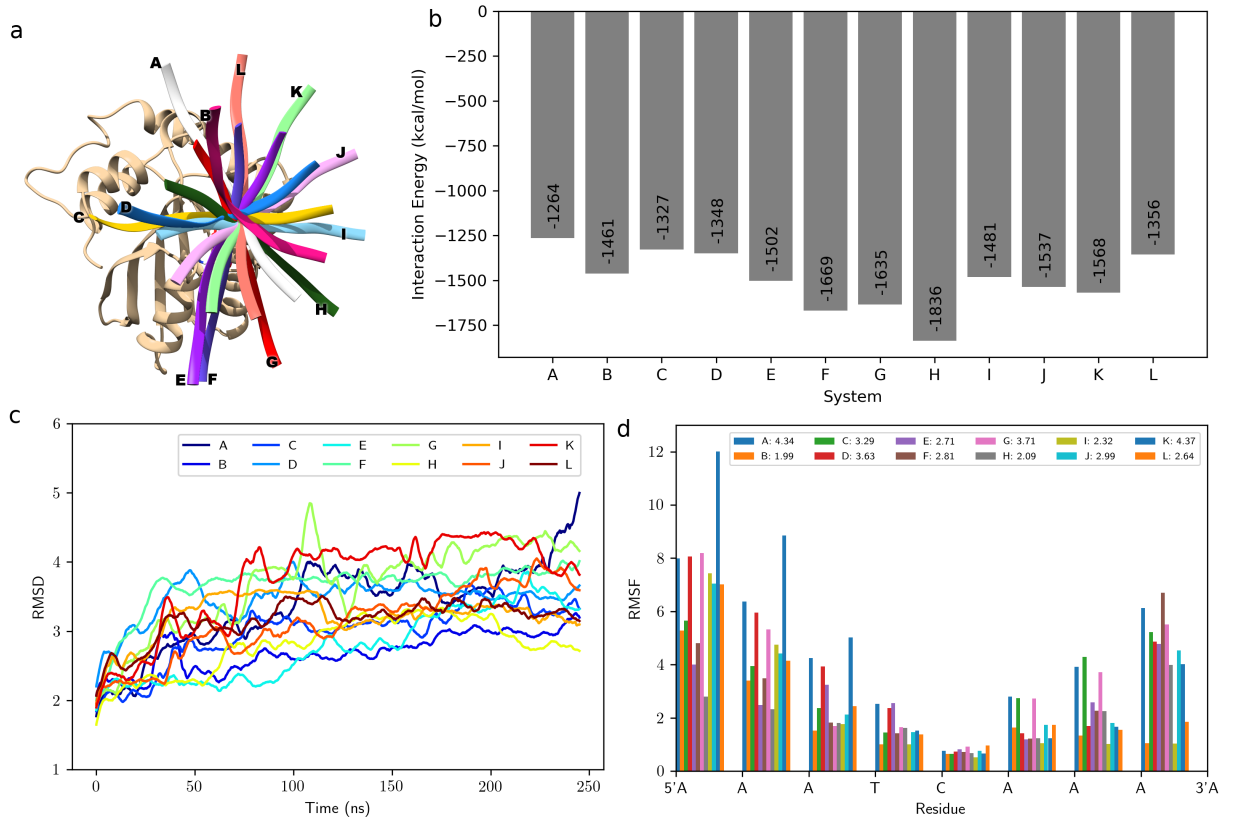


Figure 1: **a)** Twelve possible orientations of ssDNA substrate on A3H, **b)** Protein-substrate interaction energies, **c)** RMSD over time with respect to original crystal structure, and **d)** RMSF of nucleotides in ssDNA substrate.

The initial crystal structure for monomeric A3H was obtained from the RCSB Protein Data Bank (pdbid: 5W45).⁶ The sequence was confirmed to match the A3H-HapI consensus

via Uniprot.⁷ Protonation states of ionizable residues were assigned with H++, followed by system preparation with *tLeap* in AmberTools16.^{8,9} All systems were neutralized with Cl⁻ and solvated in water using a minimum distance of 12 Å between the protein surface and the edge of the box. The parameter sets employed were ff14SB,¹⁰ OL15,¹¹ YIL,¹² and TIP3P.¹³

Molecular dynamics (MD) simulations were run in the NVT ensemble at 300K after minimization (50 steps with steepest descent followed by 450 steps with conjugate gradient) and iterative thermalization (20 equally spaced stages from 10K to 300K at 12,500 steps per stage) using a Berendsen thermostat.¹⁴ Positional restraints with a force constant of 25.0 kcal mol⁻¹ Å⁻² were applied to the Zn²⁺, coordinating residues, water in the active site, and the deoxycytidine, as the active site dissociated in unrestrained simulations. The cytidine base was held in the active site with a distance restraint of 15.0 kcal mol⁻¹ Å⁻². Every system was simulated for 250 ns (in triplicate) with a 2.0 fs timestep and SHAKE for all bonds involving hydrogen atoms with the pmemd.cuda module in AMBER18 using a cutoff distance of 8.0 Å for nonbonded interactions and the smooth particle-mesh Ewald method for long-range Coulomb effects.¹⁵⁻¹⁷ Twelve systems (termed A-L) were generated by rotating the phosphate backbone of the ssDNA strand with respect to the central dC nucleotide in the active site (**Figure 1a**). The tested ssDNA sequence comprises 5'-dAdAdAdTdCdAdAdAdA-3'. All systems were built with the Modeller software.^{18,19} Input coordinates and topologies included in Supplementary Information.

The generated trajectories were analyzed with cpptraj²⁰ and an in-house FORTRAN90 program²¹⁻²⁴ for energy decomposition analysis (EDA). The EDA results suggest that system H has the most favorable total interaction energy (**Figure 1b**). The ssDNA backbone for system B is oriented in the opposite 5'-3' direction to system H and has a lower interaction energy than adjacent systems. This suggests that this orientation of the phosphate backbone allows for more favorable interactions with protein residues. Root mean squared deviation (RMSD) analysis suggests that systems B and H have the smallest deviations with respect to the original structure, suggesting a more stable initial orientation of the ssDNA (**Figure**

1c). This is further supported by an RMS fluctuation (RMSF) analysis focused on the individual nucleotides in the substrate (**Figure 1d**). The average RMSF of the substrate for each system indicates that systems B and H have the smallest average fluctuations for the ssDNA substrate (2.0 Å and 2.1 Å respectively). Analysis of residue-wise interactions with the entire ssDNA substrate suggests that systems B and H have the most favorable interactions. Loop 1 ($_{17}\text{RRLRRPYYPYPRKALL}_{30}$) shows significant favorable interactions with the substrate.

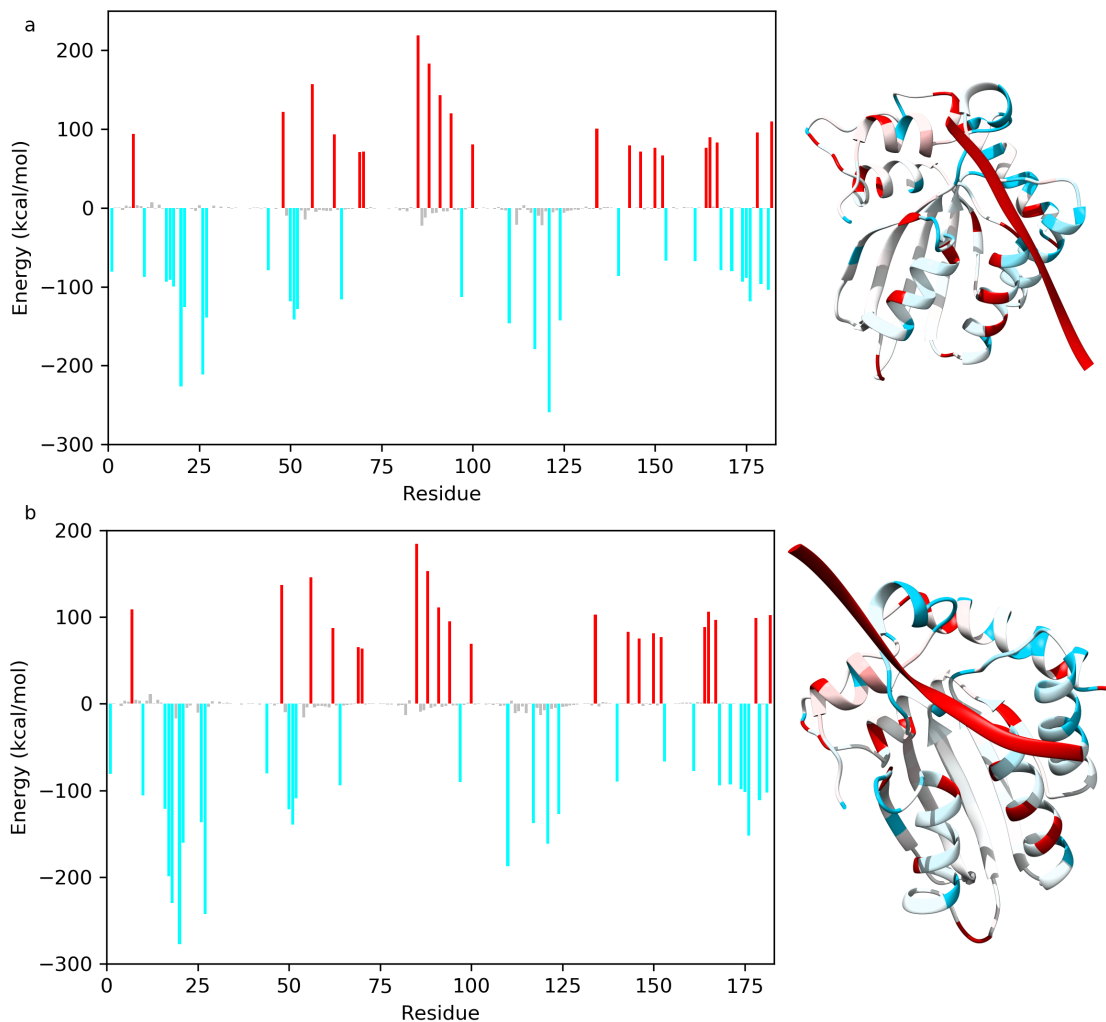


Figure 2: Non-bonded interaction energy between each residue and the ssDNA for a) System B and b) System H, with favorable(unfavorable) residue-substrate interactions in blue(red).

Systems B and H were run for an additional 250 ns from the end of the initial simulation

with the restraints on the target dC removed to observe the stability of the enzyme-substrate complex. Interatomic distances between the reacting carbon on the cytidine and the water oxygen in the active site were measured over the trajectories to determine which system maintained a more stable binding. The ssDNA substrate in System B remained bound in the active site the entire duration of the simulation (average 3.2 Å). Conversely system H exhibited sporadic loss of binding. This suggests that while the total interaction energy between the protein surface and the substrate is more favorable with system H, the local interactions with the target cytidine are more stable with system B.

Based on the above results, system B was selected for subsequent simulations to investigate the selectivity of A3H for the reported consensus sequence. Six new systems were generated by modifying the nucleotides that flank the central dC to investigate the effects of different bases in place of the preferred dTdCdA motif. Thymine recognition was tested by generating three system: 5'-dAddCdA-3', 5'-dCdCdA-3', and 5'-dGdCdA-3' substrates. Adenine recognition was tested by 5'-dTdCdC-3', 5'-dTdCdG-3', and 5'-dTdCdT-3' substrates.

EDA was performed to compare the differences (if any) in protein/substrate interaction between the 5'-dTdCdA-3' motif and all other systems. The total non-bonded interaction between the protein and the nucleotides on the 5' position suggest that dT is favored over dG (-239 kcal mol⁻¹) and dA (-38 kcal mol⁻¹), but less than dC (+81 kcal mol⁻¹). The 5' dT nucleotide shows favorable interactions with R17, R21 and R26 when compared with those from purine nucleotides, however these interactions are generally unchanged when dC is in the 5' flanking position. dT also has more favorable interactions with S86 (between -1 and -5 kcal mol⁻¹) and less favorable interactions with S87 (between +1 and +5 kcal mol⁻¹) compared with the other nucleotides (see **Figure 3b**). The total non-bonded interaction of the protein with nucleotides in the 5' position shows that dA at this position is 5 kcal mol⁻¹ less favorable than dT; dG is 61 kcal mol⁻¹ less favorable, and dC is 68 kcal mol⁻¹ less favorable.

The total non-bonded interaction between the entire ssDNA strand and the protein is

more favorable with dA in the 3' flanking position by between 31 and 245 kcal/mol compared with all other nucleotides at the same position. Our results suggest that 3'-dA shows strong attractive interactions with R17, R21, R26, K51, and K52. When compared with the other nucleotides in the same position, the arginines interact more favorably with dA by at least 5 kcal mol⁻¹ (see **Figure 3c**). These arginines are on Loop 1, which has previously been shown to be involved in DNA binding and recognition.²⁵⁻²⁷ In contrast, the lysines show slight preference for all three other bases. A3H-R26 is structurally homologous to A3A-R28 and A3B-R211, which have been previously reported as key residues that drive selectivity in the respective A3s, and both preferentially act upon a 5'-dTdC-3' substrate.^{2,28,29} The non-bonded interaction between the individual nucleotides at the 3' position and the entire protein suggest that dC and dT are less favored by 33 kcal mol⁻¹ and 55 kcal mol⁻¹ respectively, while dG shows < 1 kcal mol⁻¹ difference in interaction compared with dA. These results are consistent with previous experimental results showing similar selectivity for 5'-dTdCdA-3' and 5'-dTdCdG-3'.⁵

A3G, A3F and AID have a homologous loop, corresponding to residues 313 to 322 in A3G. Previous studies have shown that A3G_{CTD} 5'-dCdC-3' selectivity is driven by this loop and can be modified by mutating to the homologous AID or A3F sequences.^{3,30,31} Based on our results, the DNA binding orientation precludes the interaction of the ssDNA substrate with the region in A3H that is homologous to the A3G 313-322 loop. One more difference observed between A3H and other A3s relates to the structure of the bound ssDNA substrate. In other A3s it has been reported that the DNA adopts a hairpin conformation.² In A3H, the substrate can orient in a hairpin conformation in a monomeric system, however, A3H forms an RNA-mediated dimer which obstructs this orientation.

Dimer simulations were carried out to test the possibility of the B, H and hairpin ssDNA orientations in the active site. The details of the system setup and simulations are reported in Ref.³² The dimer structure shows that the active sites in each monomer are located near the RNA-mediated dimer interface of A3H. This location effectively prevents the ssDNA

substrate from adopting a hairpin conformation. When a dimer system containing the superposed hairpin structure of A3A in both active sites was considered, the simulation was unstable due to strongly repulsive interactions between the ssDNA substrates and the RNA. Conversely, a track of positively charged surface residues is observed connecting the two active sites in the A3H dimer (**Figure 3**). For the B and H systems, the initial structures were simulated with two separate strands, one in each active site. During the initial stages of the simulations, both strands were observed to come together and thus a single strand spanning both active sites through the RNA interface was also considered as shown in **Figure 3a**.

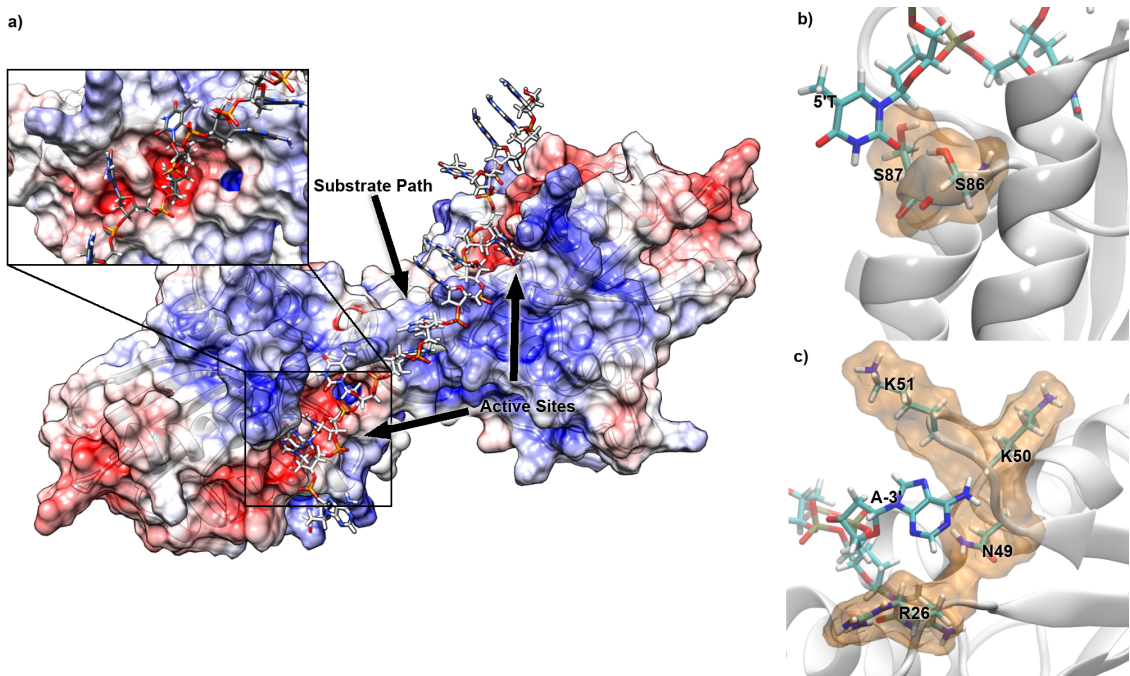


Figure 3: **a)** Electrostatic potential surface of A3H RNA-mediated dimer with ssDNA. Negative charges shown in red, positive charges shown in blue. Inset is active site with nucleophilic property and substrate cytidine in position. **b)** Residues S86 and S87 interacting with the 5' thymine. **c)** 3' adenine in a pocket formed by R26, N49, K50, and K51,

In conclusion, we have performed computational simulations to investigate the binding orientation and substrate selectivity indicators for A3H. Our results suggest that the preferred binding orientation aligns the ssDNA substrate along a track that provides favorable interactions with the target cytidine and the two flanking residues, including three arginines that significantly favor the 3' flanking adenine. Our results are consistent with previous

experimental reports on substrate binding and consensus sequence selectivity. Additionally some of the A3H residues predicted to be related to selectivity are homologous with selectivity residues reported for other A3s. These results also provide possible targets for mutagenesis to investigate the role of the selectivity filters for the consensus sequence of A3H.

Acknowledgement

This work was supported by NIH grant R01GM108583. Computational time was provided by the University of North Texas CASCaM’s CRUNTCh3 high-performance cluster partially supported by NSF grant CHE-1531468.

Supporting Information Available

This material is available free of charge via the Internet at <http://pubs.acs.org/>.

References

- (1) Liu, M.; Mallinger, A.; Tortorici, M.; Newbatt, Y.; Richards, M.; Mirza, A.; Van Montfort, R. L.; Burke, R.; Blagg, J.; Kaserer, T. Evaluation of APOBEC3B Recognition Motifs by NMR Reveals Preferred Substrates. *ACS Chemical Biology* **2018**, *13*, 2427–2432.
- (2) 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. et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nature structural & molecular biology* **2017**, *24*, 131–139.

- (3) Ebrahimi, D.; Alinejad-Rokny, H.; Davenport, M. P. Insights into the motif preference of APOBEC3 enzymes. *PLoS ONE* **2014**, *9*.
- (4) Conticello, S. G. The AID/APOBEC family of nucleic acid mutators. *Genome Biology* **2008**, *9*.
- (5) Starrett, G. J.; Luengas, E. M.; McCann, J. L.; Ebrahimi, D.; Temiz, N. A.; Love, R. P.; Feng, Y.; Adolph, M. B.; Chelico, L.; Law, E. K. et al. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nature communications* **2016**, *7*, 12918.
- (6) Ito, F.; Yang, H.; Xiao, X.; Li, S.-X.; Wolfe, A.; Zirkle, B.; Arutiunian, V.; Chen, X. S. Understanding the Structure, Multimerization, Subcellular Localization and mC Selectivity of a Genomic Mutator and Anti-HIV Factor APOBEC3H. *Scientific Reports* **2018**, *8*, 3763.
- (7) Bateman, A.; Martin, M.; O'Donovan, C.; Magrane, M.; Alpi, E.; Antunes, R.; Bely, B.; Bingley, M.; Bonilla, C.; Britto, R. et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **2017**, *45*, D158–D169.
- (8) Schafmeister, C. E. A. F.; Ross, W. S.; Romanovski, V. LEAP. 1995.
- (9) Case, D.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Greene, D.; Homeyer, N. et al. Amber16. 2017.
- (10) 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. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.
- (11) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER

- Force Fields Improves the Description of Z- and B-DNA. *Journal of Chemical Theory and Computation* **2015**, *11*, 5723–5736.
- (12) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. Reparameterization of RNA χ torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *Journal of Chemical Theory and Computation* **2010**, *6*, 1520–1531.
- (13) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.
- (14) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.
- (15) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* **2012**, *8*, 1542–1555, PMID: 22582031.
- (16) 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. *Journal of Chemical Theory and Computation* **2013**, *9*, 3878–3888, PMID: 26592383.
- (17) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593.
- (18) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **1993**, *234*, 779 – 815.

- (19) Fiser, A.; Do, R. K. G.; Šali, A. Modeling of loops in protein structures. *Protein Science* **2000**, *9*, 1753–1773.
- (20) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084–3095, PMID: 26583988.
- (21) Cisneros, G. A.; Wang, M.; Silinski, P.; Fitzgerald, M.; Yang, W. Theoretical and experimental determination on two substrates turned over by 4-oxalocrotonate tautomerase. *J. Phys. Chem. A* **2006**, *110*, 700–708.
- (22) Dewage, S. W.; Cisneros, G. A. Computational Analysis of Ammonia Transfer Along Two Intramolecular Tunnels in Staphylococcus aureus Glutamine-Dependent Amidotransferase (GatCAB). *The Journal of Physical Chemistry B* **2015**, *119*, 3669–3677, PMID: 25654336.
- (23) Elias, A. A.; Cisneros, G. A. In *Biomolecular Modelling and Simulations*; Karabancheva-Christova, T., Ed.; Advances in Protein Chemistry and Structural Biology; Academic Press, 2014; Vol. 96; pp 39 – 75.
- (24) Graham, S. E.; Syeda, F.; Cisneros, G. A. Computational Prediction of Residues Involved in Fidelity Checking for DNA Synthesis in DNA Polymerase I. *Biochemistry* **2012**, *51*, 2569–2578, PMID: 22397306.
- (25) Logue, E. C.; Bloch, N.; Dhuey, E.; Zhang, R.; Cao, P.; Herate, C.; Chauveau, L.; Hubbard, S. R.; Landau, N. R. A DNA sequence recognition loop on APOBEC3A controls substrate specificity. *PLoS ONE* **2014**, *9*.
- (26) Shi, K.; Carpenter, M. A.; Kurahashi, K.; Harris, R. S.; Aihara, H. Crystal Structure of the DNA Deaminase APOBEC3B Catalytic Domain. *The Journal of biological chemistry* **2015**, *290*, 28120–30.

- (27) Lu, X.; Zhang, T.; Xu, Z.; Liu, S.; Zhao, B.; Lan, W.; Wang, C.; Ding, J.; Cao, C. Crystal structure of DNA cytidine deaminase APOBEC3G catalytic deamination domain suggests a binding mode of full-length enzyme to single-stranded DNA. *The Journal of biological chemistry* **2015**, *290*, 4010–21.
- (28) 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. *Journal of Chemical Theory and Computation* **2019**, *15*, 637–647.
- (29) 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. et al. The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. *Journal of Molecular Biology* **2013**, *425*, 4442–4454.
- (30) 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.; Stivers, J. T. Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *Journal of Biological Chemistry* **2010**, *285*, 40956–40964.
- (31) Carpenter, M. A.; Rajagurubandara, E.; Wijesinghe, P.; Bhagwat, A. S. Determinants of sequence-specificity within human AID and APOBEC3G. *DNA Repair* **2010**, *9*, 579–587.
- (32) Hix, M. A.; Wong, L.; Flath, B.; Chelico, L.; Cisneros, G. A. Induced mutagenesis by the DNA cytosine deaminase APOBEC3H Haplotype I protects against lung cancer. <https://www.biorxiv.org/content/10.1101/2020.02.28.970509v1> **2020**,

Graphical TOC Entry

