

Article

Coevolution of Lentiviral Vif with Host A3F and A3G: Insights from Computational Modelling and Ancestral Sequence Reconstruction

David Nicolas Giuseppe Huebert ^{1,2}, Atefeh Ghorbani ¹, Shaw Yick Brian Lam ² and Mani Larijani ^{1,2,*}

¹ Immunology and Infectious Diseases Program, Division of Biomedical Sciences, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada; david_huebert@sfu.ca (D.N.G.H.); atefeh_ghorbani@sfu.ca (A.G.)

² Structural Biology and Immunology Program, Department of Molecular Biology and Biochemistry, Faculty of Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; brian_lam_2@sfu.ca

* Correspondence: mlarijani@sfu.ca; Tel.: +1-(778)-782-7276

Abstract: The evolutionary arms race between host restriction factors and viral antagonists provides crucial insights into immune system evolution and viral adaptation. This study investigates the structural and evolutionary dynamics of the double-domain restriction factors A3F and A3G and their viral inhibitor, Vif, across diverse primate species. By constructing 3D structural homology models and integrating ancestral sequence reconstruction (ASR), we identified patterns of sequence diversity, structural conservation, and functional adaptation. Inactive CD1 (Catalytic Domain 1) domains displayed greater sequence diversity and more positive surface charges than active CD2 domains, aiding nucleotide chain binding and intersegmental transfer. Despite variability, the CD2 DNA-binding grooves remained structurally consistent with conserved residues maintaining critical functions. A3F and A3G diverged in loop 7' interaction strategies, utilising distinct molecular interactions to facilitate their roles. Vif exhibited charge variation linked to host species, reflecting its coevolution with A3 proteins. These findings illuminate how structural adaptations and charge dynamics enable both restriction factors and their viral antagonists to adapt to selective pressures. Our results emphasize the importance of studying structural evolution in host–virus interactions, with implications for understanding immune defense mechanisms, zoonotic risks, and viral evolution. This work establishes a foundation for further exploration of restriction factor diversity and coevolution across species.

Keywords: Apobec3; Vif; lentiviruses; host–viral coevolution; viral antagonists; deamination; computational structure



Academic Editor: Alexander Gorbalenya

Received: 31 December 2024

Revised: 3 March 2025

Accepted: 5 March 2025

Published: 10 March 2025

Citation: Huebert, D.N.G.; Ghorbani, A.; Lam, S.Y.B.; Larijani, M.

Coevolution of Lentiviral Vif with Host A3F and A3G: Insights from Computational Modelling and Ancestral Sequence Reconstruction. *Viruses* **2025**, *17*, 393. <https://doi.org/10.3390/v17030393>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Across species, the dynamic interactions between viruses and host organisms represent an ongoing evolutionary arms race, shaped by long periods of host–virus coevolution. This intricate relationship is particularly well studied in retroviruses where host restriction factors, such as those in primates, evolve to inhibit viral elements, while lentiviral proteins adapt to circumvent host defenses [1–3]. A prominent example of this interplay is the interaction between the host mutational enzyme apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3 (APOBEC3 or A3) and the lentiviral viral infectivity factor (Vif) [4]. APOBEC3 proteins act as restriction factors that induce mutations in viral genomes, while Vif counteracts their effects by targeting APOBEC3 proteins for degradation through the ubiquitination pathway. Lentiviruses are found in a wide range of animals,

including ruminants, equines, felines, and primates, though notably absent in New World monkeys (NWMs) [5]. The zoonotic transmission of these viruses from one host species to another involves three critical steps: (1) exposure to the virus, (2) spillover into a new host species, and (3) establishment and spread within the new host population [6]. Successful transmission depends on the virus's ability to recognize host cell receptors for entry and evade host innate restriction factors. Human immunodeficiency virus (HIV) exemplifies this process. HIV-1 originated from simian immunodeficiency virus (SIV) strains found in chimpanzees (*Pan troglodytes*) and gorillas (*Gorilla gorilla*), while HIV-2 originated from SIV strains infecting sooty mangabeys (*Cercocebus atys*). While SIV has crossed into humans on at least 11 occasions, only since the 1970s has it given rise to the epidemic forms of HIV [7–11]. In contrast, SIV has been infecting great apes (GAs) and Old World monkeys (OWMs) for at least 30,000 years, resulting in diverse clinical outcomes depending on the host species [12]. This evolutionary backdrop provides a unique framework to study the cellular and molecular interactions that govern host–virus dynamics.

HIV primarily infects immune cells, including T helper cells, macrophages, and certain dendritic cells [13–18]. During viral replication, APOBEC3 proteins from the previous host are incorporated into budding virions. Once the virus enters a new host cell, it encounters a range of restriction factors acting at different stages of the viral life cycle. These include tetherin, which traps virions at the cell surface [19–21]; SERINC3 and SERINC5, which alter viral envelope rigidity [22–25]; TRIM5 α , which causes premature uncoating of the viral genome [26–28]; and SAMHD1, which depletes deoxynucleotide pools necessary for reverse transcription [29–31]. APOBEC3 proteins, in particular, deaminate deoxycytidine (dC) residues in single-stranded DNA (ssDNA) generated during reverse transcription, converting them into deoxyuridine (dU) [4,13–18,32–34]. This mutagenesis can result in either complete inactivation of the viral genome or mutations that enable the virus to evade host innate and adaptive immune detection [35–37]. Despite these formidable defenses, viruses have evolved specific mechanisms to counteract specific restriction factors [4,19–34]. Viral Vif, for instance, binds to APOBEC3 proteins in conjunction with host CBF- β , recruiting an E3 ubiquitin ligase complex to polyubiquitinate APOBEC3 proteins, targeting them for degradation before they can incorporate into the virion [35–37]. While APOBEC3 proteins are most studied in the context of retrotransposons and retroviruses, they also exhibit antiviral activity against DNA viruses such as herpes simplex virus (HSV), hepatitis B virus (HBV), and parvoviruses [38]. Additionally, there is evidence suggesting that APOBEC3 proteins may restrict certain RNA viruses, including coronaviruses, though their precise mechanisms remain unclear. Collectively, these restriction factors act as barriers to zoonotic transmission, which viruses must overcome to establish infection in new hosts [6,8].

For lentiviruses like HIV and SIV to successfully adapt to a new host species, they must evade the host's unique restriction barriers, such as APOBEC3 proteins [6]. In primates, the APOBEC3 family underwent remarkable diversification through gene duplication, resulting in seven paralogs (A3A–H) with some containing two fused domains [39–43]. This diversification is thought to have been driven by evolutionary pressures from retroelements and potentially retroviruses. APOBEC3 genes originally evolved from activation-induced cytidine deaminase (AID), an enzyme involved in antibody diversification in jawed vertebrates [44–48]. The initial diversification of APOBEC3 genes occurred approximately 170 million years ago in placental mammals, giving rise to three ancestral subtypes (A3Z1, A3Z2, and A3Z3). Positive selection across taxa has driven further divergence of APOBEC3 genes, often in response to viral pressures, leading to corresponding evolutionary changes in lentiviral Vif [48–54]. The expansion of APOBEC3 proteins, particularly the A3A–H paralogs [15,40,55–58], is hypothesized to be a result of this coevolutionary arms race with retrotransposons and retroviruses [5,55,59–62]. Some APOBEC3 proteins, notably A3A,

A3C, and A3H, retain single deaminase domains similar to their ancestral counterparts, while others, such as A3B, A3D, A3F, and A3G, have evolved double domains [34,63–65]. In these double-domain APOBEC3 proteins, the N-terminal domain (CD1) is typically inactive but suggested to play a crucial role in substrate binding, three-dimensional diffusion, facilitating ssDNA intersegmental transfer, and binding ssRNA to enter budding virions, while the catalytically active C-terminal domain (CD2) acts through direct mutation. These structural adaptations enhance their antiviral activity, particularly against retroviruses. Among the APOBEC3 family, A3D, A3F, A3G, and A3H exhibit the most potent antiviral effects and are common targets of lentiviral Vif along with A3C [6,32]. Furthermore, A3C, A3D, and A3F are all A3Z2s that maintain a similar Vif-binding site, while A3H has a similar one offset by one α -helix, and A3G-CD1 has a unique binding site that, unlike the others, was recently found to require a nucleic acid intermediate [4,66,67].

The mechanism of APOBEC3-mediated restriction involves binding viral genomic RNA or ssDNA, inhibiting reverse transcription, and introducing mutations into proviral DNA [68]. Each APOBEC3 protein exhibits sequence specificity in its deaminase activity. For example, A3F prefers to mutate *TTC* motifs, while A3G targets *CCC* motifs [69]. The extent and location of these mutations can determine their effects, ranging from complete viral inactivation to the generation of escape mutations that confer drug resistance or immune evasion [70,71]. To counteract these effects, Vif interacts with APOBEC3 proteins to either slow their catalytic rate or promote their degradation prior to viral budding where APOBEC requires binding of ssRNA [5,17,58,72,73]. Importantly, Vif exhibits species specificity, being most effective against APOBEC3 proteins of its native host which underscores the coevolutionary dynamics between host and virus [74,75]. Thus, in addition to a race against both endogenous and zoonotic retroviruses, this adversarial relationship with Vif is believed to have contributed to the rapid evolution and diversification of A3 enzymes in primates [62]. Structurally, APOBEC3 proteins possess a catalytic pocket containing a zinc ion (Zn^{2+}) coordinated by a histidine and two cysteine residues with a catalytic glutamate facilitating the deamination reaction. Although the CD1 domain is inactive, it retains the structural components necessary for catalysis and often oligomerizes with other APOBEC3 proteins to enhance DNA editing efficiency through processivity [69,76–89]. Four loops ($\ell 1'$, $\ell 3'$, $\ell 5'$, and $\ell 7'$) in CD2 form a U-shaped DNA-binding groove with the catalytic pocket positioned at its center with a gatekeeper residue overtop that is able to both occlude or stabilize the entry of DNA in the active CD2 domains [80,82,88–90]. Conversely, these loops close around the pocket in A3G-CD1 [84]. Loop 7, in particular, determines sequence specificity through its unique composition [69,76–89]. For instance, A3F contains hydrophobic residues (YYFW), while A3G has hydrophilic residues (YDDQ). The structural arrangement of APOBEC3 domains can vary between globular and dumbbell-shaped configurations, influencing their interaction with nucleic acids and other proteins [91].

The dynamic interplay between host APOBEC3 proteins, viral Vif, and viral nucleic acids highlights the evolutionary arms race that shape host–virus interactions relative to both zoonotic and the following intraspecies transmissions. These coevolutionary pressures have driven the diversification of APOBEC3 proteins in primates and lentiviral adaptations to overcome host defenses. By utilising multifaceted computational approaches to examine the structural and evolutionary variations in A3F, A3G, and Vif across primates and their lentiviruses, we can gain deeper insights into the mechanisms underlying viral zoonosis and host immunity [92]. Although structural computational approaches are only one way of expanding on this diversification, it allows us to observe interactions that have yet to be resolved through more traditional means. Computational approaches, such as those used by King et al., which were solved 2 years prior to the X-ray crystal structure, have proven valuable in predicting protein structures and interactions, offering a

framework for studying these complex interactions [93–95]. This paper aims to extend our understanding of APOBEC3 proteins and their role in shaping the evolutionary dynamics of host–virus relationships.

2. Materials and Methods

2.1. Acquisition and Alignment of Sequences

We attempted to find all the sequences for all primate A3F and A3G, and all of the sequences of Vif across HIV and SIV (Figure 1). Sequences for all species were attained from NCBI or Ensembl [96,97]. Look at Tables S1–S3 for the individual sources. Alignments were created using Strap bioinformatics suite and WebLogo [98–101].

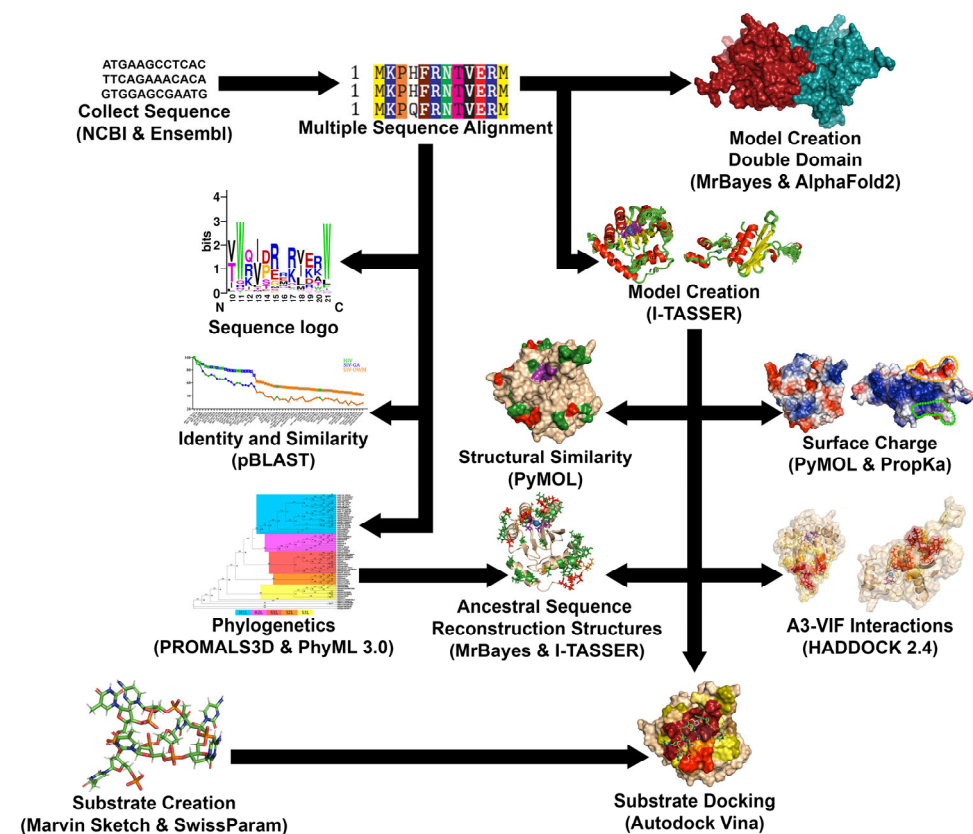


Figure 1. Computational workflow of primate A3FG and HIV/SIV Vif.

2.2. Production of Similarity Graphs and Phylogenetic Trees

Graphs of identity and similarity were created utilising NCBI BLASTp under default conditions (as they were sufficient for pair alignment) (Figure 1) [102]. This was done by comparing human genes and HIV-1M NL4-3 genes to all other species and collecting data on the identity, similarity, and gaps in the sequence. This was performed on the full proteins as well as on the domains which had a cutoff for human A3F, which was CD1 M1 to I187 and CD2 L188 to E373, while human A3G was CD1 M1 to I192 and CD2 L193 to N384, and the cutoffs were maintained across other species.

The phylogenetic trees were created utilising the amino acid sequences. Firstly, a multiple sequence alignment was created using PROMALS3D to account for not only the sequence but the 3D structures to which they are related [103]. This was performed under default conditions utilising the sequences of A3F, A3G, or Vif we had found as well as the known protein crystal structures of APOBECs or Vif to aid the construction of the alignment. A3G utilised structures PDB:5K81 [84] and 3E1U [85] for CD1 and CD2, respectively. A3F utilised 2 structures, CD2 structures PDB:3WUS [81] and 5HX5 [104],

as no structures exist of A3F-CD1. Vif utilised 2 separate structures, PDB:4N9F [105] and 6P59 [106]. The alignment was then used in the program TranslatorX to create this alignment again in the context of nucleic acids as opposed to the original amino acid alignment (<http://translatorx.co.uk>, accessed on 8 August 2018) [107]. This alignment was then put through the PhyML SMS suite under default conditions to choose a substitution model to calculate the likelihood of relationship appropriate to each protein [108]. The substitution models chosen were JTT [109], HIVw [110], and GTR [111] for A3F, A3G, and Vif, respectively. Using these amino acid substitution models, the trees were constructed through default conditions of the PhyML 3.0 suite utilising maximum likelihood [112].

2.3. Creation of 3D Structural Homology Models

Creation of protein structures of AID/APOBECs through X-ray crystallography and NMR has been generally difficult due to a number of reasons including their genotoxicity, high charged surface area, need for truncations, and mutations for crystallisation. As such, 3D structural homology models were created for each protein utilising Swissmodel to aid in finding proper structural templates as well as create 3D structural homology models to look at the possible pocket conformations (Figure 1) [113–117]. The I-TASSER suite (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>, accessed on 8 December 2016) was then used under default conditions to create high quality 3D structural homology models of each species' A3FG domains as well as Vif across viruses [118–122]. Many templates were utilised for 3D structural homology model creation, with A3F-CD1 3D structural homology models being created from templates 3VOW, 5K81, 5K82, 5K83, and 3VM8 [84,123,124], and A3F-CD2 models from templates PDB:4IOU, 3WUS, 5HX5, 5W2M, and 5ZVB [79,81,82,104,125]. The templates of PDB:5K81, 5K82, and 5K83 [84] were used for A3G-CD1, and 3E1U, 3IR2, 3V4J, 4ROW, and 4ROV [85,86,88,126] for A3G-CD2. To form structures of each A3 catalytic domain, independent domain cut-offs were noted through the context of humans as A3F-CD1 M1 to I187, and A3F-CD2 L188 to E373, in A3G-CD1 M1 to I192 and A3G-CD2 L193 to N384, and the cut-offs were maintained across other species.

Vif 3D structural homology models were created utilising templates PDB:4N9F [105] and 6NIL [127]. Vif 3D structural homology models were taken from M1 to D172 in the context of HIV-1M NL4-3, and this was maintained for 3D location based on the cut-off of current structures at these locations.

Structural templates of all structures were obtained through the protein data bank (<http://www.rcsb.org>, accessed on 30 September 2016), and all templates and 3D structural homology models were viewed through PyMOL v1.7.6 (<http://www.pymol.org/>, accessed on 15 September 2016) [128,129]. These models were chosen with Zn in the pocket of A3s and Vif's HCCH Zn-binding site in HIV/SIV, which were put in place with I-TASSER's protein–ligand-binding site program COACH [130,131]. These 3D structural homology models were checked for their Cscore (confidence of structure based on threaded template alignments), TM-Score (compared to known structure but insensitive to local error), and RMSD (compared to known structure but sensitive to local error), which were calculated for confidence and similarity with known structures through the I-TASSER Suites (Tables S4–S12). We also constructed Ramachandran plots for each 3D structural homology model to see the ratios of favoured, allowed, and outlier residues where these models maintained greater than 96% favourable results and allowed amino acid angles [132].

A3F and A3G CD2 models were also chosen for having an open catalytic pocket for later DNA docking experiments. Representative species A3s were selected for later dockings so that there would be a human and a representative of each taxon. We then chose the species for having known A3F and A3G sequences and for having known HIV/SIV

Vif antagonists, with the exception of the NWMs that have no lentiviral antagonists, so we chose the 2 most diverse species from humans. These representatives were *Pan troglodytes* for GAs and *Papio Anubis* for OWMs, and for A3F *Saimiri boliviensis boliviensis* and A3G, and *Sanguinus labiatus* for NWMs.

The creation of double-domain 3D structural homology models could later be attained through the use of AlphaFold2 after its creation, utilising the monomer_casp14 default conditions with gpu relaxation and all templates up to 2022 on Compute Canada's Narval server [133–135].

2.4. Analysis of Structural Surfaces and Charge

Due to variations in the identity and similarity of these proteins and the integral nature of charge in both DNA and inhibitor binding, we set out to determine the charges of these proteins (Figure 1). Analyses of the charges of these proteins were obtained utilising two methods. Firstly, charges of individual protein 3D structural homology models were created using PropKa 3.0 on the PDB2PQR Server (<https://server.poissonboltzmann.org/>, accessed on 10 September 2018) with CHARMM forcefield at pH7 to ascertain residue protonation states and, from this, calculate the total unfolded and folded charges of the proteins at any pH [136–139]. Thus, the program was able to determine the protonation state of any individual amino acid based on its interactions with the residues or solvent exposure around it using the charges and atom sizes of the CHARMM forcefield as reference. We tabulated the 4 3D structural homology models (created via I-TASSER) of each domain across all the species at pH 7 to understand the charges. Secondly, PyMOL v1.7.6 vacuum electrostatics were used to visualize the charge by their regions.

2.5. Protein-Protein Docking

For an attempt to collate data on the binding sites between Vif, A3F, and A3G, we utilised the HADDOCK2.4 servers under default conditions (Figure 1) [140,141]. This process began through random rotations of the starting structures and rigid body energy minimization, semi-rigid annealing in torsion angle space, and final refinement in Cartesian space with explicit solvent. However, we added further restraints by selecting specific residues of each of the proteins that are known binding sites. For the sites of binding, we utilised those highlighted by Delviks-Frankenberry et al. 2020 for A3G-CD1 and A3F-CD2, and Wang et al. 2018 for the Vif-binding sites [4,41]. These highlighted regions were then chosen again across primates' A3FGs and HIV/SIV Vif based on their 3D location of amino acids in comparison to the original. Default conditions were the most appropriate for docking as it allowed for an efficient use of resources while also allowing the factoring in of course and fine grain interactions. Furthermore, the parameters allow for a rigid interface to become a semi-flexible interface where the interacting residues (within 5 Å) are more motile while still retaining their core structure to prevent deformation. To aid in this, all hydrogens were removed and then replaced at the end of the process. From an initial 1000 complexes, the ten to fourteen bindings with the lowest energy profiles were chosen, all of which had at least 30 Å² of surface. This is produced through multiple refinements and molecular dynamics steps using OPALs force field to ascertain the positions of the flexible residues, with a final resolution utilising TIP3P water models for final refinements. From this point, PyMOL 1.7.6 was utilised to highlight any amino acid between A3F or A3G and Vif by selecting all amino acids within 4 Å of each other and then these data were collated into charts, graphs, and 3D coloured spectrum analyses of these surfaces.

2.6. Creation of DNA Substrates

Docking substrates were created as 7mer nucleotide sequences similar to our previous studies (Figure 1) [142,143]. They were constructed with a 3-base motif in the centre of

the sequence as well as repeating nucleotides on either end to give added stability. The substrates created were *taTTCat* for A3F and *ttCCctt* for A3G. These were chosen since they are well documented as the preferred substrate motifs in humans [69,80,144–147]. Together, we created both these preferred substrate motifs of A3FG as well as two residues on either side, extending not only the stability of the sequence but also the length of the DNA to better extend down the DNA-binding groove during later docking protocols. The creation of substrates was initiated using ChemDraw Prime v.16.0 (<https://perkinelmer-chemdraw-prime.software.informer.com/16.0/>, accessed on 19 December 2016) to make the 2D version of the substrate and then Marvin Sketch v.5.11.5 (<http://www.chemaxon.com/products/marvin/marvinsketch/>, accessed on 10 January, 2017) to assemble it into 3D. From this point, SwissParam (<http://swissparam.ch>, accessed on 1 November 2018) was used to generate the docking parameters through CHARMM22 force field to account for electrostatic and VDW potentials [148].

2.7. Analysis of DNA-Protein Docking Results

After the creation of both the proteins and substrates, four open 3D structural homology models of A3F-CD2, A3G-CD2 for humans and a representative of each primate taxon were chosen (Figure 1). These representatives were *Pan troglodytes* for GAs and *Papio Anubis* for OWMs, and for A3F *Saimiri boliviensis boliviensis* and A3G, *Sanguinus labiatus* for NWMs. These were chosen due to the ability to obtain representatives of the same species (except in NWMs) that, across OWMs and NWMs, were distinct with numerous adaptations compared to humans, while in GAs we wanted a species as close to humans as possible. This research was enabled in part by support provided by sharcnet's graham cluster of the University of Waterloo (www.sharcnet.ca, accessed on 17 May, 2019) and Compute Canada (www.computecanada.ca, accessed on 17 May 2019). We utilised AutoDock Vina (<https://ccsb.scripps.edu/>, accessed on 17 May 2019) to ascertain the most energetically favourable conformations of the substrate across the protein's surface using default parameters (exhaustiveness 8 and energy range 3 kcal/mol) with random seeds utilising its Monte Carlo-based search algorithms [149,150]. This was performed in a 30 Å³ box around the α -carbon of the tryptophan of loop 6 in the centre of the substrate groove's face of the protein, adjacent to the catalytic pocket. This permits the entire substrate groove surface to be interacted with by the DNA docking protocols. For each 3D structural homology model, 11 docking modes were produced per run, and 30 runs were completed in total, leading to 330 docking modes per model. UCSF chimera v.1.11.2 (<https://www.cgl.ucsf.edu/chimera>, accessed on 15 June 2019) was used to view the conformations of the substrate and its interactions with A3F and A3G [151]. At this point, we tabulated whether the DNA entered the catalytic pocket or not and which nucleobase or portion of backbone entered the catalytic pocket. The data were compiled by grouping the runs in subsets of 6 compiled with all the other models of the species and then averaging across 5 sets per species to gain statistics of which nucleosides fell in the pocket. At this point, all conformations that had the cytidine in the 5th positions within the catalytic pocket were compiled into a list and made into pdb files and placed in PyMOL to ascertain all amino acid residues with which they interacted within 4 Å. This was then compiled for each species based on each amino acid to determine to what extent each secondary structure and residue affected the binding of DNA when cytidine was in the catalytic pocket. This was then able to be used for tabulation and 3D heatmaps.

2.8. Ancestral Sequence Reconstruction

For the creation of ancestral sequences, we had already selected our extant species, created a multiple sequence alignment utilising amino acids based on protein structural

data, and created a phylogenetic tree (Figure 1). To understand how these proteins evolved, we also required a multiple sequence alignment of the nucleotides and as such used TranslatorX to create the alignment again in the context of nucleic acids (<http://translatorx.co.uk>, accessed on 8 August 2018) [107]. We utilised MrBayes/2.3.7 employing Markov chain Monte Carlo (MCMC) and used the phylogeny generated prior as a guide tree, as well as the nucleotide alignment, to form our ancestral calculations [152–155]. The reason why we chose to use this program is because with Bayesian inferences we were able to determine, based on the known relationships of primate species, the probability of any potential extinct organisms' sequences being a certain residue. We utilised the GTR substitution model where the nucleotide positions of the codons were unlinked, and each run was continued until the standard deviation of split frequencies (estimated marginal posterior probability of a phylogeny splitting into nodes) of <0.01 was realised with a potential scale reduction factor (estimated factor which can reduce the scale of the distribution if simulations were allowed infinite number of iterations) for all parameters of around 1.0. With our phylogenetic guide tree, we chose to calculate dependent on the three nodes from which all great apes, Old World monkeys, and New World monkeys, respectively, descended. This analysis was conducted for each ancestral group with eight nchains (Markov chains) and nrns three times and combined as one phylogeny to create the final result. After the creation of the phylogenies for ancestral sequence reconstruction (ASR), the number of possible sequences with nonsynonymous effects per domain was assessed with the aid of MrBayes Helper (<https://seyedjavad.shinyapps.io/MrBayes-Helper/>, accessed on 17 March 2022). Three-dimensional structural homology models were created for these possibilities by utilising I-TASSER servers; however, if greater than 20 sequences were available, only the top 20 were subjected to 3D structural homology modelling [118–120]. After this, they were placed through the same processes and placed through PropKa [130–132,136–139].

3. Results

3.1. Evolution of A3F, A3G, and Vif

3.1.1. Vif Contains More Variation than A3FGs in Its Binding Regions

Both A3FGs and Vif have evolved rapidly due to the ongoing host–pathogen race with viruses [6,49,156]. The location of Vif binding on A3FGs varies. While both A3G and A3F have an inactive CD1 and active CD2, Vif binds A3G on the inactive CD1 at the junction of loop 7 ($\ell 7$) and $\alpha 4$ [4]. In contrast, Vif binds A3F on the active CD2 across the C-term of $\alpha 2$, into $\ell 4$, and in the middle of $\alpha 3$. Accommodating these diverse binding surfaces, Vif has evolved separate binding surfaces corresponding to A3F and A3G [41]. HIV-1 Vif binds A3F via three disparate regions that come together in 3D, whereas one linear motif constitutes the region that binds A3G. Based on this, we were intrigued to see how these sequences compare across primates and lentiviruses.

Analysing A3F-Vif-binding interactions, we note a host of predicted adaptations amongst both the host and viral proteins. Human A3F-CD2 residues L255, 260DD261, P265, and E289 interact with HIV-1M NL4-3 Vif residues in three binding regions: (1) 14DRMR17, (2) G71, 73TGERDW79, and (3) 171EDRWNK176. Across primates, we notice that positions 255 and 265 are fully conserved, position 289 is highly conserved, while positions 260–261 are variable (Figure 2A). However, when looking to HIV and SIV, we note diversity in A3F-binding regions across residues 14–17, 71, 74–79, and 171–176 (Figure 2B). Of these, the most conserved regions are residues 17 (R/K); 74 (T/P); 75 (G/P); 76 (E); and 77 (R/K). In comparison, very few similarities are seen throughout the third binding region of Vif. Intriguingly, when comparing HIV-1 and HIV-2 lineages, all retain similar motifs across residues 17 and 75–79 corresponding to region 1 and 2, respectively, but not in the third binding region (Figure S1A–C). Thus, when examining the A3F-Vif-binding interactions

across species, we note the largest variability localized within Vif, suggesting multiple attempts to circumvent A3s of varying species.

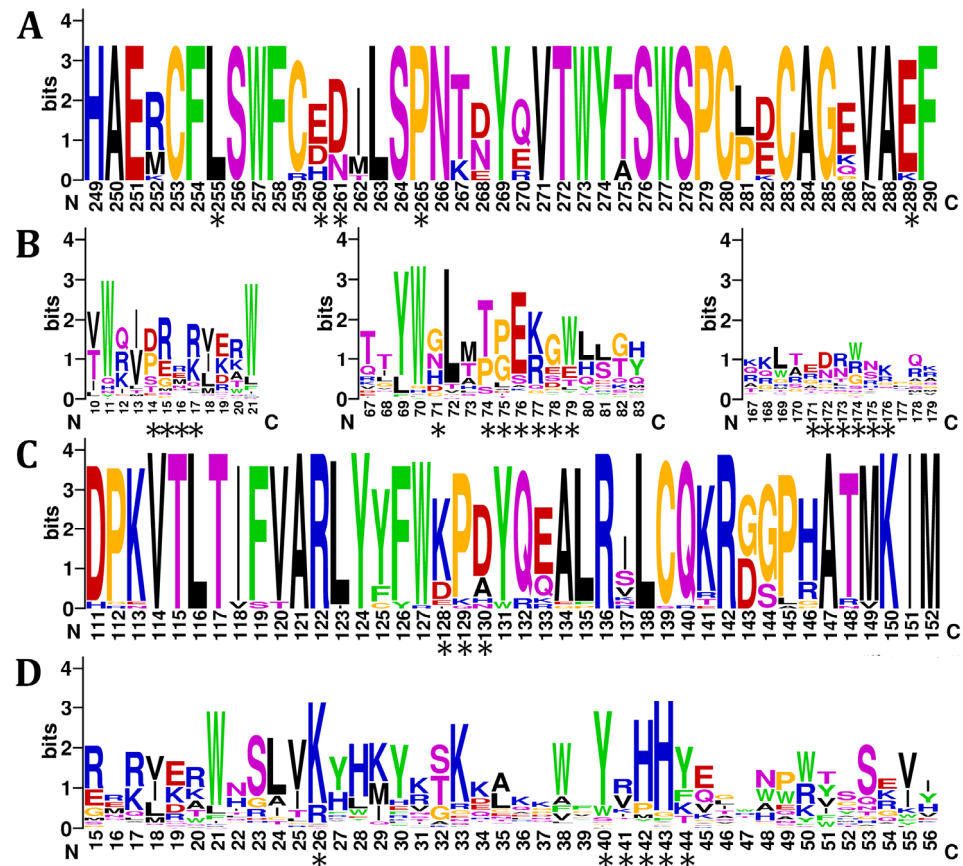


Figure 2. Amino acid sequence logos. All sequences are shown here with heights based on their conservation across alignments of all species across different proteins. Asterisks denote known residues that interact with their targets with either an APOBEC [4] or with Vif [41], respectively. The sequences are A3F denoting Vif-binding sites (A), three charts of Vif with A3G-binding sites (B), A3G with Vif-binding sites (C), and Vif with A3G-binding sites (D). Figure created in WebLogo.

Adaptations can also be seen within the A3G-Vif-binding interactions. Human A3G-CD1 residues from 128–130 (128DPD130) binds HIV-1M NL4-3 Vif residues K26 and 40YRHHY44. Across all primate A3G-CD1, the most common motif is 128KPD130; however, the most common sequence among GA is 128DPD130 (Figure 2C). Across primate Vif, we see high levels of positive residues at residue 26, aromatics at 40 and 44, and mainly histidine at 42 and 43, with other amino acids also seen at lower levels (Figure 2D). When comparing HIV-1 and HIV-2 lineages, we note residues 26, 40, 42, and 43 remain constant, while residues 41 and 44 are quite variable (positive residue/V and Y/K, respectively) (Figure S1D–F). This suggests that there is more consensus in binding of A3G among HIVs than binding of A3F.

3.1.2. Rapid Genetic Evolution of Vif and A3FGs

To explore the evolutionary relationship between Vif and A3FG proteins across species, we analysed their sequence variations. We hypothesized that Vif genes exhibit greater diversity than A3 genes due to their direct targeting by A3s, the error-prone nature of reverse transcription, and selective pressures to overcome host barriers [6,157]. Our analysis revealed that Vif genes vary significantly with similarity to HIV-1M NL4-3 ranging from 94.3% to 73.1% in HIV-1 and SIV of GA, while other lineages, including HIV-2, showed 61.6% to 41.3% similarity (Figure 3A). *Cercopithecus atys* Vif genes displayed 54.3% to 48.4%

similarity. This variability extends even within Vif genes of viruses infecting the same species. Comparatively, Vif evolution aligns with patterns observed in Env (Figure S2C) and other restriction factor antagonists (Figure S2D–G) but contrasts with the higher conservation seen in Gag and Pol genes (Figure S2A,B). This suggests that while many lentiviral genes, particularly restriction factor antagonists, have undergone significant adaptation, Gag and Pol remain more constrained over 1 million years of evolution [49].

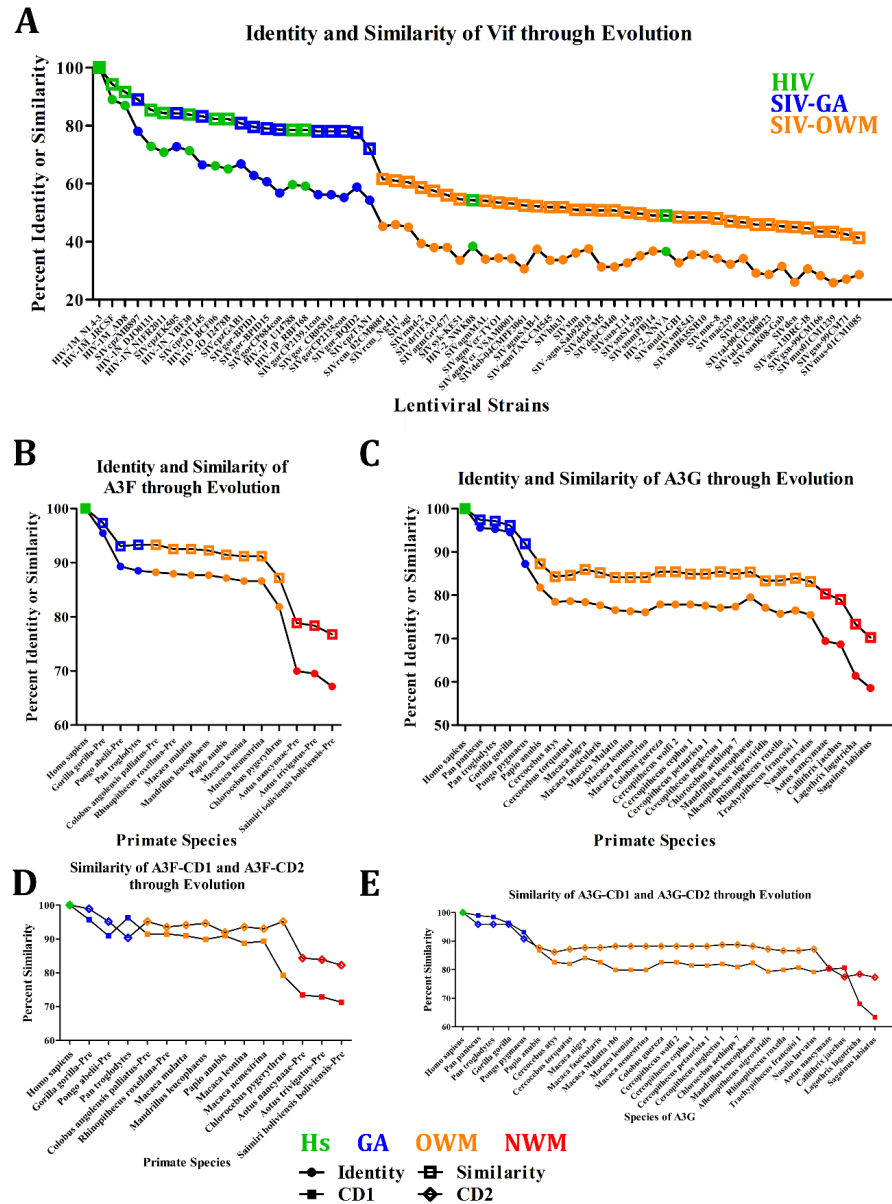


Figure 3. Identity and similarity of Vif and APOBEC proteins. Shown here are the identity and similarities of the amino acid sequences across species of Vif (A), A3F (B), and A3G (C), as well as similarities of A3F (D) and A3G (E) dependent on domain compared to HIV-1M NL4-3 and *H. sapiens*, respectively. The colour scheme of A3FGs is based on their taxa, while Vif is based on the taxa that is infected, where green denotes humans, blue denotes great apes, orange denotes Old World monkeys, and red denotes New World monkeys.

In contrast, A3 genes, particularly A3F and A3G, exhibit less diversity than Vif but still show notable variation across primates. A3F similarity decreases across taxa: 97.3% to 93.3% in GA, 93.3% to 87.2% in Old World monkeys (OWMs), and 78.8% to 76.7% in New World monkeys (NWMs) (Figure 3B). Similarly, A3G exhibits declining similarity: 97.4% to 91.9% in GAs, 87.2% to 83.2% in OWMs, and 80.3% to 79.2% in NWMs (Figure 3C). Comparable

levels of divergence are seen in tetherin (Figure S3B) and TRIM5 α (Figure S3D). These levels of divergence are greater than those observed in related immune proteins like AID, which only shows a decline from 99.5% to 97.5% across all taxa (Figure S3A). Similarly, there are low divergence proteins like SamHD1 (Figure S3C) and toll-like receptors (Figure S3E–I). This suggests that some restriction factors have diverged more over 33 million years than other immune proteins have over ~400 million years [49,158].

We then wondered if differing evolutionary pressures existed on the catalytic domains (CD1 and CD2) of A3F and A3G. For A3F, CD1 showed greater divergence than CD2 across taxa: similarity declined from 96.3% to 90.9% in GAs, 91.4% to 79.3% in OWMs, and 73.4% to 71.3% in NWMs (Figure 3D). In contrast, A3F-CD2 similarity ranged from 98.9% to 90.3% in GAs, 95.2% to 91.9% in OWMs, and 84.3% to 82.3% in NWMs. Interestingly, in some cases, such as *Pan troglodytes*, CD2 displayed greater divergence than CD1, hinting at species-specific evolutionary pressures. For A3G, CD1 generally showed greater variability, with similarity declining from 98.9% to 93.1% in GAs, 86.8% to 79.2% in OWMs, and 80.6% to 63.4% in NWMs (Figure 3E). Conversely, A3G-CD2 remained more conserved: 95.9% to 90.8% in GAs, 88.7% to 86.2% in OWMs, and 80.5% to 77.3% in NWMs. This divergence in A3G-CD1 is particularly pronounced in species like *Lagothrix lagotricha* and *Saguinus labiatus*, while GAs and *Callithrix jacchus* had greater diversity in CD2. Despite these divergences, the catalytic residues in both A3F and A3G remain 100% conserved across species, underscoring the importance of preserving enzymatic function. The observed differences suggest that non-catalytic CD1 domains, being inactive, are more prone to evolutionary adaptation than CD2, which plays a catalytic role. Notably, A3F-CD2, targeted by Vif, is less affected by evolutionary pressures than CD1. Conversely, in A3G, although CD1 is the primary target of Vif, CD2 exhibits greater diversity in GAs, reflecting selective pressures that extend beyond Vif interaction alone.

When comparing A3F and A3G, their CD1 domains exhibit similar levels of divergence across most taxa, except for highly divergent NWM species. CD2 domains, however, differ between the two proteins, with A3G-CD2 displaying greater divergence than A3F-CD2 in OWM and NWM. This divergence may relate to their substrate specificities: A3G uniquely targets CCC motifs, while A3F acts on TTC motifs, aligning with other APOBEC3 proteins like A3A, A3B, A3C, A3D, and A3H [69,144,145,159–162]. Together, these findings highlight how selective pressures shape A3 gene evolution differently across domains and species. The non-catalytic CD1 domains, being less constrained by functional requirements, are more susceptible to mutations and adaptations. However, catalytic CD2 domains also experience specific pressures, particularly in A3G, where unique interactions with viral and host factors drive greater diversification. These patterns suggest that A3 proteins balance the need for functional conservation with the flexibility to adapt to varying evolutionary challenges.

3.1.3. Overall Structure of Vif from HIV Lineages and Primate A3FGs Maintained

To investigate how sequence variations in Vif and A3FGs affect their structures, we hypothesized that Vif would exhibit structural differences, as studies have suggested variations in binding and Zn requirements [163]. Few Vif structures have been resolved, with existing examples including HIV-1 and SIVrcm, leaving gaps for others such as HIV-2 and many SIV lineages [105,106,127]. Notably, mutational studies suggest that HIV-2, which infected humans via *Cercocebus atys*, may not require Zn binding by the HCCH domain and might interact with A3s differently [164]. Using I-TASSER, we predicted structures of A3FGs and Vif [118–120]. For lentiviral Vifs infecting great apes (GAs), their tertiary structures are largely conserved with respect to the known Vif structures (PDB:6NIL and 8E40), with most residues overlapping across viruses (Figure 4A) [66,127]. Variations primarily occur at the N-terminus, which forms a loop wrapping around the β -sheet, and

at the α -turn near the HCCH motif, where HIV-2 features a loop instead. In Old World monkey (OWM) Vifs, structural differences are more pronounced, particularly in all looped regions, except $\ell 9$, and the α -turn and α -helices, except for $\alpha 2$ and $\alpha 3$ (Figure 4B). For instance, $\alpha 1$ frequently disrupts prematurely at its termini with complete breakdown in SIV_{Volc}, and in some SIVs, the disruption occurs at the α -turn near Zn transitions into an α -helix. Variability in $\alpha 4$ includes differences in length or extra turns compared to HIV Vifs. Current models lack resolution for the Vif C-terminus, which is also missing from known structures. Examining known binding sites on a 3D structural homology model of HIV-1M NL4-3 (Figure 4C), regions interacting with A3F (cyan), A3G (hot pink), A3H (orange), and Zn in the HCCH domain (purple) show significant variation in SIVs infecting OWMs [41]. However, residues interacting with the E3 ubiquitin ligase complex remain conserved. This suggests that while HIV and its precursors share similarities in A3FG-binding regions, SIV Vifs from OWMs show considerable differences.

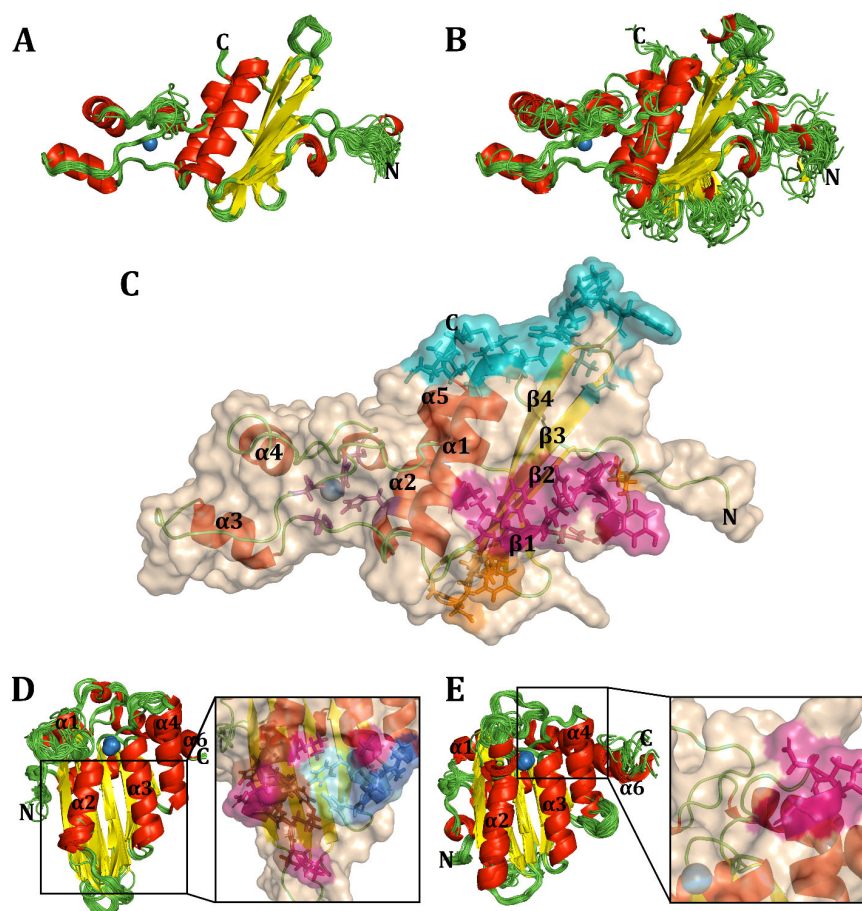


Figure 4. Structures of Vif and A3FGs across species. All structures are shown in cartoon format with colours based on secondary structure with loops (green), α -helices (red), and β -strands (yellow), except for zinc, which is shown as a sphere (sky blue). (A,B) Vif is shown for great apes as well as Old World monkeys. (C) Vif 3D structural homology model of HIV-1M NL4-3 shown as above with surface coloured (wheat) and the HCCH motif interacting with Zn shown as sticks (purple). The binding sites for A3F (cyan), A3G (hot pink), and A3H (orange) are coloured separately as noted in Wang et al. 2018 [41]. (D,E) The structures of A3F-CD2 and A3G-CD1, respectively, with their Vif-binding regions highlighted as in Delviks-Frankenberry et al. 2020 [4]. (D) The highlighted region of A3F shows the regions known to bind Vif (magenta), regions associated with Vif (brown), regions known to bind CBF- β (slate), and regions associated with CBF- β (cyan). (E) The known Vif-binding sites are shown as magenta.

We also compared structures of A3F-CD2 and A3G-CD1, hypothesizing these domains would share conserved core structures due to the evolutionary lineage of cytidine and adenosine deaminases [48]. Many APOBEC structures are available, predominantly from humans and macaques, including A3G-CD1 and A3F-CD2 [79–84,104,125,127,165,166]. Across primates, A3F-CD2 and A3G-CD1 display remarkably conserved base structures despite evolutionary divergence (Figure 4D,E). A3F-CD2 shows the most variation in loop regions, especially $\ell 3'$, whereas A3G-CD1's $\ell 3$ is shorter, similar to AID structures [94]. Unlike A3F-CD2, A3G-CD1 is inactive, with a smaller catalytic pocket and an extended turn near the catalytic Zn. Additionally, A3G-CD1 features a C-terminal extension of $\alpha 2$, consistent with prior findings [84]. Binding sites noted in Delviks-Frankenberry et al. for Vif on A3F and A3G remain structurally conserved, with minimal structural variation observed at the CBF- β -binding site [4]. These findings indicate that APOBEC regions bound by Vif have retained their core structures while evolving surface modifications.

3.2. Structural Evolution of A3F and A3G

3.2.1. The Amino Acid Variation of APOBEC3F and APOBEC3G

The dual-domain APOBEC proteins A3F and A3G contain the inactive pseudocatalytic CD1 domain and the active catalytic CD2 domain. Vif targets A3F-CD2 and A3G-CD1 specifically for degradation [80,81,167]. Alignments of A3F (Figure 5A) and A3G (Figure 5B) show similarities between the two domains, including conserved catalytic residues (HxE and CxxC motifs) in both active and inactive domains. Earlier studies focused on A3F and A3G in humans, African green monkeys, and macaques but revealed evolutionary pressures, such as sequence variation at position 127 in A3G but not in A3F [168]. Variation exists not only across species, but also polymorphisms exist within certain species [49,53].

Primates are categorized into three taxa: great apes (GAs), Old World monkeys (OWMs), and New World monkeys (NWMs). We hypothesized that alignments of A3F and A3G would show evolutionary adaptations while maintaining functionality. Alignments included 15 species for A3F (4 GAs, 8 OWMs, and 3 NWMs) (Figure 5A) and 15 species for A3G (5 GAs, 6 OWMs, and 4 NWMs) (Figure 5B). Additionally, 38 A3G sequences were identified with 23 additional OWM sequences (Figure S4). Adaptations were most pronounced between GAs and NWMs, separated by 33 million years of evolution, yet each species retained unique adaptations within otherwise conserved regions [49,53]. A key feature distinguishing A3F-CD1, A3F-CD2, and A3G-CD1 (A3Z2A subtype) from A3G-CD2 (A3Z1 subtype) is the presence or absence of specific motifs. A WF motif in α -helix 2, five residues after the catalytic glutamate, and an RI/RL motif at the start of $\ell 7$ characterize these subtypes [169]. Variations include a shift to RF in A3F-CD1 of some OWMs (e.g., macaques and *Rhinopithecus roxellana*) and an RI motif in A3G-CD1 of *Saguinus labiatus*.

Structural elements interacting with ssDNA, particularly loops 1', 3', 5', and 7', showed varying degrees of conservation and adaptation across taxa [4]. Loop 1 of the CD1's, from $\alpha 1$ to the gateway residue, contains polar and positive residues with a hydrophobic PIL motif in the middle. In A3F-CD1, loop 1 sequences in NWMs and *Chlorocebus pygerythrus* are distinct. A3F-CD2's loop 1 is more conserved and positively charged, while A3G-CD2's loop 1 is highly variable across taxa. Gatekeeper residues also vary. In A3F-CD2 and A3F-CD1, GAs and OWMs typically retain arginine, while NWMs use lysine or glutamine. A3G-CD1 gatekeepers are arginine, except in macaques (leucine), while A3G-CD2 consistently uses histidine. Loop 3, spanning four residues in A3F-CD1 and A3G-CD1, is variable after the initial valine. A3F-CD2's loop 3, 9 residues long, is conserved except in NWM, whereas A3G-CD2's 13-residue loop 3 is highly variable but remains polar and positively charged. Loop 5 is typically SWSPC across APOBEC proteins but is SWTPC in A3F-CD1, similar to A3B-CD1 in most primates. Some NWM variants of A3F-CD1 and A3G-CD1 display

SWNPC, resembling A3D [166,170]. Loop 7, associated with substrate specificity, contains an RLYYFW planar motif in most A3G-CD1 and both A3F domains. In contrast, A3G-CD2's loop 7 is RIYDDQ, exhibiting a more polar and negatively charged composition.

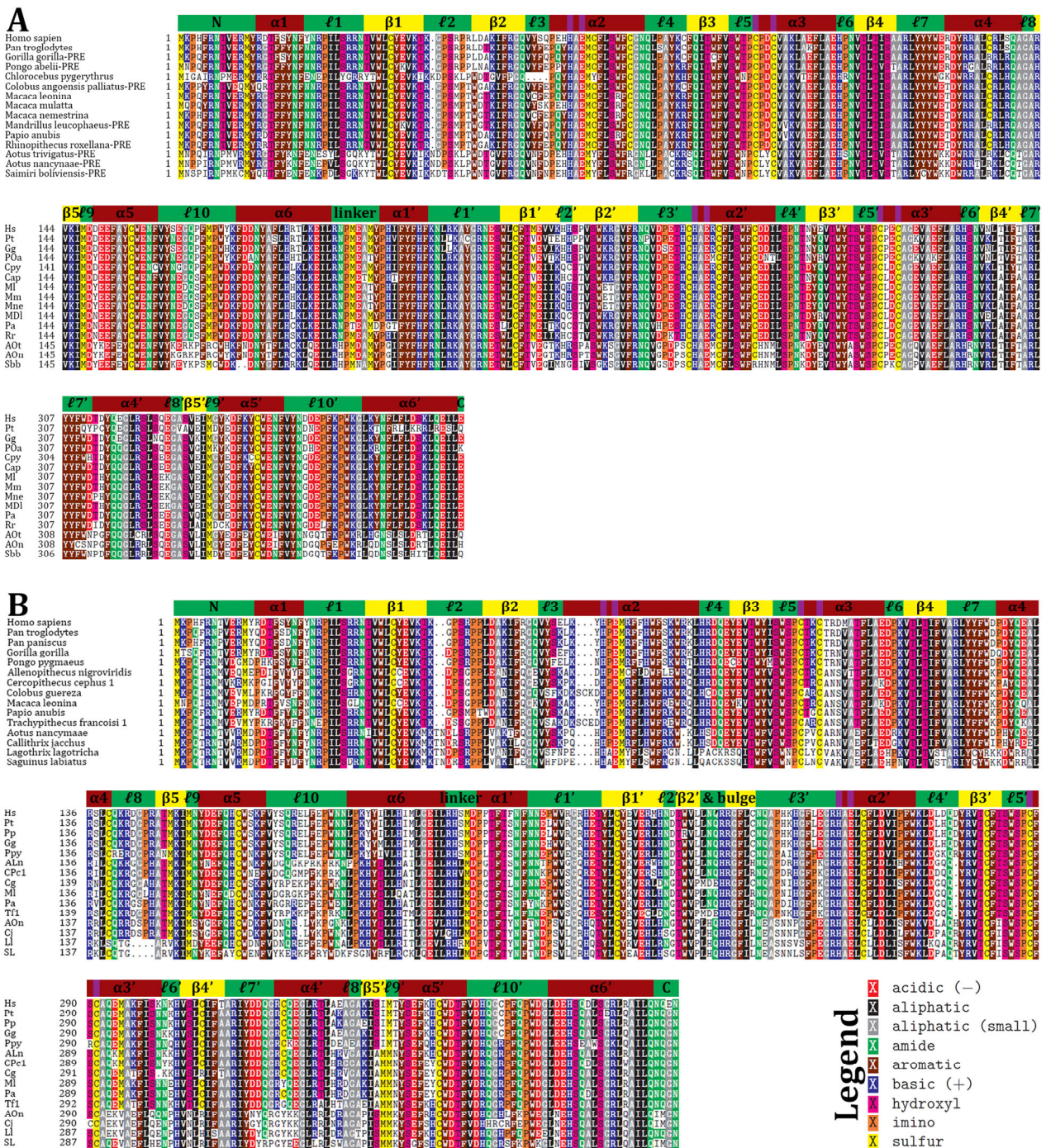


Figure 5. APOBEC3 alignments. Amino acid alignments of A3F (A) and A3G (B).

3.2.2. Structural Variation of A3F and A3G Across the Species

To understand the impact of sequence diversity on the 3D structures of A3F and A3G, we examined structural variations across species. Prior research established that human APOBEC3 proteins share a similar overall structure, including insights into A3F-CD2's DNA-binding mechanism, the positioning of its catalytic residues, and the Zn²⁺ cofactor pocket [79,81,82,104]. Furthermore, the negative patch on α-helices 2', 3', and 4', which

interacts with HIV-1 Vif, is also well characterized. However, structural data for A3F-CD1 and the Z2a subtype, which includes A3BD-CD1, were absent at the project's onset [80]. Thus, insights were inferred from related Z2 domains, such as A3C and A3G-CD1. We hypothesized that while Vif-binding regions might exhibit significant variation across taxa, the DNA-binding and oligomerization sites would remain relatively conserved.

Ribbon structures of A3F-CD1 and CD2 were generated and superimposed for comparison across species, including 60 models for each domain (Figure 6A,B). Both domains share six α -helices, five β -strands, and ten loops, along with N-terminal and C-terminal loop extensions on CD1 and CD2, respectively. This aligns exactly with the known experimental models of A3F-CD1 [79–82,104,125,127]. While α -helices and β -strands showed high structural conservation, loops were more variable. For instance, loop 3 in CD1, which is four residues long with a three-residue extension on α 2, differed markedly from CD2, where it extends to nine residues and protrudes into solution without rigid structural support. Despite these variations, key catalytic residues and the Zn^{2+} coordination pocket remained invariant, highlighting their evolutionary stability.

Surface analyses of A3F-CD1 and A3F-CD2 revealed extensive adaptations across the three taxa, including changes around the catalytic pocket and on distal surfaces (Figure 6C). On CD2, variations appeared on the protein's back, away from the DNA-binding groove. Notably, human A3F-CD2's DNA-binding region (especially loops 5' and 7') was highly conserved, with minimal changes at the C-terminal end of loop 7'. Surprisingly, the human Vif-binding region displayed little variation; only two residues, I262M and C259R, diverged in New World monkeys (NWMs) and *S. boliviensis*, respectively. Most of these adaptations occur on loops and α -helices that exist on the surface of the protein. When focusing on specific taxa, *Pan troglodytes* exhibited localized changes in CD1 near the catalytic pocket (e.g., loop 3) but fewer elsewhere. In CD2, adaptations were more pronounced away from the DNA-binding groove and on β -strands at the back of the protein, though the critical Vif-binding residues (291LAR293, E321) remained largely conserved, with the exception of E321K in *Macaca* and *Mandrillus* lineages. Similarly, Old World monkeys (OWMs) like *Papio anubis* showed adaptations in β -strands and the loop 7' distal end at D313H but retained stability in the DNA-binding groove. NWMs, such as *S. boliviensis*, displayed widespread structural changes, including within the Vif-binding region (C259R and I262M). Despite these modifications, the core structure of A3F and its DNA-binding regions remained intact across species.

Given the sequence diversity of A3G, we investigated its 3D structure, hypothesizing that Vif-binding regions would exhibit greater variation than DNA-binding regions. Ribbon structures of A3G-CD1 and A3G-CD2, superimposed across 144 and 152 models, respectively, revealed a similar structural framework to A3F (Figure 7A,B). Both domains retained six α -helices, five β -strands, and ten loops. While α -helices and β -sheets were conserved across species, including with respect to the experimental structures, loop 3 in both domains displayed significant backbone variation [84–86,165,166,171,172]. In humans, loop 3 in CD1 is 4 residues long, interacting with a three-residue extension of α 2, whereas CD2's loop 3' extends to 13 residues.

Surface analysis revealed substantial interspecies adaptations in A3G (Figure 7C). In humans, CD1 showed significant changes around the catalytic pocket and the back of the domain, while CD2 exhibited fewer variations near the DNA-binding groove and the back. Loop 7', critical for substrate specificity, remained mostly conserved except for minor variations at its proximal end away from the groove. In *P. troglodytes* and other GAs, CD1 displayed extensive adaptations, particularly in the Vif-binding region (e.g., Y19D, P129Q), while CD2 showed clustering of changes away from the DNA-binding groove. OWMs, particularly *P. anubis*, showed variations in CD1's ℓ 3 across all residues and Vif-binding

regions, including D128K, while maintaining CD2's DNA-binding region stability. NWMs, such as *S. labiatus*, displayed widespread changes across both domains, particularly in the Vif-binding region (Y59H, F126Y, D128K, and P129K). Although the DNA-binding region of A3G-CD2 remained largely conserved, NWM's modifications in loop 7' could influence substrate specificity.

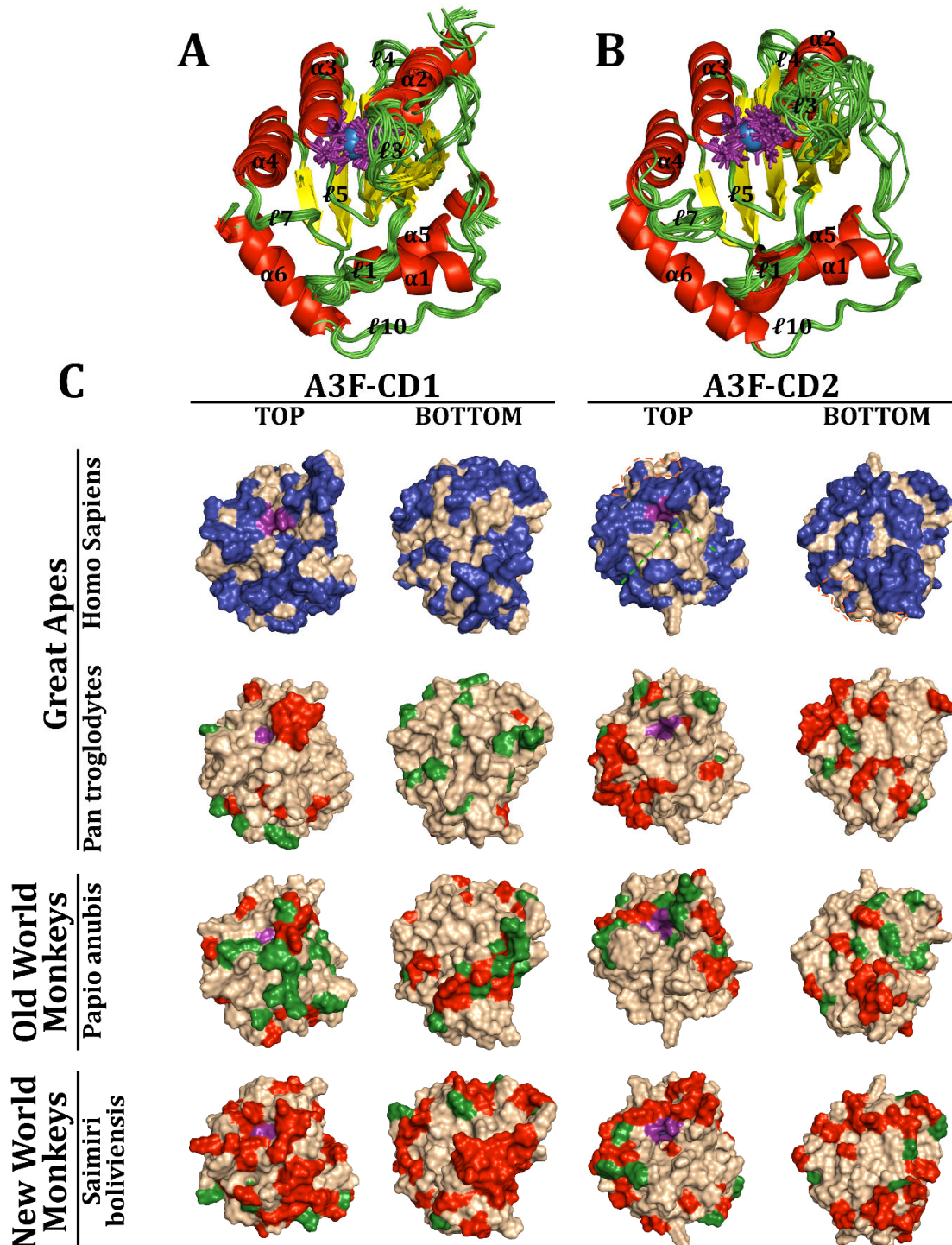


Figure 6. APOBEC3F structure and diversity. Ribbon structures of A3F-CD1 (A) and A3F-CD2 (B). The surface structures (C) of A3F-CD1 and CD2 with representatives of GAs, OWMs, and NWMs are shown from the top (for the DNA-binding groove), and the bottom. The colour scheme shows purple for the catalytic residues, wheat for the unaltered amino acids, and blue for all amino acids in humans that have adaptations across primates while across other primate representatives show adaptations in comparison to humans in red and adaptations compared to other members of the same taxa in green.

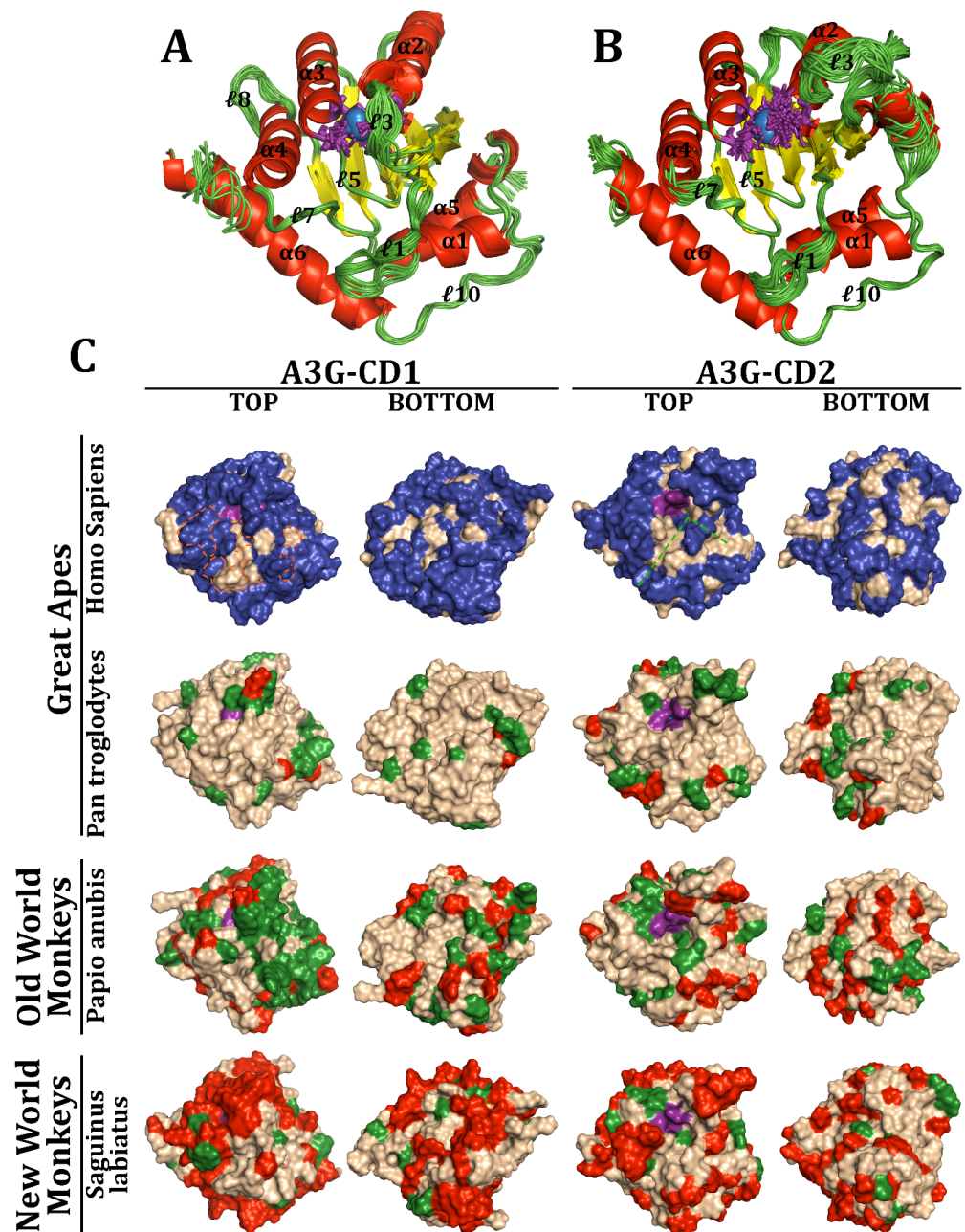


Figure 7. APOBEC3G structure and diversity. Ribbon structures of A3G-CD1 (A) and A3G-CD2 (B). The surface structures (C) of A3G-CD1 and A3G-CD2 with representatives of GAs, OWMs, and NWMs are shown from the top (for the DNA-binding groove), and the bottom. The colour scheme shows purple for the catalytic residues, wheat for the unaltered amino acids, and blue for all amino acids in humans that have adaptations across primates while across other primate representatives show adaptations in comparison to humans in red and adaptations compared to other members of the same taxa in green.

Across both A3F and A3G, the catalytic core and DNA-binding regions are strikingly conserved, reflecting their critical role in maintaining protein functionality. Variations are more pronounced in surface residues, particularly in the Vif-binding regions, which may reflect evolutionary pressures from host–pathogen interactions. In NWMs, extensive adaptations in both A3F and A3G highlight the evolutionary divergence over 33 million years [49,53]. The structural stability of the catalytic pocket and DNA-binding grooves across taxa underscores their functional importance. Meanwhile, surface adaptations, particularly in loops and Vif-binding regions, suggest evolutionary flexibility,

enabling species-specific responses to selective pressures. Together, these findings illuminate the evolutionary dynamics of APOBEC3 proteins and their role in antiviral defense across primates.

3.2.3. The Role of Charge Variations in Vif and A3 Proteins

Understanding the charge of proteins, much like their structural characteristics, is critical for grasping how they interact with other proteins and their environment. Protein charge plays a pivotal role in determining the dynamics of these interactions. In this context, we hypothesized that lentiviruses affecting the GA lineage would exhibit similarities in charge profiles, while the A3FGs would show respective charge similarities due to their overarching structural commonalities.

To investigate, we began with Vif proteins. Our analysis revealed substantial similarities in Vif charge profiles across SIVs infecting GA lineages, whereas there was 2 to 4 times greater variation within HIV and SIVs of OWMs (Figure 8A). Specifically, the charge at pH 7 for SIVs of *Pan troglodytes* and *Gorilla gorilla* ranged from 8.29 to 11.35, for SIVs of *Cercocebus atys* from 7.27 to 11.50, and for HIV in humans from 7.85 to 15.51. In contrast, other SIVs of OWMs displayed a broader range from 0.74 to 13.79. This variability highlights the extensive diversity in SIVs infecting various OWMs with relatively higher similarities among *Cercocebus atys*, *Pan troglodytes*, and *Gorilla gorilla*. When HIV entered human populations, this diversity expanded, suggesting that charge variation was essential for viral survival and replication in humans, even though A3s exhibited limited changes in charge.

Next, we examined the structural impact of these proteins on their charge regions over evolution. Structural analysis of the binding regions in HIV-1M NL4-3 Vif showed a prominent negative region around the A3F-binding site, while HIV-2 NNVA Vif exhibited a smaller negative region, with much of the surrounding area being positive or slightly positive (Figure S5A,B). In HIV-1M NL4-3, the negative charges originated from residues D14, E76, D78, and 171ED172, countered by positive charges from R77, R15, and R17, the latter two interacting directly with E76 and 171ED172, respectively. Conversely, HIV-2 NNVA's positive charges came from R18 and R174, with a negative charge from D175 and neutral charges from W73 and W81. The A3G-binding regions in both HIV-1M NL4-3 and HIV-2 NNVA Vif remained positively charged due to lysines and histidines. Interestingly, Vif proteins across lentiviruses demonstrated charge variation in their A3FG-binding regions, with fewer changes noted between SIVgor BPID and HIV-1M NL4-3 or HIV-2 NNVA and SIVsm SL92b, reflecting their close phylogenetic relationships (Figure S6).

Shifting focus to the A3 proteins, we examined how surface charge variation influenced the domain charges of these proteins. Based on prior knowledge of human A3 proteins, we hypothesized that the CD1 domains of A3F and A3G would remain positively charged, while CD2 domains would exhibit negative charges. PropKa was employed to measure the charge of 3D structural homology models for CD1, CD2, and dual-domain structures at pH 7, for A3F in Figure 8B and A3G in Figure 8C, respectively. In A3F and A3G, CD1 was consistently more positive than CD2, except in *Allenopithecus nigroviridis* (OWM) of A3G, where CD1 and CD2 had charges of 4.48 and 6.61, respectively. Remarkably, *Pongo abelii* A3F domains showed the smallest charge difference between CD1 and CD2 (1.265), contrasting with average differences of 9.166 for A3F (ranging 2.13–22.86) and 12.68 for A3G (ranging 1.265–13.06). The largest difference occurred in A3G of *Macaca nigra* (22.85). CD1 charges across A3F ranged from 3.25 in *Colobus angolensis* (OWM) to 12.86 in *Aotus nancynaa* (NWM), while CD2 charges ranged from slightly positive to negative, typically between 2.86 in *Aotus trivigatus* and -5.92 in humans. Notably, excepting *Allenopithecus nigroviridis*, CD1 and CD2 charges varied from 21.52 in *Macaca nigra* to 7.65 in *Gorilla gorilla*,

and 4.46 in *Cercopithecus wolffi* to -4.29 in humans, respectively. Dual-domain models of A3F often exhibited intermediate charges between CD1 and CD2, except in *Pongo abelii*, *Pan troglodytes*, and the three NWM species. For A3G, dual-domain charges typically aligned with or exceeded CD1 charges, a pattern particularly evident among *Cercopithecus* species. Analysing ancestral sequence reconstructions (ASRs) of the most likely ancestor and extant taxa (Figure 8D, and Tables S13 and S14), we found that A3F domains in GAs and OWMs were similar with NWM domains being more positive. In contrast, A3G domains in GAs and OWMs were more positive than their NWM counterparts. A3F domain charges were variable across GAs and OWMs but less so in NWMs, while A3G charges were consistent within GAs and NWMs but varied significantly in OWMs.

These charge distributions have functional implications. Positive CD1 domains, though non-catalytic, facilitate nucleotide interaction, RNA-directed virion incorporation, and DNA strand traversal through intersegmental transfer [35–37,63]. The charged surface of human A3F-CD1 (Figure S7B,F) features three positive regions: (A) the N-terminal loop, (B) loop 1 and the C-terminal of $\alpha 6$ in the DNA-binding groove, and (C) loops 2 and 4 with the back of β -strands 2–5. These patterns are not maintained in NWMs, where region (B) is less positive and loop 7 and $\alpha 4$ become positive along with the negative regions. Human A3G-CD1's charged surface (Figures S5C and S7D,H) retains a large positive region around the catalytic pocket that extends from the junction of $\alpha 6$ and loop 10, across loop 1, the side of $\beta 1$, and around loop 3 and α -helices 2 and 3 to loops 2 and 8, and other small positive areas around β -strands 4 and 5. A3G-CD1's three negative regions include (A) across $\alpha 1$ and $\alpha 5$ with $\ell 10$, (B) $\ell 7$, $\ell 10$ and $\alpha 4$, and the C-term of $\alpha 3$. These patterns are conserved across OWMs but differ in NWMs, where positive regions (B) become less positive and loop 7 and $\alpha 4$ gain positive charges. Across OWMs and NWMs, charge locations shifted significantly, with some negative regions in OWMs, like *Macaca nigra*, nearly absent (Figure S7A).

CD2 domains are pivotal for DNA binding and sequence specificity [173]. Human A3F-CD2's surface (Figures S5D and S7C,G) includes one major positive region and two negative regions. Around the DNA-binding groove were the positive region on $\ell 1'$, $\ell 10'$, $\alpha 1'$, $\alpha 5'$, and $\alpha 6'$, with negative charges around $\ell 3'$, $\ell 7'$, $\ell 10'$, $\alpha 2'$, $\alpha 3'$, $\alpha 4'$, and $\alpha 5'$. This distribution is largely consistent across GAs and OWMs, while in NWMs, positive and negative regions are redistributed. Human A3G-CD2 (Figure S7E,I) exhibits five positive and four negative regions, with substantial conservation across taxa. The charged structures around the DNA-binding groove consist of positive charges across $\ell 1'$, $\alpha 3'$, $\alpha 4'$ and $\alpha 6'$, with negative charges across $\ell 1'$, $\ell 3'$, $\ell 7'$, $\ell 10'$, $\alpha 1'$, $\alpha 3'$, $\alpha 4'$, $\alpha 5'$, and $\alpha 6'$. The CD2 domain charges across A3F and A3G are divergent, which may be a function of the differences between A3 subtypes and may be in part due to how they interact with their respective substrates (Figure S7A).

Finally, we examined charge trends across ASR and extant taxa using PropKa (Figure 8E–H). In A3F-CD2 domains, charges increased in GAs and OWMs but not in NWMs (Figure 8F). In A3F-CD1, GAs and NWMs showed charge increases (Figure 8E). Conversely, A3G-CD1 charges decreased in GAs and OWMs but increased in NWMs (Figure 8G). A3G-CD2 domains showed no significant charge shifts in OWMs or NWMs, though OWM charges expanded in range while GA charges decreased (Figure 8H). Notably, charge diversification in OWM A3G domains aligns with positive selection and OWM expansion [49]. Overall, our findings reveal that Vif and A3 charge regions have been adapted not only in areas where they interact, but also across the entirety of their surface with exceptions to the A3 DNA-binding regions.

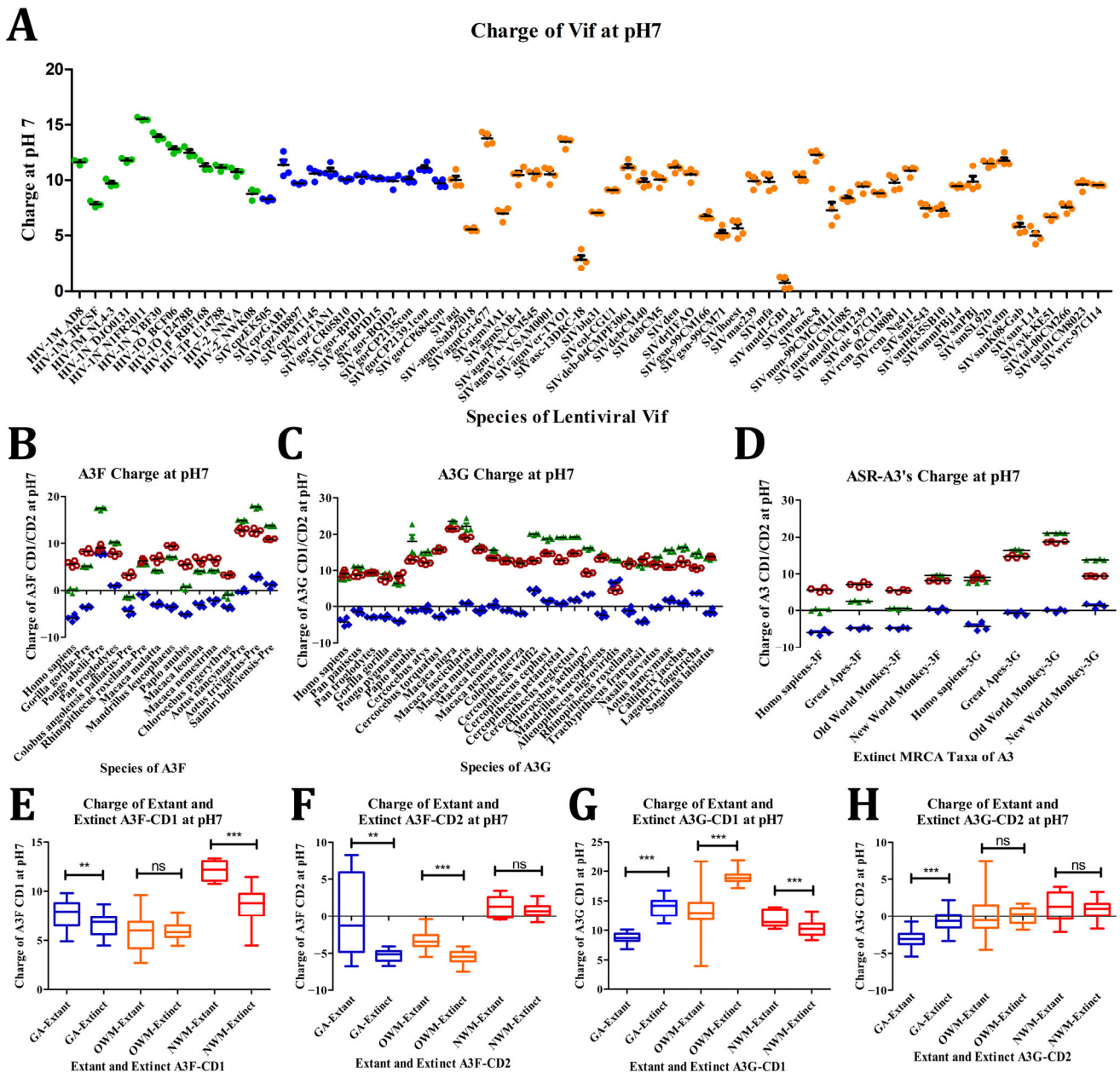


Figure 8. Charge of Vif and APOBEC3FG. The charge of Vif (A), A3F (B), A3G (C), and A3FG ASR models (D) at pH7. Vif (A) graph is coloured by HIV/SIV host with humans (green), great apes (blue), Old World monkeys (orange), and New World monkeys (red), while A3FG models are coloured by domain with CD1 (red), CD2 (blue), and double domain (green). The charges of extant and extinct ASR models were produced for A3F-CD1 (E), A3F-CD2 (F), A3G-CD1 (G), and A3G-CD2 (H), where they are coloured by taxa with great apes (blue), Old World monkeys (orange), and New World monkeys (red). [ns = not significant, *** $p < 0.0001$, (E) ** $p = 0.0019$, (F) ** $p = 0.0024$].

3.3. Overall A3 and Vif-Binding Modes Remain Similar Despite Adaptation

As we analysed the structural properties of protein cores and their surface charges, our focus shifted to understanding their interactions. Using HADDOCK 2.4, we sought to predict how structural and surface charge variations influenced binding between Vif and APOBEC3 (A3) proteins, specifically A3F and A3G. We hypothesized that these interactions would vary significantly due to differences in surface structure and charge, despite their shared functions in host–virus dynamics. To our surprise, despite sequence variability, secondary structures across these proteins and their modes of interactions remained largely

conserved, as indicated by our conformational analyses (Figure 9, and Tables S15–S22). This consistency suggests that the binding interfaces retain functional conservation even amidst evolutionary divergence.

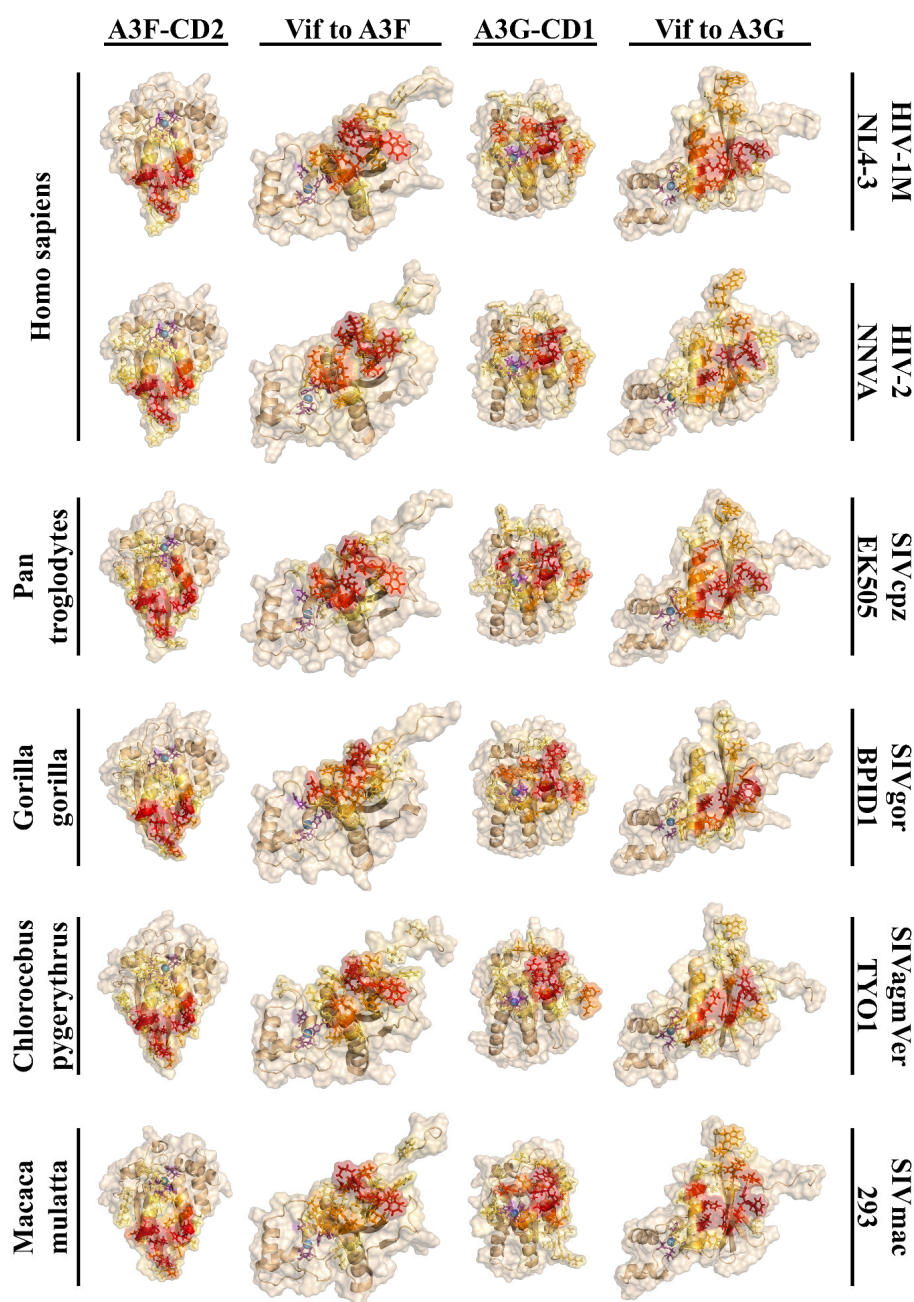


Figure 9. Haddock heatmaps of A3FGs and Vif across primates and viruses. Heatmaps of interactions recorded when docking host A3FG with HIV/SIV Vif. At the top are shown the columns of APOBECs as well as Vif interacting with either A3F or A3G. On the left and right are the primates and the viral strains, respectively. All 3D structural homology models within each column are shown at the same angle for ease of viewing interactions. Cartoons are shown with transparent surfaces, where zinc is shown as a sky-blue sphere; the HCCH motif of Vif and the catalytic residues of A3 are shown as purple sticks; and interactions are shown on a spectrum. The spectrum is shown based on percent interaction with wheat (0%), yellow-orange (25%), orange (50%), red (75%), and firebrick (100%). Sticks of residues were shown under the condition that the residues were interacted with at least 10% of the time.

In A3F, interactions with Vif primarily occurred on $\alpha 2$, $\ell 4$, and $\alpha 3$, with additional binding on $\ell 6$ (Table S15). On Vif, the primary interaction sites were concentrated on $\ell 5$, supported by minor interactions on $\alpha 1$, $\ell 1$, and $\ell 11$ (Table S16). These interactions were similar across multiple species, but there were notable differences in intensity and frequency, highlighting the evolutionary tuning of these interfaces. Detailed analysis of the A3F-Vif interface revealed host-specific variations (Tables S19 and S20). For A3F-CD2, minor interactions were observed on $\ell 2$. In GA species, this involved residues 226K/EHH228, whereas interactions in OWM were less pronounced. Significant interactions were concentrated on residues 259CD/ED261 and 263LSPN266 with supporting interactions from S256 and I262. Additional binding was observed on residues 298EF290 and 293RHSN296, while E286 in humans had lesser interactions than K286 in *Pan troglodytes*, showing differential interaction patterns. The Vif-binding interface included three primary regions. The first region, involving residues 14DRMR17 in HIV-1 lineages and 16PGD/R18 in HIV-2 lineages, consistently interacted across species (>50% of the time). These were supported by minor interactions on $\ell 1$ from residues such as Q12 (HIV-1 lineages) and R/K14 (HIV-2 lineages). The second region, centered on the 76ER/K/Q77 motif in $\ell 5$, exhibited high interaction frequencies (>70%). Interestingly, we noted a reduction in interactions involving G71 in HIV-1 lineages, whereas the corresponding residue N74 in HIV-2 lineages showed stronger binding. The third major region included residues 171ED172 in HIV-1 lineages and 174RD/N175 in HIV-2 lineages, both maintaining significant interaction levels. During our time on this project, a CryoEM structure of A3F-CD2 and Vif was resolved, for which we calculated the interactions (Figure S8A,B and Tables S19 and S20) [127]. We found that our A3F-CD2 interactions aligned exactly with those of the CryoEM structure, while the Vif portion picked up interactions from H43 and Y44 that were only seen some times in our structures.

For A3G, Vif interactions were concentrated on $\alpha 4$ and $\ell 7$, with additional binding observed on $\ell 1$, $\ell 3$, and $\alpha 3$ (Table S17). While all lentiviruses had high interactions on $\alpha 1$, there was a combination of interactions between $\beta 1$, α -turn 1, and $\ell 3$, as well as minor but notable interaction in $\beta 3$ and $\ell 5$ (Table S18). Despite some variability, these interactions were broadly conserved across A3Gs and lentiviruses (Tables S21 and S22). Strong interactions were noted on the 127WDPD130 motif, with further strong binding on residues Y125 and Y59 (except in *Chlorocebus pygerythrus*). Additionally, residues 131YQE/Q133 on $\alpha 4$ and 98TK99 and R102 on GAs showed interactions. Loops around the catalytic pocket, particularly $\ell 1$, displayed heightened interaction, with OWMs showing elevated binding on I26 and S28. The Vif residues for A3G-binding revealed critical interaction points. Known residues such as Y40 and the 42HHYE45 motif in HIV-1 lineages displayed strong binding, with *G. gorilla* uniquely exhibiting interactions on R41. Similar patterns were observed in HIV-2 lineages. Other high-interaction regions included residues K26 and $\alpha 1$ residues N19, 22KR23, and 30YI31, though these varied among lentivirus species. Across all species, residues W70 and 79WH80 consistently interacted with A3G. HIV-2 exhibited additional binding on W51, C54, and Y85, while in SIVagmVer-TYO1, interactions involving W51 and C54 were replaced by M53. During this project, six CryoEM structures of A3G and Vif were resolved, which we analysed (Figure S8, and Tables S21 and S22) [66,67,172]. A3F-Vif structures aligned with our dockings, however, with only one known structure we were unable to analyse with regards to its frequency of interactions. However, between A3G and Vif, there were six structures also containing ssRNA or dsRNA. For the most part, our results including frequency align with theirs save for interactions between A3G $\ell 1$ and Vif $\alpha 1$ C-term occurring through the RNA.

These interaction patterns highlight a functional necessity for conserved binding topologies, even as specific residues diverge across host species. Despite evolutionary

pressure leading to sequence variation, the overall interaction networks between Vif and A3 proteins maintain structural integrity. This balance likely ensures viral adaptation while preserving critical binding capabilities. The results underscore that differences in binding between Vif and A3 proteins are often subtle, manifesting at the level of single amino acids rather than broader structural shifts. This precision reflects the evolutionary co-adaptation of host antiviral defenses and viral countermeasures. While the primary interaction regions are conserved, species-specific modifications optimize binding efficiency, providing insights into the molecular dynamics underlying host–virus interactions. These findings contribute to a broader understanding of how structural and charge variability impacts protein–protein interactions in the context of viral evolution.

3.4. A3FG-DNA Interactions

3.4.1. DNA Conformations and the Catalytic Pocket of A3F and A3G

A3F-CD2 and A3G-CD2 display significant diversity in structural homology, surface configuration, and charge across species. However, their core structures and DNA-binding regions remain remarkably conserved. This led us to investigate whether their preferred DNA motifs interact consistently across species. For A3F, we used the motif *taTTCat* and docked it to observe nucleotide interactions with the catalytic pocket (Figure 10A). In A3F-CD2 across *Homo sapiens*, *Pan troglodytes*, *Papio anubis*, and *Saimiri boliviensis*, the thymidine (T4) at position 4 most frequently occupied the catalytic pocket (50.2%), while terminal nucleotides T1 (4.2%), A6 (6.6%), and T7 (5.3%) were the least frequent. After T4, interactions varied: in *H. sapiens*, adenosine A2 (16%), and cytidine C5 (13.4%) were prevalent, while in *S. boliviensis*, T3 (12.6%) and C5 (13.8%) were dominant. *P. troglodytes* showed a mix of A2 (10.0%), T3 (10.9%), and C5 (11.9%), while *P. anubis* favored A2 (14.4%) followed by T3 (8.1%), A6 (8.6%), and C5 (9.7%). Critically, deamination could only occur when a cytidine occupied position 5. Overall, nitrogenous bases entered the pocket in 48.0% of docked conformations, while the backbone (phosphates or riboses) accounted for 6.6%. DNA often remained near the pocket residues (45.2%) and was rarely positioned far from the catalytic pocket (0.2%). For A3G, the motif *ttCCCtt* was used, and docking revealed consistent interactions across *H. sapiens*, *P. troglodytes*, *P. anubis*, and *Saguinus labiatus* (Figure 10C). Cytidine C4 was the most frequently positioned nucleotide in the catalytic pocket (37.7%), followed by C5 (21.8%), which is also the nucleotide most deaminated in vitro. Terminal nucleotides T1 (7.3%), T2 (12.5%), T6 (8.9%), and T7 (10.4%) were consistent, while C3 had the lowest interaction rate (1.7%). The palindromic nature of *ttCCCtt* occasionally reversed the directionality of C3, which was reclassified as C5 for this analysis. Nitrogenous bases entered the catalytic pocket in 57.1% of conformations, with the backbone contributing 12.1%. When nucleotides did not enter the pocket (30.3%), most stayed near catalytic residues, with only 0.5% positioning far away. Despite differences in substrate preferences, A3F and A3G exhibit conserved interaction patterns with their respective DNA motifs across species. This consistency highlights the evolutionary stability of their catalytic mechanisms despite surface-level diversity.

3.4.2. Molecular Interactions of DNA Sequences with APOBEC3F

To understand the interactions between nucleotide chains and the amino acids of A3F-CD2 across species, we analysed dockings where the cytidine at position 5 (C5) of the DNA motif *taTTCat* was within the catalytic pocket. Using PyMOL, we visualized interactions, hypothesizing that despite structural variations, the conserved DNA-binding groove would result in stable interactions across species. Interaction levels were measured as the percentage of contact between DNA and APOBEC proteins across *Homo sapiens*, *Pan troglodytes*, *Papio anubis*, and *Saimiri boliviensis* (Figure 10B). DNA consistently folded

along the APOBEC DNA-binding groove, with amino acid interactions visualized in red (high interaction) to wheat (no interaction). Looking first at the secondary structure, most interactions (75.2%) occurred on protein loops: $\ell 1'$ (29.9%), $\ell 3'$ (15.3%), $\ell 5'$ (11.0%), and $\ell 7'$ (19.0%) (Table S23). Humans showed fewer interactions on $\ell 7'$ but compensated with increased interactions on $\ell 3'$. Secondary structures like $\alpha 1'$ (5.4%), $\alpha 2'$ (4.1%), $\alpha 3'$ (5.9%), and $\alpha 4'$ (4.0%) showed modest interactions, while $\beta 1'$ (1.2%) and $\beta 2'$ (2.3%) exhibited fewer interactions. Interestingly, $\alpha 6'$ had moderate interactions in great apes (e.g., 2.5% in humans, 3.1% in *P. troglodytes*), which were absent in *P. anubis* and *S. boliviensis*.

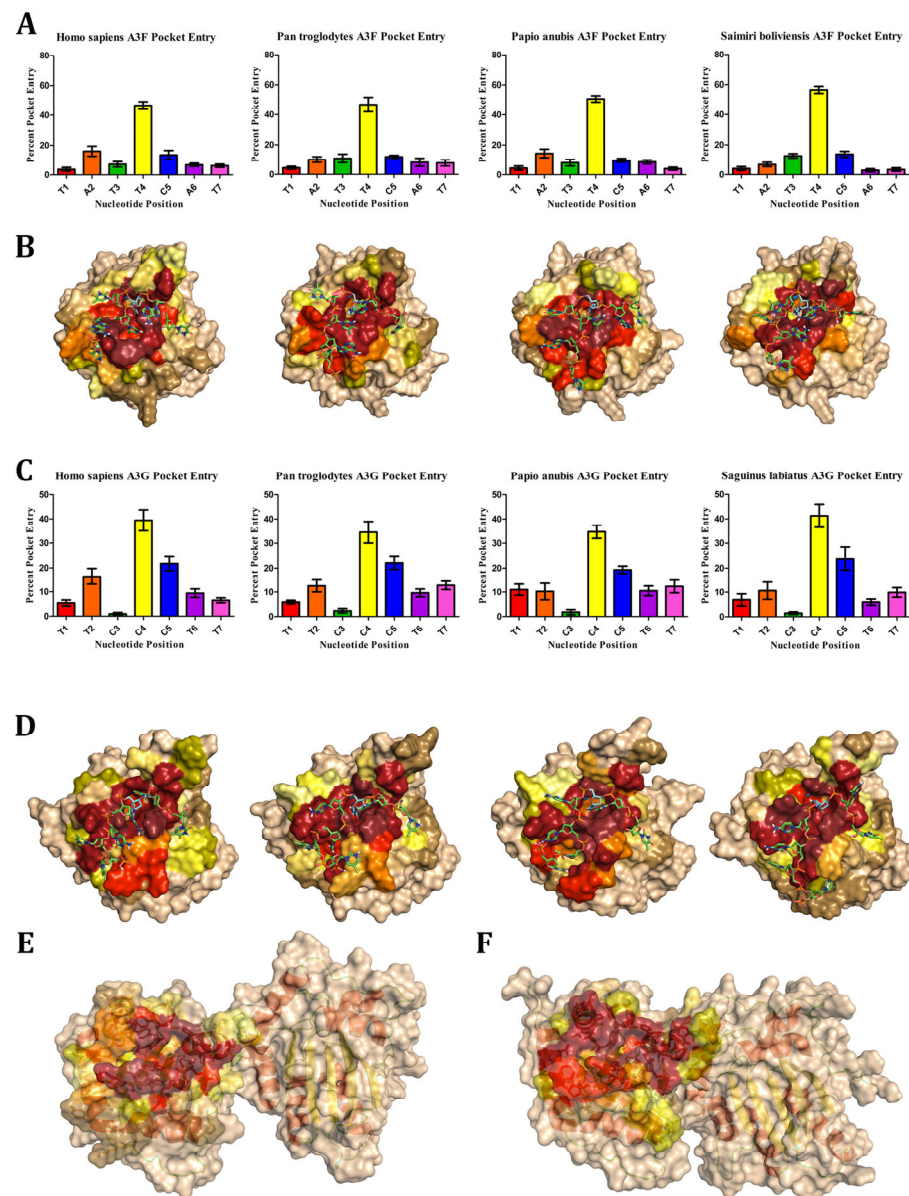


Figure 10. DNA-APOBEC interactions. Analysis of which nucleotide remained in the catalytic pocket most often in A3F species (A) and A3G species (B). Structural heatmaps of DNA-APOBEC interactions when deoxycytidine is in the catalytic pocket, shown with representatives of the DNA on the surface bound to A3F [left to right: *Homo sapiens*, *Pan troglodytes*, *Papio Anubis*, *Saimiri boliviensis*] (C), and A3G [left to right: *Homo sapiens*, *Pan troglodytes*, *Papio Anubis*, *Saguinus labiatus*] (D). The heatmap of interaction levels with DNA coloured spectrum from 100% to 0% as ruby, firebrick, red, orange, bright orange, yellow orange, light orange, tv yellow, pale yellow, olive, sand, and wheat. Heatmaps of human double-domain A3F (E) and A3G (F) are also shown using the heatmap data available from the single domain human heatmaps.

On $\ell 1'$, two key regions consistently interacted across species. In humans, residues 208RK209 and 211YGRNES216 were highly interactive, separated by A210 with moderate interactions (Figure 10B and Table S24). Notable adaptations included *P. troglodytes* substitutions R208L and Y211C. N214, a “gatekeeper” residue, interacted with DNA 100% of the time in humans, *P. anubis*, and *S. boliviensis*. Interactions extended to W217 (or L217 in *P. anubis*) in $\beta 1'$. $\ell 7'$, the substrate specificity loop, also exhibited strong interactions. Across species, residues 305RLYYFWD311 (human sequence) showed significant interaction, particularly the hydrophobic region 307YYFW310. *H. sapiens* had less interactions on R305 than all the other species. *P. troglodytes* substitutions from WD to 310QY311 raised site 311 interactions by ~20% compared to other species. Additional interactions occurred at $\alpha 4'$, particularly Y314 (F314 in *S. boliviensis*) and, to a lesser extent, D313 (C313 in *P. troglodytes*). $\ell 3'$ contained a high number of interactions at N240, V242, and catalytic H249. Moderate to low interaction was seen on Q241, 243DP244, and 246TH247, but interaction was reduced in *S. boliviensis*, which underwent significant adaptations (e.g., human 243DPETH247 to *S. boliviensis* 242GSDPS246). Shifts in interactions were observed at R239 of $\beta 2'$, which showed high interaction levels in *S. boliviensis* compared to moderate levels in other species. $\ell 5'$ and $\alpha 3'$ featured conserved residues with moderate interaction levels. Notably, the sequence 277WSPCPEC283 displayed high interaction levels at W277 and catalytic C280. Some species-specific variations included substitutions at 281PE282 to 281LD282 in *P. anubis* and E281 to K282 in *S. boliviensis*. Other secondary structures, like $\alpha 1'$, had consistent moderate interaction levels at I199 (V in *P. troglodytes*, T in *P. anubis*) and 202FH203. In $\alpha 6'$, *H. sapiens* showed a moderate number of interactions with F363, while *P. troglodytes* had additional moderate interaction levels on L364 and R367. *S. boliviensis* and *P. anubis* exhibited only a low number of interactions in 363FL364 and K367.

Despite variations around the edges of the DNA-binding grooves' amino acid sequences, A3F-DNA interactions are consistent across species. Most interactions occur on specific loops and conserved residues, maintaining the functional integrity of the catalytic pocket and substrate specificity. These findings underscore the evolutionary stability of A3F's interactions with DNA, highlighting the conserved mechanisms underpinning APOBEC activity.

3.4.3. Molecular Interactions of DNA Sequences with APOBEC3G

To further explore the interactions between nucleotide chains and A3G-CD2 across taxa, we examined its preferred motif (*ttCCctt*) when the cytidine at position 5 (C5) was within the catalytic pocket. Using PyMOL, we visualized interactions and hypothesized that the conserved DNA-binding groove would result in many consistent interactions across species despite sequence diversity. As shown in Figure 10D, DNA folded across the APOBEC DNA-binding groove in 3D structural homology models with amino acids colour coded by interaction frequency (red indicating the most interactions and wheat indicating none). Analysis of secondary structures revealed that 77.8% of interactions occurred on the protein loops, predominantly $\ell 1'$ (27.4%), $\ell 3'$ (20.3%), $\ell 5'$ (11.4%), and $\ell 7'$ (18.6%) (Table S25). Loops $\ell 1'$, $\ell 3'$, and $\ell 7'$ displayed consistent interactions across *Homo sapiens*, *Pan troglodytes*, *Papio anubis*, and *Saimiri labiatus*. However, $\ell 5'$ exhibited slightly reduced interactions in *S. labiatus*. Secondary structures such as $\alpha 1'$ (4.5%), $\alpha 2'$ (4.7%), $\alpha 3'$ (3.3%), and $\alpha 6'$ (4.4%) contributed modestly, while $\beta 1'$ (1.5%), $\beta 2'$ (1.3%), and $\alpha 4'$ (2.0%) had lower interaction frequencies. Sporadic interactions occurred across other regions.

In $\ell 1'$, two regions, 210PWVRG214 and 216HET217, showed high interaction frequencies, separated by a moderate number of interactions at R215 (Figure 10D and Table S26). The first region included adaptations such as L210 in *P. troglodytes*, S213 in *P. anubis*, and both S211 and L213 in *S. labiatus*. In the second region, there were only interactions at

Q217 in *S. labiatus*. High interaction frequencies were observed universally at H216, a key “gatekeeper” residue above the catalytic pocket. This was consistent across species, continuing through the next two residues until $\beta 1'$, where Y219 exhibited moderate interaction frequencies. On $\ell 3'$, two highly interactive segments were identified: 244NQAP247 and 254EGRH256, including catalytic H257. Residues N244, 246AP247, and H257 consistently showed high interaction frequencies. Notably, *P. anubis* displayed strong interactions at Q245, while *S. labiatus* showed high interaction levels at its S248 adaptation. Moderate interaction frequencies were seen at Q245 (E242 in *S. labiatus*) and 254EG255 (K254 in *P. anubis*). Residues such as 248HK249 (248DI249 in *P. anubis* and 245NN246 in *S. labiatus*) and R256 exhibited lower interaction frequencies. On $\beta 2'$, position 243 showed moderate interaction levels across species, varying between cysteine (great apes), leucine (New World monkeys), and arginine (*P. anubis*), where interactions were strongest. Catalytic glutamate (E259) consistently maintained high interaction frequencies, while A258 had moderate interaction levels. $\ell 7'$, the substrate specificity loop, featured high interaction frequencies across the critical motif 315YDDQ318 (*H. sapiens*, *P. troglodytes*, *P. anubis*) or 312YDYR315 (*S. labiatus*). Neighbouring residues I314 and R313 displayed high to moderate and low interaction levels, respectively, influenced by their positioning within the $\ell 1' / \ell 7'$ interface. Moderate interaction levels extended to $\alpha 4'$, particularly R320 (Y318 in *S. labiatus*). In $\ell 5'$, the motif 285WSPC288 showed predominantly high or moderate interaction levels across species, though the number of interactions with S283 in *S. labiatus* was low. These interactions extended to $\alpha 3'$, where C291 maintained moderate interaction levels, while residues 289FS290 exhibited lower interaction frequencies. In $\beta 3'$, low interaction levels were observed at T283 (*H. sapiens*; I280 in *S. labiatus*). Toward the C-terminal region, $\alpha 6'$ consistently displayed high to moderate interaction frequencies at R374 and moderate interaction frequencies at L371, with other interactions occurring sporadically.

The interaction analysis across secondary structures and amino acid residues revealed a highly conserved A3G-DNA-binding interface across taxa. Despite sequence diversity and structural adaptations, the DNA-binding groove consistently mediated interactions, maintaining functional integrity and catalytic activity. These findings underscore the evolutionary conservation of A3G's DNA-binding mechanism, highlighting its critical role in APOBEC function.

3.4.4. Comparison of A3F and A3G Double Domains Across Ancestral Proteins

One issue with understanding how DNA binds double-domain APOBEC proteins is that the double-domain proteins can exist in globular states, dumbbell states, and transitional states (between the two), which affects how these proteins are able to traverse the genome [63,65,91,165]. This means that DNA can interact with these domains differently dependent on which form is present at the time. As such, we created double-domain 3D structural homology models of these proteins in AlphaFold2, across which the A3Fs (Figure 10E) tended to maintain a dumbbell and the A3Gs (Figure 10F) maintained globular conformations; yet, in both cases, these 3D structural homology models were formed with a similar rotation in orientation with respect to CD2. This rotation is more similar to the recent experimental A3G structures than to the original ones produced in 2020 by Maiti et al. and Yang et al. [65,165,166,172]. Although these 3D structural homology models are static, these conformations would be in constant flux. The conformations of both grooves of A3F and A3G of humans, coloured by their interactions with DNA, show that in these conformations, the nearest place for the DNA to interact would be across the β -sheet of the CD1 domain, which has previously been shown to be a putative RNA-binding region in AID [174]. These are typified by positive charges sandwiched between the positive loops 2', 4', 6', and 8' as well as the negative charges across the $\alpha 5'$, $\alpha 6'$, and loop 10' with positive or

polar charges across the β -sheet (Figure S7F,H). Together, this shows an alternate path that ssDNA could take along the β -sheets of A3FG-CD1 domains across all species that in AID is known to bind RNA. This β -sheet region is similar across all ancestral A3F-CD1s and A3G-CD1 of GAs, while ancestral A3G-CD1s of OWMs and NWMs have modifications to this region (Figures S9 and S10).

4. Discussion

As seen in SIV/HIV, zoonosis requires adaptation to host zoonotic barriers, thereby forming viral variants and later new viral species [6,8]. Such pressures on viruses and hosts can be seen most starkly in the interactions between host barriers and viral antagonists. Throughout this study, we have attempted to understand the coevolution of hosts and viruses through the computational analysis of the evolution of Vif, A3F, and A3G, as well as their interactions with their targets through antagonism and DNA, respectively. Most structures of APOBECs and Vif from prior studies have been limited to humans and several other species, such as *macaca mulatta*, without a comprehensive look across all primate and viral species [67,104,127,165,166]. Furthermore, these studies predominantly showed A3F, A3G, and Vif with mutations or truncations including only one A3 domain present at a time, and only recently have double-domain A3G crystal structures been created. As such, it was difficult to understand the qualities and ramifications of how interactions between these macromolecules across hosts and viruses in a single species, let alone across the evolution of all known primates and their known lentiviral antagonists, occurred. To this end, we noted the high level of genetic diversity across species, including at the A3-Vif-binding sites especially in Vif (Figures 2, 3 and 5), and determined that there are adaptations across the surface of A3s but not in the DNA-binding groove (Figures 4, 6 and 7); that charges have changed across evolution with respect to known Vif-bound domains of A3s, as well as the diversity of charges that exist across Old World monkeys (Figure 8); and the similarities in binding interactions of A3s with Vif and DNA across species given this level of diversity (Figures 9 and 10). Throughout this study, we expanded on these key structural inferences to the coevolutionary dynamic of the A3-Vif host-viral system.

Viruses, with their short life cycles and error-prone reverse transcription, evolve rapidly under host selective pressures [49,157]. Host zoonotic barriers, such as APOBEC3s (A3s), SamHD1, tetherin, and SERINC3s, exert strong evolutionary pressure, driving viral adaptations [6]. For example, the stability of host A3s compared to viral Vif binding sites (Figure 2) illustrates this dynamic. Notably, A3G-Vif-binding regions are more conserved than A3F-Vif-binding regions, despite A3G-CD1 being non-catalytic (Figures 2 and S1). All Vif-bound A3 domains, except A3H (A3Z3 family), belong to the A3Z2 domain type (e.g., A3B-CD1, A3C, A3F-CD2, A3G-CD1) [4,41,175]. While human A3CDF proteins share conserved binding sites with Vif, A3G's unique binding site likely requires fewer viral adaptations to inhibit its function. As Zennou et al. (2006) demonstrated, different species exhibit variable A3 mutational rates and Vif inhibition capacities, reflecting species-specific selective pressures [168]. This may lead to diverse Vif adaptations targeting distinct A3 family members, depending on polymorphisms and their effectiveness against viral defenses [53,176]. Primate A3FG and HIV/SIV phylogenies used prior to creating ASR (Figures S11 and S12) highlight the extensive diversification of both host and viral proteins, which is consistent with findings from previous studies on divergence patterns [5,49,52,176,177]. Reduced gene identity and similarity in A3 sequences (Figures 3 and S2 and S3) further validate these observations. As Sawyer et al. posited, A3G undergoes Darwinian positive selection at viral binding sites, a phenomenon also observed in A3F, tetherin, TRIM5 α , and SamHD1, indicating that restriction factors adapt to diverse viral pressures [49].

This study focused on A3F and A3G restriction factors, examining their structural features, surface charge distributions, and interactions with preferred DNA motifs. Amino acid alignments highlight regions of perfect conservation alongside numerous non-synonymous adaptations, especially within catalytic loops, potentially affecting DNA interactions (Figure 5). Interestingly, some sites universally conserved across human A3s [87], such as A3F positions R30, I117, F159, F167, P169, M195, and I301, as well as A3G positions R30, I118, L135, F172, P174, W175, R215, F262, L271, V305, I309, A312, and I337, exhibit variability across taxa. This suggests diverse evolutionary pressures shaping these proteins, leading to conservation across humans that was lacking across some taxa or even in some cases species, although it cannot be determined what pressures caused all of these adaptations.

The amino acid diversity of A3F and A3G spans a broad range (97.3–76.7% and 97.4–70.2%, respectively), contrasting with the highly conserved immune gene AID (99.5–97.5%), which is essential for immunoglobulin DNA somatic hypermutation and class switch recombination, having been conserved back to lampreys (Figures 2B,C and S3A) [158,178–181]. A3F and A3G, however, function on the front lines of immune defense, targeting lentiviral Vif (both directly and indirectly), and possibly extinct viral antagonists [62]. Evolutionary pressure is more pronounced on CD1 domains, which lack direct catalytic activity, than on CD2 domains, which must retain DNA-binding grooves and catalytic pockets for activity (Figure 3D,E). This dichotomy aligns CD1s more closely with TRIM5 α in diversity and CD2s with SamHD1 and tetherin for A3F and A3G, respectively. CD1's role in RNA binding for HIV/SIV virion packaging and intersegmental DNA transfer likely contributes to its higher diversity compared to the catalytically active CD2 [35,61,78,182]. Notably, lentiviruses do not infect all primates, such as New World monkeys, suggesting species-specific pressures beyond just lentiviral ones [5]. Other retroelements, such as Line-1, are also targeted by A3 proteins, with A3F showing greater effects in humans than A3G [61]. These extant and possibly extinct pressures likely drive the observed protein diversity across taxa. The distinct functional capacities of A3F and A3G, as shown in their species-dependent ability to reduce infectivity, further underline their evolutionary specialization [168]. Overall, CD1 diversity supports its broader functional requirements, while CD2 conservation reflects its critical catalytic role in maintaining sequence specificity and DNA-binding stability.

While innate TLRs and AID, which facilitate antibody creation, show that not all parts of the immune system are similarly pressured, the fact that SamHD1 (which depletes dNTPs) has less positive selection than other restriction factors shown here may denote that presence of antagonists is not the sole determinant of evolution (Figure S3) [29–31]. Instead, the positive selection may be due to both immuno-evasion and methods of action, which may place it in direct contact with the viral proteins or genome, as seen in the other restriction factors [20,26]. On the other hand, it is unsurprising that lentiviruses would have high diversity, especially as HIV-1 is noted for its high mutation rate [157]. Nonetheless, noticeably in HIV-1, higher mutation rates have been seen occurring in regions of Gag, Pol, and Env in comparison to restriction factors, which could be due to differences between a zoonotic virus attempting to adapt to a new host relative to an endemic virus resisting pressures in a current host (Figure S2). Fortunately, we were able to obtain a large repertoire of HIV/SIV and primate proteins to compare for these experiments. However, this has led to a limitation where there is high variance in the C-term of Vif alignments, as well as some proteins where we were not able to have sufficient alignment to determine percent identity and similarity. Notwithstanding, our results indicate selective pressures across restriction factors and their antagonists, both past and current, that have led to greater diversities than are commonly seen, even in comparison to other immune genes.

While linear sequences provide valuable insights, protein structure is essential for understanding host–virus interactions. Structural studies of A3FGs and Vif, though limited across species, have shown remarkable conservation in the core structure of these proteins, as evidenced by Figure 3B,C [84,127]. This aligns with findings that all AID/APOBEC family members exhibit a high degree of structural similarity, further supported by ASR 3D structural homology models (Figure 10E,F) [183,184]. The cytidine deaminase superfamily, which includes AID/APOBEC proteins, maintains a conserved core structure across various deaminases, such as guanine, guanosine, and adenosine deaminases, as well as dimeric and trimeric cytidine deaminases [46,48,185–193]. Within APOBECs, A3s are classified into Z1, Z2, and Z3 domains, highlighting structural similarities across lineages [80]. For example, A3F and A3G are primarily composed of Z2 domains, except for A3G-CD2, which belongs to Z1. Z2 domains can be further categorized into Z2a (A3F-CD1), Z2b (A3F-CD2), and Z2c (A3G-CD1). Structural comparisons reveal that the pseudocatalytic domains of A3F and A3G (Figures 6A and 7A) retain the APOBEC fold observed in active domains (Figures 6B and 7B). However, modifications in the loops and surface of CD1 disrupt the DNA-binding groove characteristic of catalytically active APOBECs [80,85,88,93–95,194,195]. These observations reinforce the evolutionary structural conservation of A3 proteins while highlighting domain-specific adaptations that may influence their functional roles in host–virus dynamics.

Unfortunately, in the case of Vif, there are very few known structures and none of HIV-2; as such, it is more surprising to us that HIV-2 is able to maintain this overall structure given it only retained around 50% sequence similarity [105,127]. Smith et al. 2014 showed that Vif of HIV-2 was able to bind CUL5, even when Zn was not present, and bind to a separate region of A3F and A3G [164]. As for the structures of many SIV-Vifs that infect OWMs, there is very little literature on their structure, save for the structure of SIVrcm-Vif, which has many similarities in overall structure with HIV-1 Vif in the regions that have been determined [106]. As Binning et al. stated, the variations in these unstructured looped regions could have effects on their targeting and would be important, along with a maintained protein fold; however, the disruption of this structure in SIVolc means there is a possibility of a highly variable Vif in this species.

The strength of our 3D structural homology modelling system is that it can accurately predict the structures of these proteins even with low percent identity; however, these are limited by the number of structures that have been previously found. This accuracy has been previously shown in our laboratory when we were able to accurately predict the structure of AID, which was later confirmed by X-ray crystallography [95,146]. As such, it limits our knowledge of the C-term of the protein which has not yet been fully elucidated. When we observe the charges of these proteins in Figures 8A–D and S5–S7, we see that both Vif and A3s have varied in their charges, but it is important to note that this is only considering the charges of the Vif-sensitive domains of A3FGs and all the regions that we can model using the 3D structural homology of Vif. Knowing that SIVs of OWMs have existed longer than those in GAs, as well as the known divergence between OWMs, it makes sense that charge ranges in these Vifs are highly divergent; however, while SIV of *Pan troglodytes* and *Gorilla gorilla* come from the same source, their low charge diversity also is understandable [6]. Intriguingly, the range in charges for HIVs may be due to the multiple lineages that were formed by four separate zoonotic events in the HIV-1s, while HIV-2 comes from nine events from sooty mangabeys. Furthermore, variance in the charges may indicate not only the necessity of adapting to host A3s but also other host mechanisms of the ubiquitination system. Conversely, A3FGs have ranges in their charges across all taxa that could be a result of selection from multiple viruses over evolutionary time, just as today they are known to interact with retroviruses, retrotransposons, HSV,

HBV, parvovirus, coronaviruses, and endogenous retroviruses [38,62]. This may explain the trends seen in Figure 8E–H across A3FGs, comparing their extant and ASR charges as different pressures occurred prior to what we know of as the evolution of HIV/SIV [49]. Our phylogenetic analysis employed multiple Bayesian runs rather than relying solely on maximum likelihood or maximum parsimony, enhancing the accuracy of ancestral sequence reconstruction (ASR) and 3D structural homology models. However, due to the extreme diversity of HIV/SIV Vif, we were unable to confidently derive ASR sequences for 3D modelling.

In analysing A3-Vif interactions (Figure 9), we observed that binding regions aligned with the established literature, though amino acid-level variations influenced the nature of these interactions [4,41,175]. This suggests that Vif's ability to bind A3s across species requires fine-tuned modifications, particularly at non-interacting proximal sites, to maintain functionality. However, it is of note that a limitation of the method we utilised is that completely blind docking is not possible. Regions of interactions must be chosen that could affect our results, which is why we reasoned that a larger aggregate approach would give a clearer picture of the data. Lentiviruses rely on these interactions to navigate zoonotic and interspecies barriers, including polymorphisms [53,156]. Our findings corroborate previous research that Vif interacts with A3F primarily through the α 2-3 cleft, though our data highlight a crucial role for CBF- β in stabilizing these interactions [127]. Without CBF- β , Vif compensates by stabilizing alternative binding sites. Interestingly, HIV/SIV-Vif can effectively bind A3G independently of CBF- β , maintaining stable interactions at known surfaces, which suggests reduced reliance on CBF- β for A3G binding [167]. In contrast, HIV-2 Vif targets distinct regions of A3F and A3G, although our results suggest these interactions may converge on similar binding locations [164]. For A3F, binding to modified regions on the first site may initially occur, but subsequent CBF- β binding could destabilize the interaction, preventing degradation. Conversely, HIV-2 Vif-A3G binding appears viable, albeit with diminished strength due to reduced interactions at Vif residues Y41 and H45. Interestingly, Vif also targets PPP2R5 to induce G2/M cell cycle arrest [196]. While this interaction spans a larger region of Vif than A3 binding, it shares four residues with A3F's binding site and three with A3G's. Although CBF- β binds Vif in this context, it does not directly interact with PPP2R5—similar to A3G binding but without requiring a nucleic acid intermediate. Notably, these shared residues are more evolutionarily conserved than their surroundings, suggesting that PPP2R5 binding may constrain Vif's adaptability in interacting with A3F and A3G.

Detailed structural observations further elucidate Vif-binding specificity. For A3F, Vif targets CD2 on two negatively charged and hydrophobic patches predominantly located on α 2', α 3', and α 4' helices [79–82,104,197]. These binding sites are comparable to those seen in A3C and A3H [80,81,123,198,199]. In A3G, Vif binds near what would have been the catalytic groove if the domain were active, unlike other APOBECs [156,167,198]. Across species, A3G exhibits significantly more adaptations in its Vif-binding regions than A3F, reflecting a broader divergence in A3G-binding sites (Figures 6C and 7C). While A3F-binding regions also show occasional adaptations, these occur less frequently compared to A3G [51,52,156,167,177,200]. This diversity highlights the evolutionary arms race between host A3 proteins and viral Vif, where both must adapt to sustain their respective functions. For A3F, this involves precise interactions with Vif in the α 2-3 cleft and stabilization through CBF- β , while A3G demonstrates a more autonomous binding capability with Vif, particularly in its catalytic groove-adjacent region. The variability in Vif-binding adaptations across species emphasizes the dynamic interplay of these proteins under selective pressures, ensuring the persistence of lentiviral survival strategies amidst diverse host defenses.

Our understanding of binding is currently hampered by lack of clear 3D structural homology models of the full-length Vif since all 3D structural homology models are missing the C-term in which particular parts of both Vif-binding regions are contained (Figure S8 and Tables S21 and S22) [175]. Furthermore, the lack of CBF- β across multiple species in A3F leads to a lack of information on other host–virus interactions seen in humans through 6NIL [127]. However, given the differences between our human 3D computational models and those of the CryoEM structure, it is very likely that these interfaces remain similar across all species. It was also discovered recently that Vif-A3G binding requires ssRNA for proper binding, which is also integral to the method used for A3G to be transported to the budding virion [65–67,172]. This could mean that some of the residues we noted on A3G ℓ 1 and Vif α 1 are not as interacted with as we thought and instead are used to support RNA above them during this vulnerable period for the APOBEC. Notwithstanding this, we created multiple bindings of A3s and Vifs across species thousands of times to understand the repertoires of interactions. Moreover, our results in aggregate clearly show high similarity of structure through evolution across highly divergent proteins with intriguing binding similarities and variances, denoting regional and residue interactions on the surfaces of binding partners (Figure 9). Thus, the shifting of some of these interactions to RNA present during incorporation could provide a much stronger binding among all of the other residues seen across this study. This evolutionary arms race is also complicated by the slow lifespans of primates that use A3s as barriers to zoonosis, while viral Vif is able to consistently adapt from individual to individual. It could be said that once a zoonotic disease takes hold, the arms race is already partially lost; however, A3s may still remain a barrier against other zoonotic viruses attempting replication as well as retrotransposons [5,201]. A3 proteins restrict diverse DNA and RNA viruses, including some that may no longer exist [38]. Recent findings reveal a unique interaction between Epstein-Barr Virus (EBV) BORF2 and APOBEC3B across its DNA-binding groove, underscoring the complex evolutionary pressures shaping A3 diversification [202].

The structural conservation of APOBEC3 (A3) proteins, particularly A3F and A3G, across species reveals significant evolutionary adaptations in their DNA-binding mechanisms. CD1 and CD2 domains exhibit distinct properties, with CD1 showing highly positive charges and structural variability, while CD2 maintains a conserved DNA-binding groove. These differences suggest functional specialization, where CD1 facilitates interactions with DNA and other APOBEC molecules, enhancing synergistic antiviral effects, while CD2 supports DNA movement across the groove. One of the reasons A3F and A3G may have kept an inactive domain was to allow for greater movement across DNA by jumping from region to region [17,35,77,79,86,88,91,182]. Key structural motifs highlight evolutionary conservation and variability. The NRPI motif in CD1, essential for dimerization through R24 and creating a positively charged surface, however, in both A3G-CD1 and A3F-CD1 among some primates, exists as a lysine or in NWMs and *Chlorocebus pygerythrus* as a glutamate [165,166]. A3F-CD2 regions exhibit highly negative patches (ℓ 4', α 3', α 4'), contrasting with positively charged ℓ 1' residues over the catalytic pocket, supporting DNA binding. This charge distribution significantly affects DNA interactions. CD1 remains highly positive across taxa, while CD2 is more negative where charge dichotomy likely reflects an adaptive balance to facilitate DNA binding and catalysis. For instance, A3F-CD1 shows enhanced positive interactions attracting DNA phosphates, while A3G-CD1 employs hydrophobic residues in already polar regions, demonstrating domain-specific functional optimization.

The binding preference for specific DNA motifs reflects structural adaptations. A3F interacts predominantly with TTC motifs through hydrophobic π - π interactions, while A3G binds CCC motifs via polar and charged interactions between DNA bases and amino acids. Notably, the substrate specificity loop ℓ 7' (307YYFW310 in A3F; 315YDDQ318 in

A3G) is critical for maintaining motif recognition and DNA binding. All the while, the 305RL306 of A3F and 313RI314 of A3G maintain the structure of this loop and connect it with $\ell 1'$. These conserved interactions underscore the importance of these loops in regulating enzymatic activity and substrate specificity. Across all species, $\ell 1'$, $\ell 3'$, $\ell 5'$, and $\ell 7'$ loops dominate DNA interactions, contributing over 75% of contacts in both proteins (Figures 10B,D and S8, and Tables S23–S26) [80,82,88–90]. The loops form the DNA-binding groove with $\ell 1'$ shaping the U-like groove that stabilizes DNA binding and are critical for the efficiency of A3's enzymatic mechanism. Specific residues, such as R213/R215 of A3F and A3G and the adjacent pocket gatekeeper positions (214/216), play structural roles by supporting the loop with the backbone of $\ell 1'$ and $\alpha 1'$ and enabling interactions with DNA backbones over the pocket, respectively [80,82,88–90,93–95,123,142,194,195,203–205]. Variations in these residues, such as asparagine in A3F and histidine in A3G, adapt domain functionality to distinct DNA-binding needs. The α -helices, though less involved than loops, provide critical structural support. For example, $\alpha 4'$ in A3F enables additional π - π interactions, stabilizing the sandwiching of π - π bonding in $\ell 7'$, while $\alpha 6'$ at position R374 in A3G interacts strongly with DNA phosphates, anchoring the DNA at the 5' end. These helices complement the loops, enhancing DNA stabilization and catalytic efficiency.

APOBEC3 proteins exhibit extensive structural conservation, albeit positionally diverse, yet they retain flexibility to adapt to diverse functions. The linker region between domains, which comprises eight residues in A3F and five residues in A3G, allows for dynamic conformational changes, as seen in Gorle et al. [91]. This flexibility enables catalytic pockets to adopt multiple orientations, facilitating diverse interactions with nucleotides, RNA, and reverse transcriptase. Homology models and previous studies reveal three major conformations: both catalytic pockets aligned, pockets in opposite orientations, or intermediate states [65,66,91,165,166,172]. These conformations impact DNA-binding efficiency and interactions. For example, double-domain APOBECs often position catalytic pockets near each other, creating a dual-binding interface for DNA or RNA. This configuration enhances interactions at key motifs, such as the 24RPIL28 region in A3G-CD1, which plays a role in virion packaging and catalysis [65,165,166]. Variations in this motif, such as glutamate or lysine adaptations in A3F of New World monkeys, and Maiti et al.'s substitution of R24A, revealed changes in RNA interactions occur in this region as well as reduced catalysis. The APOBECs during the uncoating and transcription phase of the lentiviral life cycle have the potential of interacting with reverse transcriptase, RNA, and DNA, which given double-domain APOBEC's ability for dimerization and intersegmental transfer, means it may interact with multiple nucleotide chains or their regions at a time [63,165]. In the double-domain structures for which we have modelled 3D structural homology, we can see the possibility of a second route across the β -sheet of the CD1 domains, as the CD1 pocket and groove are blocked behind loop 3. While in this conformation, DNA could interact across this region which maintains similar structural surfaces and relative charges.

Docking studies further illustrate functional conservation despite structural variations. Preferred DNA motifs (5'-*taTTCat* for A3F and 5'-*ttCCCtt* for A3G) consistently position specific nucleotides within the catalytic pocket, particularly at position 4. Although docking algorithms differ in predicted nucleotide stability, the conserved binding patterns across species highlight the robustness of APOBEC-DNA interactions. However, it is also of note that due to the degrees of freedom that occur in DNA as compared to other small molecules, certain conformations may occur more often due to algorithmic scalability issues. Differences in catalytic efficiency between A3F and A3G align with these observations, with A3G displaying greater nucleotide and backbone interactions. The loops and helices collectively maintain the structural integrity and functionality of A3F and A3G. In $\ell 7'$, hydrophobic residues in A3F stabilize interactions with *TTC* motifs, while hydrophilic

residues in A3G facilitate binding to CCC motifs. These substrate-specific adaptations illustrate the evolutionary tuning of APOBEC3 proteins to optimize their antiviral activities. Interactions at $\ell 1'$ further stabilize the DNA-binding groove, with positive residues in A3F enhancing phosphate interactions, and hydrophobic residues in A3G contributing to overall stabilization. Comparative analyses of $\alpha 1'$, $\alpha 2'$, and $\alpha 3'$ reveal their role in shaping the catalytic pocket and stabilizing DNA. For example, $\alpha 1'$ residues in A3F engage in hydrophobic and π -bonding interactions, while planar and polar residues in A3G strengthen DNA stabilization. $\alpha 2'$ and $\alpha 3'$ are integral in their shaping of the catalytic pocket. However, $\alpha 4'$ of A3F Y314 allowed for extra π - π bonding with the nucleotides stabilizing the sandwiching of π - π bonding in $\ell 7'$. These interactions are complemented by $\alpha 6'$ residues such as R374 in A3G, which anchor DNA through phosphate interactions, underscoring the helices' supporting role in DNA binding. The observed structural variations across species highlight APOBEC3 proteins' evolutionary adaptability while preserving critical functions. For example, the conformational flexibility enabled by the linker region allows APOBECs to interact with multiple nucleotide chains or different regions simultaneously, enhancing their antiviral efficacy. Structural models also suggest alternative DNA-binding routes when catalytic grooves are blocked, ensuring continued DNA processing. Despite structural shifts, conserved motifs, charge distributions, and substrate-binding loops demonstrate the evolutionary pressures to maintain functionality. This balance of flexibility and conservation enables APOBEC3 proteins to efficiently bind and process DNA, contributing to their role in antiviral defense.

5. Conclusions

Our study contributes significantly to the growing field of host–viral coevolution by providing insights into the evolutionary and structural dynamics of A3F, A3G, and Vif proteins. As the first research to construct extensive 3D structural homology models for these proteins across species, it offers a unique perspective on their sequential and three-dimensional evolution. By integrating structural modelling with ancestral sequence reconstruction (ASR), we uncover shifts in extinct proteins, broadening our understanding of the evolutionary arms race between restriction factors and their viral antagonists. The diversity observed in A3F and A3G, particularly in their sequence alignments and structural adaptations, reveals a clear evolutionary strategy. While CD1 exhibits greater sequence diversity than CD2, its higher positive charge facilitates binding nucleotide chains for inter-segmental transfer and virion packaging. Despite this diversity, the DNA-binding grooves across taxa maintain structural and functional consistency, with conserved residues or their equivalents playing critical roles in DNA interactions. Notably, the greatest variation between A3F and A3G lies in $\ell 7'$, where hydrophobic and π - π interactions dominate in A3F, while A3G relies on polar and charged interactions. The substrate-binding grooves remain evolutionarily stable across taxa, suggesting an essential role in protein functionality across evolutionary lineages. Our findings underscore the importance of examining both structural and evolutionary aspects to understand the mechanisms driving host–virus dynamics. Future studies should expand sequence data and structural models for restriction factors, viral antagonists, and accessory proteins across more species. Investigating these proteins' coevolution will enhance our knowledge of immune system development and inform predictions about viral adaptation and zoonotic risks. This work represents a novel approach to studying protein evolution by combining structural and sequence analyses, paving the way for deeper insights into coevolutionary processes and their implications for human health.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v17030393/s1>, Figure S1: Amino acid sequence logos of Different HIV lineages; Figure S2: Identity and Similarity of Viral proteins; Figure S3: Identity and Similarity of Host proteins; Figure S4: A3G Full Alignment; Figure S5: Charged surfaces of Human A3FG and HIV Vif; Figure S6: Surface charge of Vif; Figure S7: Charges of A3FG across species and domains; Figure S8: A3FG and Vif CryoEM binding surfaces; Figure S9: Structures of ancestral sequence reconstruction; Figure S10: Dual-domain APOBEC3FG from humans and ASR; Figure S11: A3FGs host phylogeny; Figure S12: Viral Vif phylogenies; Table S1: Sequences of A3F across Species from NCBI or Ensemble; Table S2: Sequences of A3G across Species from NCBI; Table S3: Sequences of full Lentiviral genomes; Table S4: A3F-CD1 Models created through I-TASSER, Zn Coordinates, Templates, Confidence Scores and Ramachandran Plot information; Table S5: A3F-CD2 Models created through I-TASSER, Zn Coordinates, Templates, Confidence Scores and Ramachandran Plot information; Table S6: A3G-CD1 Models created through I-TASSER, Zn Coordinates, Templates, Confidence Scores and Ramachandran Plot information; Table S7: A3G-CD2 Models created through I-TASSER, Zn Coordinates, Templates, Confidence Scores and Ramachandran Plot information; Table S8: Ancestral A3F-CD1 Models created through I-TASSER, Zn Coordinates, Templates, Confidence Scores and Ramachandran Plot information; Table S9: Ancestral A3F-CD2 Models created through I-TASSER, Zn Coordinates, Templates, Confidence Scores and Ramachandran Plot information; Table S10: Ancestral A3G-CD1 Models created through I-TASSER, Zn Coordinates, Templates, Confidence Scores and Ramachandran Plot information; Table S11: Ancestral A3G-CD2 Models created through I-TASSER, Zn Coordinates, Templates, Confidence Scores and Ramachandran Plot information; Table S12: HIV/SIV Vif Models created through I-TASSER, Zn Coordinates, Templates, Confidence Scores and Ramachandran Plot information; Table S13: A3F-Identity and Similarity dependent on Most Likely Ancestral Sequence Reconstruction; Table S14: A3G-Identity and Similarity dependent on Most Likely Ancestral Sequence Reconstruction; Table S15: Interactions of A3F2 on Vif by secondary structure; Table S16: Interactions of Vif on A3F2 by secondary structure; Table S17: Interactions of A3G1 on Vif by secondary structure; Table S18: Interactions of Vif on A3G1 by secondary structure; Table S19: Interactions of A3F2 on Vif by Amino Acid; Table S20: Interactions of Vif on A3F2 by Amino Acid; Table S21: Interactions of A3G1 on Vif by Amino Acid; Table S22: Interactions of Vif on A3G1 by Amino Acid; Table S23: Overall interactions of A3F with *taTTCat* against different species across Secondary Structures; Table S24: Interactions across the species of A3F with individual Amino Acids; Table S25: Overall interactions of A3G with *tCCCt* against different species across Secondary Structures; Table S26: Interactions across the species of A3G with individual Amino Acids.

Author Contributions: Conceptualization, D.N.G.H. and M.L.; methodology, D.N.G.H.; software, D.N.G.H.; validation, D.N.G.H., A.G. and M.L.; formal analysis, D.N.G.H., A.G. and S.Y.B.L.; data curation, D.N.G.H.; writing—original draft preparation, D.N.G.H.; writing—review and editing, D.N.G.H. and M.L.; visualization, D.N.G.H.; supervision, M.L.; project administration, D.N.G.H.; funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Canadian Institute of Health Research (CIHR), grant number MOP111132, and the International Development Research Center (IDRC), grant number 108405001. The Natural Sciences and Engineering Research Council of Canada provides a discovery grant to M.L. The Leukemia & Lymphoma Society of Canada (LLSC) provides an operating grant to M.L., and supported D.H. with a fellowship, 2018–2021.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, Mani Larijani, upon reasonable request.

Acknowledgments: The authors are grateful to the aid of Gladdale Huebert as well as the other members of the Larijani Lab for editing this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

APOBEC	Apolipoprotein B mRNA editing polypeptide like
A3F	APOBEC3F
A3G	APOBEC3G
Vif	Viral infectivity factor
GAs	Great apes
OWMs	Old World monkeys
NWMs	New World monkeys

References

1. Kluge, S.F.; Sauter, D.; Kirchhoff, F. SnapShot: Antiviral restriction factors. *Cell* **2015**, *163*, 774–774 e1. [[CrossRef](#)] [[PubMed](#)]
2. Kirmaier, A.; Krupp, A.; Johnson, W.E. Understanding restriction factors and intrinsic immunity: Insights and lessons from the primate lentiviruses. *Future Virol.* **2014**, *9*, 483–497. [[CrossRef](#)] [[PubMed](#)]
3. Colomer-Lluch, M.; Ruiz, A.; Moris, A.; Prado, J.G. Restriction Factors: From Intrinsic Viral Restriction to Shaping Cellular Immunity Against HIV-1. *Front. Immunol.* **2018**, *9*, 2876. [[CrossRef](#)] [[PubMed](#)]
4. Delviks-Frankenberry, K.A.; Desimmie, B.A.; Pathak, V.K. Structural Insights into APOBEC3-Mediated Lentiviral Restriction. *Viruses* **2020**, *12*, 587. [[CrossRef](#)]
5. Nakano, Y.; Aso, H.; Soper, A.; Yamada, E.; Moriwaki, M.; Juarez-Fernandez, G.; Koyanagi, Y.; Sato, K. A conflict of interest: The evolutionary arms race between mammalian APOBEC3 and lentiviral Vif. *Retrovirology* **2017**, *14*, 31. [[CrossRef](#)]
6. Sauter, D.; Kirchhoff, F. Key Viral Adaptations Preceding the AIDS Pandemic. *Cell Host Microbe* **2019**, *25*, 27–38. [[CrossRef](#)]
7. Song, B.; Javanbakht, H.; Perron, M.; Park, D.H.; Strelau, M.; Sodroski, J. Retrovirus restriction by TRIM5alpha variants from Old World and New World primates. *J. Virol.* **2005**, *79*, 3930–3937. [[CrossRef](#)]
8. Marx, P.A.; Apetrei, C.; Drucker, E. AIDS as a zoonosis? Confusion over the origin of the virus and the origin of the epidemics. *J. Med. Primatol.* **2004**, *33*, 220–226. [[CrossRef](#)]
9. Sharp, P.M.; Hahn, B.H. Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* **2011**, *1*, a006841. [[CrossRef](#)]
10. Ahuka-Mundeye, S.; Ayoub, A.; Mbala-Kingebeni, P.; Liegeois, F.; Esteban, A.; Lunguya-Metila, O.; Demba, D.; Bilulu, G.; Mbenzo-Abokome, V.; Inogwabini, B.I.; et al. Novel multiplexed HIV/simian immunodeficiency virus antibody detection assay. *Emerg. Infect. Dis.* **2011**, *17*, 2277–2286. [[CrossRef](#)]
11. Faria, N.R.; Rambaut, A.; Suchard, M.A.; Baele, G.; Bedford, T.; Ward, M.J.; Tatem, A.J.; Sousa, J.D.; Arinaminpathy, N.; Pepin, J.; et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **2014**, *346*, 56–61. [[CrossRef](#)] [[PubMed](#)]
12. Klatt, N.R.; Silvestri, G.; Hirsch, V. Nonpathogenic simian immunodeficiency virus infections. *Cold Spring Harb. Perspect. Med.* **2012**, *2*, a007153. [[CrossRef](#)] [[PubMed](#)]
13. Siriwardena, S.U.; Chen, K.; Bhagwat, A.S. Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases. *Chem. Rev.* **2016**, *116*, 12688–12710. [[CrossRef](#)] [[PubMed](#)]
14. Meulendyke, K.A.; Croteau, J.D.; Zink, M.C. HIV life cycle, innate immunity and autophagy in the central nervous system. *Curr. Opin. HIV AIDS* **2014**, *9*, 565–571. [[CrossRef](#)]
15. Hultquist, J.F.; Lengyel, J.A.; Refsland, E.W.; LaRue, R.S.; Lackey, L.; Brown, W.L.; Harris, R.S. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J. Virol.* **2011**, *85*, 11220–11234. [[CrossRef](#)]
16. Song, C.; Sutton, L.; Johnson, M.E.; D’Aquila, R.T.; Donahue, J.P. Signals in APOBEC3F N-terminal and C-terminal deaminase domains each contribute to encapsidation in HIV-1 virions and are both required for HIV-1 restriction. *J. Biol. Chem.* **2012**, *287*, 16965–16974. [[CrossRef](#)]
17. Feng, Y.; Baig, T.T.; Love, R.P.; Chelico, L. Suppression of APOBEC3-mediated restriction of HIV-1 by Vif. *Front. Microbiol.* **2014**, *5*, 450. [[CrossRef](#)]
18. Feng, Y.; Love, R.P.; Ara, A.; Baig, T.T.; Adolph, M.B.; Chelico, L. Natural Polymorphisms and Oligomerization of Human APOBEC3H Contribute to Single-stranded DNA Scanning Ability. *J. Biol. Chem.* **2015**, *290*, 27188–27203. [[CrossRef](#)]
19. Pereira, E.A.; daSilva, L.L. HIV-1 Nef: Taking Control of Protein Trafficking. *Traffic* **2016**, *17*, 976–996. [[CrossRef](#)]

20. Harris, R.S.; Hultquist, J.F.; Evans, D.T. The restriction factors of human immunodeficiency virus. *J. Biol. Chem.* **2012**, *287*, 40875–40883. [[CrossRef](#)]
21. Mack, K.; Starz, K.; Sauter, D.; Langer, S.; Bibollet-Ruche, F.; Learn, G.H.; Sturzel, C.M.; Leoz, M.; Plantier, J.C.; Geyer, M.; et al. Efficient Vpu-Mediated Tetherin Antagonism by an HIV-1 Group O Strain. *J. Virol.* **2017**, *91*, 20. [[CrossRef](#)] [[PubMed](#)]
22. Firrito, C.; Bertelli, C.; Vanzo, T.; Chande, A.; Pizzato, M. SERINC5 as a New Restriction Factor for Human Immunodeficiency Virus and Murine Leukemia Virus. *Annu. Rev. Virol.* **2018**, *5*, 323–340. [[CrossRef](#)] [[PubMed](#)]
23. Ghimire, D.; Rai, M.; Gaur, R. Novel host restriction factors implicated in HIV-1 replication. *J. Gen. Virol.* **2018**, *99*, 435–446. [[CrossRef](#)] [[PubMed](#)]
24. Usami, Y.; Wu, Y.; Gottlinger, H.G. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature* **2015**, *526*, 218–223. [[CrossRef](#)]
25. Rosa, A.; Chande, A.; Ziglio, S.; De Sanctis, V.; Bertorelli, R.; Goh, S.L.; McCauley, S.M.; Nowosielska, A.; Antonarakis, S.E.; Luban, J.; et al. HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature* **2015**, *526*, 212–217. [[CrossRef](#)]
26. Zheng, Y.H.; Jeang, K.T.; Tokunaga, K. Host restriction factors in retroviral infection: Promises in virus-host interaction. *Retrovirology* **2012**, *9*, 112. [[CrossRef](#)]
27. Jia, X.; Zhao, Q.; Xiong, Y. HIV suppression by host restriction factors and viral immune evasion. *Curr. Opin. Struct. Biol.* **2015**, *31*, 106–114. [[CrossRef](#)]
28. Wu, F.; Ourmanov, I.; Riddick, N.; Matsuda, K.; Whitted, S.; Plishka, R.J.; Buckler-White, A.; Starost, M.F.; Hirsch, V.M. TRIM5alpha restriction affects clinical outcome and disease progression in simian immunodeficiency virus-infected rhesus macaques. *J. Virol.* **2015**, *89*, 2233–2240. [[CrossRef](#)]
29. Zhang, Z.; Zheng, L.; Yu, Y.; Wu, J.; Yang, F.; Xu, Y.; Guo, Q.; Wu, X.; Cao, S.; Cao, L.; et al. Involvement of SAMHD1 in dNTP homeostasis and the maintenance of genomic integrity and oncotherapy (Review). *Int. J. Oncol.* **2020**, *56*, 879–888. [[CrossRef](#)]
30. Coggins, S.A.; Mahboubi, B.; Schinazi, R.F.; Kim, B. SAMHD1 Functions and Human Diseases. *Viruses* **2020**, *12*, 382. [[CrossRef](#)]
31. Chen, S.; Bonifati, S.; Qin, Z.; St Gelais, C.; Wu, L. SAMHD1 Suppression of Antiviral Immune Responses. *Trends Microbiol.* **2019**, *27*, 254–267. [[CrossRef](#)] [[PubMed](#)]
32. Harris, R.S.; Dudley, J.P. APOBECs and virus restriction. *Virology* **2015**, *479–480*, 131–145. [[CrossRef](#)] [[PubMed](#)]
33. Borzooee, F.; Asgharpour, M.; Quinlan, E.; Grant, M.D.; Larijani, M. Viral subversion of APOBEC3s: Lessons for anti-tumor immunity and tumor immunotherapy. *Int. Rev. Immunol.* **2018**, *37*, 151–164. [[CrossRef](#)] [[PubMed](#)]
34. Olson, M.E.; Harris, R.S.; Harki, D.A. APOBEC Enzymes as Targets for Virus and Cancer Therapy. *Cell Chem. Biol.* **2018**, *25*, 36–49. [[CrossRef](#)]
35. Morse, M.; Huo, R.; Feng, Y.; Rouzina, I.; Chelico, L.; Williams, M.C. Dimerization regulates both deaminase-dependent and deaminase-independent HIV-1 restriction by APOBEC3G. *Nat. Commun.* **2017**, *8*, 597. [[CrossRef](#)]
36. Chelico, L.; Pham, P.; Calabrese, P.; Goodman, M.F. APOBEC3G DNA deaminase acts processively 3' → 5' on single-stranded DNA. *Nat. Struct. Mol. Biol.* **2006**, *13*, 392–399. [[CrossRef](#)]
37. Monajemi, M.; Woodworth, C.F.; Zipperlen, K.; Gallant, M.; Grant, M.D.; Larijani, M. Positioning of APOBEC3G/F mutational hotspots in the human immunodeficiency virus genome favors reduced recognition by CD8+ T cells. *PLoS ONE* **2014**, *9*, e93428. [[CrossRef](#)]
38. Milewska, A.; Kindler, E.; Vkovski, P.; Zeglen, S.; Ochman, M.; Thiel, V.; Rajfur, Z.; Pyrc, K. APOBEC3-mediated restriction of RNA virus replication. *Sci. Rep.* **2018**, *8*, 5960. [[CrossRef](#)]
39. Renner, T.M.; Belanger, K.; Goodwin, L.R.; Campbell, M.; Langlois, M.A. Characterization of molecular attributes that influence LINE-1 restriction by all seven human APOBEC3 proteins. *Virology* **2018**, *520*, 127–136. [[CrossRef](#)]
40. OhAinle, M.; Kerns, J.A.; Li, M.M.; Malik, H.S.; Emerman, M. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* **2008**, *4*, 249–259. [[CrossRef](#)]
41. Wang, J.; Shaban, N.M.; Land, A.M.; Brown, W.L.; Harris, R.S. Simian Immunodeficiency Virus Vif and Human APOBEC3B Interactions Resemble Those between HIV-1 Vif and Human APOBEC3G. *J. Virol.* **2018**, *92*, 14. [[CrossRef](#)] [[PubMed](#)]
42. Shapiro, M.; Meier, S.; MacCarthy, T. The cytidine deaminase under-representation reporter (CDUR) as a tool to study evolution of sequences under deaminase mutational pressure. *BMC Bioinform.* **2018**, *19*, 163. [[CrossRef](#)]
43. Jarmuz, A.; Chester, A.; Bayliss, J.; Gisbourne, J.; Dunham, I.; Scott, J.; Navaratnam, N. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **2002**, *79*, 285–296. [[CrossRef](#)] [[PubMed](#)]
44. Harris, R.S.; Liddament, M.T. Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.* **2004**, *4*, 868–877. [[CrossRef](#)]
45. Holland, S.J.; Berghuis, L.M.; King, J.J.; Iyer, L.M.; Sikora, K.; Fifield, H.; Peter, S.; Quinlan, E.M.; Sugahara, F.; Shingate, P.; et al. Expansions, diversification, and interindividual copy number variations of AID/APOBEC family cytidine deaminase genes in lampreys. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E3211–E3220. [[CrossRef](#)]
46. Conticello, S.G.; Langlois, M.A.; Yang, Z.; Neuberger, M.S. DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv. Immunol.* **2007**, *94*, 37–73. [[CrossRef](#)]

47. Barreto, V.M.; Magor, B.G. Activation-induced cytidine deaminase structure and functions: A species comparative view. *Dev. Comp. Immunol.* **2011**, *35*, 991–1007. [[CrossRef](#)]
48. Conticello, S.G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **2008**, *9*, 229. [[CrossRef](#)]
49. Sawyer, S.L.; Emerman, M.; Malik, H.S. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* **2004**, *2*, E275. [[CrossRef](#)]
50. Bishop, K.N.; Holmes, R.K.; Sheehy, A.M.; Davidson, N.O.; Cho, S.J.; Malim, M.H. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr. Biol.* **2004**, *14*, 1392–1396. [[CrossRef](#)]
51. Compton, A.A.; Emerman, M. Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses. *PLoS Pathog.* **2013**, *9*, e1003135. [[CrossRef](#)] [[PubMed](#)]
52. Compton, A.A.; Hirsch, V.M.; Emerman, M. The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. *Cell Host Microbe* **2012**, *11*, 91–98. [[CrossRef](#)] [[PubMed](#)]
53. Krupp, A.; McCarthy, K.R.; Ooms, M.; Letko, M.; Morgan, J.S.; Simon, V.; Johnson, W.E. APOBEC3G polymorphism as a selective barrier to cross-species transmission and emergence of pathogenic SIV and AIDS in a primate host. *PLoS Pathog.* **2013**, *9*, e1003641. [[CrossRef](#)] [[PubMed](#)]
54. 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*, 11925–11940. [[CrossRef](#)]
55. Chen, H.; Lilley, C.E.; Yu, Q.; Lee, D.V.; Chou, J.; Narvaiza, I.; Landau, N.R.; Weitzman, M.D. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.* **2006**, *16*, 480–485. [[CrossRef](#)]
56. Gallois-Montbrun, S.; Kramer, B.; Swanson, C.M.; Byers, H.; Lynham, S.; Ward, M.; Malim, M.H. Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules. *J. Virol.* **2007**, *81*, 2165–2178. [[CrossRef](#)]
57. Ito, S.; Nagaoka, H.; Shinkura, R.; Begum, N.; Muramatsu, M.; Nakata, M.; Honjo, T. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1975–1980. [[CrossRef](#)]
58. Wichroski, M.J.; Robb, G.B.; Rana, T.M. Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies. *PLoS Pathog.* **2006**, *2*, e41. [[CrossRef](#)]
59. Bouzidi, M.S.; Caval, V.; Suspene, R.; Hallez, C.; Pineau, P.; Wain-Hobson, S.; Vartanian, J.P. APOBEC3DE Antagonizes Hepatitis B Virus Restriction Factors APOBEC3F and APOBEC3G. *J. Mol. Biol.* **2016**, *428*, 3514–3528. [[CrossRef](#)]
60. Liang, W.; Xu, J.; Yuan, W.; Song, X.; Zhang, J.; Wei, W.; Yu, X.F.; Yang, Y. APOBEC3DE Inhibits LINE-1 Retrotransposition by Interacting with ORF1p and Influencing LINE Reverse Transcriptase Activity. *PLoS ONE* **2016**, *11*, e0157220. [[CrossRef](#)]
61. Feng, Y.; Goubran, M.H.; Follack, T.B.; Chelico, L. Deamination-independent restriction of LINE-1 retrotransposition by APOBEC3H. *Sci. Rep.* **2017**, *7*, 10881. [[CrossRef](#)] [[PubMed](#)]
62. Ito, J.; Gifford, R.J.; Sato, K. Retroviruses drive the rapid evolution of mammalian APOBEC3 genes. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 610–618. [[CrossRef](#)] [[PubMed](#)]
63. Adolph, M.B.; Love, R.P.; Chelico, L. Biochemical Basis of APOBEC3 Deoxycytidine Deaminase Activity on Diverse DNA Substrates. *ACS Infect. Dis.* **2018**, *4*, 224–238. [[CrossRef](#)] [[PubMed](#)]
64. Yang, L.; Emerman, M.; Malik, H.S.; McLaughlin, R.N.J. Retrocopying expands the functional repertoire of APOBEC3 antiviral proteins in primates. *eLife* **2020**, *9*, e58436. [[CrossRef](#)]
65. Yang, H.; Kim, K.; Li, S.; Pacheco, J.; Chen, X.S. Structural basis of sequence-specific RNA recognition by the antiviral factor APOBEC3G. *Nat. Commun.* **2022**, *13*, 7498. [[CrossRef](#)]
66. Ito, F.; Alvarez-Cabrera, A.L.; Liu, S.; Yang, H.; Shiriaeva, A.; Zhou, Z.H.; Chen, X.S. Structural basis for HIV-1 antagonism of host APOBEC3G via Cullin E3 ligase. *Sci. Adv.* **2023**, *9*, eade3168. [[CrossRef](#)]
67. Kouno, T.; Shibata, S.; Shigematsu, M.; Hyun, J.; Kim, T.G.; Matsuo, H.; Wolf, M. Structural insights into RNA bridging between HIV-1 Vif and antiviral factor APOBEC3G. *Nat. Commun.* **2023**, *14*, 4037. [[CrossRef](#)]
68. Adolph, M.B.; Webb, J.; Chelico, L. Retroviral restriction factor APOBEC3G delays the initiation of DNA synthesis by HIV-1 reverse transcriptase. *PLoS ONE* **2013**, *8*, e64196. [[CrossRef](#)]
69. Ziegler, S.J.; Liu, C.; Landau, M.; Buzovetsky, O.; Desimmi, B.A.; Zhao, Q.; Sasaki, T.; Burdick, R.C.; Pathak, V.K.; Anderson, K.S.; et al. Insights into DNA substrate selection by APOBEC3G from structural, biochemical, and functional studies. *PLoS ONE* **2018**, *13*, e0195048. [[CrossRef](#)]
70. Wensing, A.M.; Calvez, V.; Gunthard, H.F.; Johnson, V.A.; Paredes, R.; Pillay, D.; Shafer, R.W.; Richman, D.D. 2017 Update of the Drug Resistance Mutations in HIV-1. *Top. Antivir. Med.* **2017**, *24*, 132–133. [[CrossRef](#)]
71. Nisole, S.; Saib, A. Early steps of retrovirus replicative cycle. *Retrovirology* **2004**, *1*, 9. [[CrossRef](#)] [[PubMed](#)]
72. Feng, Y.; Love, R.P.; Chelico, L. HIV-1 viral infectivity factor (Vif) alters processive single-stranded DNA scanning of the retroviral restriction factor APOBEC3G. *J. Biol. Chem.* **2013**, *288*, 6083–6094. [[CrossRef](#)] [[PubMed](#)]
73. Marin, M.; Rose, K.M.; Kozak, S.L.; Kabat, D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat. Med.* **2003**, *9*, 1398–1403. [[CrossRef](#)]

74. Mariani, R.; Chen, D.; Schrofelbauer, B.; Navarro, F.; Konig, R.; Bollman, B.; Munk, C.; Nymark-McMahon, H.; Landau, N.R. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* **2003**, *114*, 21–31. [[CrossRef](#)] [[PubMed](#)]
75. Schrofelbauer, B.; Chen, D.; Landau, N.R. A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3927–3932. [[CrossRef](#)]
76. Ara, A.; Love, R.P.; Chelico, L. Different mutagenic potential of HIV-1 restriction factors APOBEC3G and APOBEC3F is determined by distinct single-stranded DNA scanning mechanisms. *PLoS Pathog.* **2014**, *10*, e1004024. [[CrossRef](#)]
77. Ara, A.; Love, R.P.; Follack, T.B.; Ahmed, K.A.; Adolph, M.B.; Chelico, L. Mechanism of Enhanced HIV Restriction by Virion Coencapsidated Cytidine Deaminases APOBEC3F and APOBEC3G. *J. Virol.* **2017**, *91*, 19. [[CrossRef](#)]
78. Feng, Y.; Chelico, L. Intensity of deoxycytidine deamination of HIV-1 proviral DNA by the retroviral restriction factor APOBEC3G is mediated by the noncatalytic domain. *J. Biol. Chem.* **2011**, *286*, 11415–11426. [[CrossRef](#)]
79. Bohn, M.F.; Shandilya, S.M.; Albin, J.S.; Kouno, T.; Anderson, B.D.; McDougale, R.M.; Carpenter, M.A.; Rathore, A.; Evans, L.; Davis, A.N.; et al. Crystal structure of the DNA cytosine deaminase APOBEC3F: The catalytically active and HIV-1 Vif-binding domain. *Structure* **2013**, *21*, 1042–1050. [[CrossRef](#)]
80. Siu, K.K.; Sultana, A.; Azimi, F.C.; Lee, J.E. Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. *Nat. Commun.* **2013**, *4*, 2593. [[CrossRef](#)]
81. Nakashima, M.; Ode, H.; Kawamura, T.; Kitamura, S.; Naganawa, Y.; Awazu, H.; Tsuzuki, S.; Matsuoka, K.; Nemoto, M.; Hachiya, A.; et al. Structural Insights into HIV-1 Vif-APOBEC3F Interaction. *J. Virol.* **2016**, *90*, 1034–1047. [[CrossRef](#)] [[PubMed](#)]
82. Fang, Y.; Xiao, X.; Li, S.X.; Wolfe, A.; Chen, X.S. Molecular Interactions of a DNA Modifying Enzyme APOBEC3F Catalytic Domain with a Single-Stranded DNA. *J. Mol. Biol.* **2018**, *430*, 87–101. [[CrossRef](#)] [[PubMed](#)]
83. Kouno, T.; Luengas, E.M.; Shigematsu, M.; Shandilya, S.M.; Zhang, J.; Chen, L.; Hara, M.; Schiffer, C.A.; Harris, R.S.; Matsuo, H. Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G. *Nat. Struct. Mol. Biol.* **2015**, *22*, 485–491. [[CrossRef](#)] [[PubMed](#)]
84. Xiao, X.; Li, S.X.; Yang, H.; Chen, X.S. Crystal structures of APOBEC3G N-domain alone and its complex with DNA. *Nat. Commun.* **2016**, *7*, 12193. [[CrossRef](#)]
85. Holden, L.G.; Prochnow, C.; Chang, Y.P.; Bransteitter, R.; Chelico, L.; Sen, U.; Stevens, R.C.; Goodman, M.F.; Chen, X.S. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **2008**, *456*, 121–124. [[CrossRef](#)]
86. Shandilya, S.M.; Nalam, M.N.; Nalivaika, E.A.; Gross, P.J.; Valesano, J.C.; Shindo, K.; Li, M.; Munson, M.; Royer, W.E.; Harjes, E.; et al. Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure* **2010**, *18*, 28–38. [[CrossRef](#)]
87. Shandilya, S.M.; Bohn, M.F.; Schiffer, C.A. A computational analysis of the structural determinants of APOBEC3's catalytic activity and vulnerability to HIV-1 Vif. *Virology* **2014**, *471–473*, 105–116. [[CrossRef](#)]
88. Lu, X.; Zhang, T.; Xu, Z.; Liu, S.; Zhao, B.; Lan, W.; Wang, C.; Ding, J.; Cao, C. Crystal structure of DNA cytosine deaminase ABOBEC3G catalytic deamination domain suggests a binding mode of full-length enzyme to single-stranded DNA. *J. Biol. Chem.* **2015**, *290*, 4010–4021. [[CrossRef](#)]
89. Maiti, A.; Myint, W.; Kanai, T.; Delviks-Frankenberry, K.; Sierra Rodriguez, C.; Pathak, V.K.; Schiffer, C.A.; Matsuo, H. Crystal structure of the catalytic domain of HIV-1 restriction factor APOBEC3G in complex with ssDNA. *Nat. Commun.* **2018**, *9*, 2460. [[CrossRef](#)]
90. Wan, L.; Nagata, T.; Katahira, M. Influence of the DNA sequence/length and pH on deaminase activity, as well as the roles of the amino acid residues around the catalytic center of APOBEC3F. *Phys. Chem. Chem. Phys.* **2018**, *20*, 3109–3117. [[CrossRef](#)]
91. Gorle, S.; Pan, Y.; Sun, Z.; Shlyakhtenko, L.S.; Harris, R.S.; Lyubchenko, Y.L.; Vukovic, L. Computational Model and Dynamics of Monomeric Full-Length APOBEC3G. *ACS Cent. Sci.* **2017**, *3*, 1180–1188. [[CrossRef](#)]
92. Ewald, J.; Sieber, P.; Garde, R.; Lang, S.N.; Schuster, S.; Ibrahim, B. Trends in mathematical modeling of host-pathogen interactions. *Cell. Mol. Life Sci.* **2020**, *77*, 467–480. [[CrossRef](#)] [[PubMed](#)]
93. Qiao, Q.; Wang, L.; Meng, F.L.; Hwang, J.K.; Alt, F.W.; Wu, H. AID Recognizes Structured DNA for Class Switch Recombination. *Mol. Cell* **2017**, *67*, 361–373 e364. [[CrossRef](#)] [[PubMed](#)]
94. King, J.J.; Larijani, M. A Novel Regulator of Activation-Induced Cytidine Deaminase/APOBECs in Immunity and Cancer: Schrodinger's CATalytic Pocket. *Front. Immunol.* **2017**, *8*, 351. [[CrossRef](#)] [[PubMed](#)]
95. King, J.J.; Manuel, C.A.; Barrett, C.V.; Raber, S.; Lucas, H.; Sutter, P.; Larijani, M. Catalytic pocket inaccessibility of activation-induced cytosine deaminase is a safeguard against excessive mutagenic activity. *Structure* **2015**, *23*, 615–627. [[CrossRef](#)] [[PubMed](#)]
96. Sayers, E.W.; Bolton, E.E.; Brister, J.R.; Canese, K.; Chan, J.; Comeau, D.C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2022**, *50*, D20–D26. [[CrossRef](#)]
97. Martin, F.J.; Amode, M.R.; Aneja, A.; Austine-Orimoloye, O.; Azov, A.G.; Barnes, I.; Becker, A.; Bennett, R.; Berry, A.; Bhai, J.; et al. Ensembl 2023. *Nucleic Acids Res.* **2023**, *51*, D933–D941. [[CrossRef](#)]
98. Gille, C.; Fahling, M.; Weyand, B.; Wieland, T.; Gille, A. Alignment-annotator web server: Rendering and annotating sequence alignments. *Nucleic Acids Res.* **2014**, *42*, W3–W6. [[CrossRef](#)]

99. Gille, C.; Frommel, C. STRAP: Editor for STRuctural Alignments of Proteins. *Bioinformatics* **2001**, *17*, 377–378. [[CrossRef](#)]
100. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)]
101. Schneider, T.D.; Stephens, R.M. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **1990**, *18*, 6097–6100. [[CrossRef](#)] [[PubMed](#)]
102. Madden, T. The BLAST Sequence Analysis Tool. In *The NCBI Handbook [Internet]*, 2nd ed.; NCBI: Bethesda, MD, USA, 2013.
103. Pei, J.; Kim, B.H.; Grishin, N.V. PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **2008**, *36*, 2295–2300. [[CrossRef](#)] [[PubMed](#)]
104. Shaban, N.M.; Shi, K.; Li, M.; Aihara, H.; Harris, R.S. 1.92 Angstrom Zinc-Free APOBEC3F Catalytic Domain Crystal Structure. *J. Mol. Biol.* **2016**, *428*, 2307–2316. [[CrossRef](#)] [[PubMed](#)]
105. Guo, Y.; Dong, L.; Qiu, X.; Wang, Y.; Zhang, B.; Liu, H.; Yu, Y.; Zang, Y.; Yang, M.; Huang, Z. Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. *Nature* **2014**, *505*, 229–233. [[CrossRef](#)]
106. Binning, J.M.; Chesarino, N.M.; Emerman, M.; Gross, J.D. Structural Basis for a Species-Specific Determinant of an SIV Vif Protein toward Hominid APOBEC3G Antagonism. *Cell Host Microbe* **2019**, *26*, 739–747 e734. [[CrossRef](#)]
107. Abascal, F.; Zardoya, R.; Telford, M.J. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **2010**, *38*, W7–W13. [[CrossRef](#)]
108. Lefort, V.; Longueville, J.E.; Gascuel, O. SMS: Smart Model Selection in PhyML. *Mol. Biol. Evol.* **2017**, *34*, 2422–2424. [[CrossRef](#)]
109. Jones, D.T.; Taylor, W.R.; Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci. CABIOS* **1992**, *8*, 275–282. [[CrossRef](#)]
110. Nickle, D.C.; Heath, L.; Jensen, M.A.; Gilbert, P.B.; Mullins, J.I.; Kosakovsky Pond, S.L. HIV-specific probabilistic models of protein evolution. *PLoS ONE* **2007**, *2*, e503. [[CrossRef](#)]
111. Simon, T. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *Lect. Math. Life Sci.* **1986**, *17*, 57–86.
112. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [[CrossRef](#)] [[PubMed](#)]
113. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [[CrossRef](#)] [[PubMed](#)]
114. Bertoni, M.; Kiefer, F.; Biasini, M.; Bordoli, L.; Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci. Rep.* **2017**, *7*, 10480. [[CrossRef](#)] [[PubMed](#)]
115. Bienert, S.; Waterhouse, A.; de Beer, T.A.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res.* **2017**, *45*, D313–D319. [[CrossRef](#)]
116. Guex, N.; Peitsch, M.C.; Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **2009**, *30* (Suppl. S1), S162–S173. [[CrossRef](#)]
117. Benkert, P.; Tosatto, S.C.; Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* **2008**, *71*, 261–277. [[CrossRef](#)]
118. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinform.* **2008**, *9*, 40. [[CrossRef](#)]
119. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738. [[CrossRef](#)]
120. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7–8. [[CrossRef](#)]
121. Wu, S.; Zhang, Y. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* **2007**, *35*, 3375–3382. [[CrossRef](#)]
122. Yang, J.; Zhang, W.; He, B.; Walker, S.E.; Zhang, H.; Govindarajoo, B.; Virtanen, J.; Xue, Z.; Shen, H.B.; Zhang, Y. Template-based protein structure prediction in CASP11 and retrospect of I-TASSER in the last decade. *Proteins* **2016**, *84* (Suppl. S1), 233–246. [[CrossRef](#)]
123. Kitamura, S.; Ode, H.; Nakashima, M.; Imahashi, M.; Naganawa, Y.; Kurosawa, T.; Yokomaku, Y.; Yamane, T.; Watanabe, N.; Suzuki, A.; et al. The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1005–1010. [[CrossRef](#)]
124. Kitamura, S.; Ode, H.; Nakashima, M.; Imahashi, M.; Naganawa, Y.; Ibe, S.; Yokomaku, Y.; Watanabe, N.; Suzuki, A.; Sugiura, W.; et al. Crystal structure of the human APOBEC3C having HIV-1 Vif-binding interface. *Protein Data Bank* **2011**. [[CrossRef](#)]
125. Cheng, C.; Zhang, T.; Wang, C.; Lan, W.; Ding, J.; Cao, C. Crystal Structure of Cytidine Deaminase Human APOBEC3F Chimeric Catalytic Domain in Complex with DNA. *Chin. J. Chem.* **2019**, *36*, 1241–1248. [[CrossRef](#)]
126. Li, M.; Shandilya, S.M.; Carpenter, M.A.; Rathore, A.; Brown, W.L.; Perkins, A.L.; Harki, D.A.; Solberg, J.; Hook, D.J.; Pandey, K.K.; et al. First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *ACS Chem. Biol.* **2012**, *7*, 506–517. [[CrossRef](#)]

127. Hu, Y.; Desimmie, B.A.; Nguyen, H.C.; Ziegler, S.J.; Cheng, T.C.; Chen, J.; Wang, J.; Wang, H.; Zhang, K.; Pathak, V.K.; et al. Structural basis of antagonism of human APOBEC3F by HIV-1 Vif. *Nat. Struct. Mol. Biol.* **2019**, *26*, 1176–1183. [[CrossRef](#)] [[PubMed](#)]
128. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)]
129. The PyMOL Molecular Graphics System, Version 1.7.6 Schrödinger, LLC. Available online: <https://www.pymol.org/> (accessed on 30 September 2016).
130. Yang, J.; Roy, A.; Zhang, Y. BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.* **2013**, *41*, D1096–D1103. [[CrossRef](#)]
131. Yang, J.; Roy, A.; Zhang, Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* **2013**, *29*, 2588–2595. [[CrossRef](#)]
132. Lovell, S.C.; Davis, I.W.; Arendall, W.B., 3rd; de Bakker, P.I.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Calpha geometry: Phi,psi and Cbeta deviation. *Proteins* **2003**, *50*, 437–450. [[CrossRef](#)]
133. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)] [[PubMed](#)]
134. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Zidek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596. [[CrossRef](#)] [[PubMed](#)]
135. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [[CrossRef](#)] [[PubMed](#)]
136. Olsson, M.H.; Sondergaard, C.R.; Rostkowski, M.; Jensen, J.H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537. [[CrossRef](#)]
137. Bas, D.C.; Rogers, D.M.; Jensen, J.H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* **2008**, *73*, 765–783. [[CrossRef](#)]
138. Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667. [[CrossRef](#)]
139. Li, H.; Robertson, A.D.; Jensen, J.H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **2005**, *61*, 704–721. [[CrossRef](#)]
140. van Zundert, G.C.P.; Rodrigues, J.; Trellet, M.; Schmitz, C.; Kastiris, P.L.; Karaca, E.; Melquiond, A.S.J.; van Dijk, M.; de Vries, S.J.; Bonvin, A. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720–725. [[CrossRef](#)]
141. Dominguez, C.; Boelens, R.; Bonvin, A.M. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737. [[CrossRef](#)]
142. Diamond, C.P.; Im, J.; Button, E.A.; Huebert, D.N.G.; King, J.J.; Borzooee, F.; Abdouni, H.S.; Bacque, L.; McCarthy, E.; Fifield, H.; et al. AID, APOBEC3A and APOBEC3B efficiently deaminate deoxycytidines neighboring DNA damage induced by oxidation or alkylation. *Biochim. Et Biophys. Acta. Gen. Subj.* **2019**, *1863*, 129415. [[CrossRef](#)]
143. Ghorbani, A.; Khataeipour, S.J.; Solbakken, M.H.; Huebert, D.N.G.; Khoddami, M.; Eslamloo, K.; Collins, C.; Hori, T.; Jentoft, S.; Rise, M.L.; et al. Ancestral reconstruction reveals catalytic inactivation of activation-induced cytidine deaminase concomitant with cold water adaptation in the Gadiformes bony fish. *BMC Biol.* **2022**, *20*, 293. [[CrossRef](#)] [[PubMed](#)]
144. Byeon, I.J.; Ahn, J.; Mitra, M.; Byeon, C.H.; Hercik, K.; Hritz, J.; Charlton, L.M.; Levin, J.G.; Gronenborn, A.M. NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nat. Commun.* **2013**, *4*, 1890. [[CrossRef](#)] [[PubMed](#)]
145. Byeon, I.J.; Byeon, C.H.; Wu, T.; Mitra, M.; Singer, D.; Levin, J.G.; Gronenborn, A.M. Nuclear Magnetic Resonance Structure of the APOBEC3B Catalytic Domain: Structural Basis for Substrate Binding and DNA Deaminase Activity. *Biochemistry* **2016**, *55*, 2944–2959. [[CrossRef](#)] [[PubMed](#)]
146. Pham, P.; Afif, S.A.; Shimoda, M.; Maeda, K.; Sakaguchi, N.; Pedersen, L.C.; Goodman, M.F. Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification. *DNA Repair* **2016**, *43*, 48–56. [[CrossRef](#)]
147. Chen, K.M.; Harjes, E.; Gross, P.J.; Fahmy, A.; Lu, Y.; Shindo, K.; Harris, R.S.; Matsuo, H. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **2008**, *452*, 116–119. [[CrossRef](#)]
148. Zoete, V.; Cuendet, M.A.; Grosdidier, A.; Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **2011**, *32*, 2359–2368. [[CrossRef](#)]
149. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [[CrossRef](#)]
150. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898. [[CrossRef](#)]

151. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [[CrossRef](#)]
152. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)]
153. Ronquist, F. Bayesian inference of character evolution. *Trends Ecol. Evol.* **2004**, *19*, 475–481. [[CrossRef](#)] [[PubMed](#)]
154. Geyer, C.J. Markov chain Monte Carlo maximum likelihood. Computing Science and Statistics 1991. In Proceedings of the 23rd Symposium on the Interface, Seattle, WA, USA, 21–24 April 1991; pp. 156–163.
155. Meyer, X. Adaptive Tree Proposals for Bayesian Phylogenetic Inference. *Syst. Biol.* **2021**, *70*, 1015–1032. [[CrossRef](#)] [[PubMed](#)]
156. Letko, M.; Silvestri, G.; Hahn, B.H.; Bibollet-Ruche, F.; Gokcumen, O.; Simon, V.; Ooms, M. Vif proteins from diverse primate lentiviral lineages use the same binding site in APOBEC3G. *J. Virol.* **2013**, *87*, 11861–11871. [[CrossRef](#)]
157. Cuevas, J.M.; Geller, R.; Garijo, R.; Lopez-Aldeguer, J.; Sanjuan, R. Extremely High Mutation Rate of HIV-1 In Vivo. *PLoS Biol.* **2015**, *13*, e1002251. [[CrossRef](#)]
158. Quinlan, E.M.; King, J.J.; Amemiya, C.T.; Hsu, E.; Larijani, M. Biochemical Regulatory Features of Activation-Induced Cytidine Deaminase Remain Conserved from Lampreys to Humans. *Mol. Cell. Biol.* **2017**, *37*, e00077-17. [[CrossRef](#)]
159. Hix, M.A.; Cisneros, G.A. Computational Investigation of APOBEC3H Substrate Orientation and Selectivity. *J. Phys. Chem. B* **2020**, *124*, 3903–3908. [[CrossRef](#)]
160. Vasudevan, A.A.J.; Balakrishnan, K.; Gertzen, C.G.W.; Borvetó, F.; Zhang, Z.; Sangwiman, A.; Held, U.; Küstermann, C.; Banerjee, S.; Schumann, G.G.; et al. Loop 1 of APOBEC3C regulates its antiviral activity against HIV-1. *bioRxiv* **2020**, *432*, 1–67. [[CrossRef](#)]
161. Mitra, M.; Singer, D.; Mano, Y.; Hritz, J.; Nam, G.; Gorelick, R.J.; Byeon, I.J.; Gronenborn, A.M.; Iwatani, Y.; Levin, J.G. Sequence and structural determinants of human APOBEC3H deaminase and anti-HIV-1 activities. *Retrovirology* **2015**, *12*, 3. [[CrossRef](#)]
162. Silvas, T.V.; Schiffer, C.A. APOBEC3s: DNA-editing human cytidine deaminases. *Protein Sci. A Publ. Protein Soc.* **2019**, *28*, 1552–1566. [[CrossRef](#)]
163. Ghimire-Rijal, S.; Maynard, E.L., Jr. Comparative thermodynamic analysis of zinc binding to the His/Cys motif in virion infectivity factor. *Inorg. Chem.* **2014**, *53*, 4295–4302. [[CrossRef](#)]
164. Smith, J.L.; Izumi, T.; Borbet, T.C.; Hagedorn, A.N.; Pathak, V.K. HIV-1 and HIV-2 Vif interact with human APOBEC3 proteins using completely different determinants. *J. Virol.* **2014**, *88*, 9893–9908. [[CrossRef](#)] [[PubMed](#)]
165. Yang, H.; Ito, F.; Wolfe, A.D.; Li, S.; Mohammadzadeh, N.; Love, R.P.; Yan, M.; Zirkle, B.; Gaba, A.; Chelico, L.; et al. Understanding the structural basis of HIV-1 restriction by the full length double-domain APOBEC3G. *Nat. Commun.* **2020**, *11*, 632. [[CrossRef](#)] [[PubMed](#)]
166. Maiti, A.; Myint, W.; Delviks-Frankenberry, K.A.; Hou, S.; Kanai, T.; Balachandran, V.; Sierra Rodriguez, C.; Tripathi, R.; Kurt Yilmaz, N.; Pathak, V.K.; et al. Crystal Structure of a Soluble APOBEC3G Variant Suggests ssDNA to Bind in a Channel that Extends between the Two Domains. *J. Mol. Biol.* **2020**, *432*, 6042–6060. [[CrossRef](#)] [[PubMed](#)]
167. Letko, M.; Booiman, T.; Kootstra, N.; Simon, V.; Ooms, M. Identification of the HIV-1 Vif and Human APOBEC3G Protein Interface. *Cell Rep.* **2015**, *13*, 1789–1799. [[CrossRef](#)]
168. Zennou, V.; Bieniasz, P.D. Comparative analysis of the antiretroviral activity of APOBEC3G and APOBEC3F from primates. *Virology* **2006**, *349*, 31–40. [[CrossRef](#)]
169. Hayward, J.A.; Tachedjian, M.; Cui, J.; Cheng, A.Z.; Johnson, A.; Baker, M.L.; Harris, R.S.; Wang, L.F.; Tachedjian, G. Differential Evolution of Antiretroviral Restriction Factors in Pteropid Bats as Revealed by APOBEC3 Gene Complexity. *Mol. Biol. Evol.* **2018**, *35*, 1626–1637. [[CrossRef](#)]
170. 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*, 7494–7506. [[CrossRef](#)]
171. Maiti, A.; Hedger, A.K.; Myint, W.; Balachandran, V.; Watts, J.K.; Schiffer, C.A.; Matsuo, H. Structure of the catalytically active APOBEC3G bound to a DNA oligonucleotide inhibitor reveals tetrahedral geometry of the transition state. *Nat. Commun.* **2022**, *13*, 7117. [[CrossRef](#)]
172. 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*, 728–733. [[CrossRef](#)]
173. Shi, K.; Demir, O.; 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*, 49–58. [[CrossRef](#)]
174. Abdouni, H.S.; King, J.J.; Ghorbani, A.; Fifield, H.; Berghuis, L.; Larijani, M. DNA/RNA hybrid substrates modulate the catalytic activity of purified AID. *Mol. Immunol.* **2018**, *93*, 94–106. [[CrossRef](#)] [[PubMed](#)]
175. Azimi, F.C.; Lee, J.E. Structural perspectives on HIV-1 Vif and APOBEC3 restriction factor interactions. *Protein Sci. A Publ. Protein Soc.* **2020**, *29*, 391–406. [[CrossRef](#)] [[PubMed](#)]
176. Zhang, J.; Webb, D.M. Rapid evolution of primate antiviral enzyme APOBEC3G. *Hum. Mol. Genet.* **2004**, *13*, 1785–1791. [[CrossRef](#)] [[PubMed](#)]

177. Etienne, L.; Bibollet-Ruche, F.; Sudmant, P.H.; Wu, L.I.; Hahn, B.H.; Emerman, M. The Role of the Antiviral APOBEC3 Gene Family in Protecting Chimpanzees against Lentiviruses from Monkeys. *PLoS Pathog.* **2015**, *11*, e1005149. [[CrossRef](#)]
178. Liu, M.C.; Liao, W.Y.; Buckley, K.M.; Yang, S.Y.; Rast, J.P.; Fugmann, S.D. AID/APOBEC-like cytidine deaminases are ancient innate immune mediators in invertebrates. *Nat. Commun.* **2018**, *9*, 1948. [[CrossRef](#)]
179. Krishnan, A.; Iyer, L.M.; Holland, S.J.; Boehm, T.; Aravind, L. Diversification of AID/APOBEC-like deaminases in metazoa: Multiplicity of clades and widespread roles in immunity. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E3201–E3210. [[CrossRef](#)]
180. Basu, U.; Wang, Y.; Alt, F.W. Evolution of phosphorylation-dependent regulation of activation-induced cytidine deaminase. *Mol. Cell* **2008**, *32*, 285–291. [[CrossRef](#)]
181. Wakae, K.; Magor, B.G.; Saunders, H.; Nagaoka, H.; Kawamura, A.; Kinoshita, K.; Honjo, T.; Muramatsu, M. Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. *Int. Immunol.* **2006**, *18*, 41–47. [[CrossRef](#)]
182. Chen, Q.; Xiao, X.; Wolfe, A.; Chen, X.S. The in vitro Biochemical Characterization of an HIV-1 Restriction Factor APOBEC3F: Importance of Loop 7 on Both CD1 and CD2 for DNA Binding and Deamination. *J. Mol. Biol.* **2016**, *428*, 2661–2670. [[CrossRef](#)]
183. Salter, J.D.; Bennett, R.P.; Smith, H.C. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem. Sci.* **2016**, *41*, 578–594. [[CrossRef](#)]
184. Bransteitter, R.; Prochnow, C.; Chen, X.S. The current structural and functional understanding of APOBEC deaminases. *Cell. Mol. Life Sci.* **2009**, *66*, 3137–3147. [[CrossRef](#)] [[PubMed](#)]
185. Conticello, S.G.; Thomas, C.J.; Petersen-Mahrt, S.K.; Neuberger, M.S. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol.* **2005**, *22*, 367–377. [[CrossRef](#)] [[PubMed](#)]
186. Losey, H.C.; Ruthenburg, A.J.; Verdine, G.L. Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nat. Struct. Mol. Biol.* **2006**, *13*, 153–159. [[CrossRef](#)]
187. Bitra, A.; Biswas, A.; Anand, R. Structural basis of the substrate specificity of cytidine deaminase superfamily Guanine deaminase. *Biochemistry* **2013**, *52*, 8106–8114. [[CrossRef](#)]
188. Jia, Q.; Zhang, J.; Zeng, H.; Tang, J.; Xiao, N.; Gao, S.; Li, H.; Xie, W. Substrate Specificity of GSDA Revealed by Cocrystal Structures and Binding Studies. *Int. J. Mol. Sci.* **2022**, *23*, 14976. [[CrossRef](#)]
189. Dolce, L.G.; Zimmer, A.A.; Tengo, L.; Weis, F.; Rubio, M.A.T.; Alfonzo, J.D.; Kowalinski, E. Structural basis for sequence-independent substrate selection by eukaryotic wobble base tRNA deaminase ADAT2/3. *Nat. Commun.* **2022**, *13*, 6737. [[CrossRef](#)]
190. Thuy-Boun, A.S.; Thomas, J.M.; Grajo, H.L.; Palumbo, C.M.; Park, S.; Nguyen, L.T.; Fisher, A.J.; Beal, P.A. Asymmetric dimerization of adenosine deaminase acting on RNA facilitates substrate recognition. *Nucleic Acids Res.* **2020**, *48*, 7958–7972. [[CrossRef](#)]
191. Minasov, G.; Wawrzak, Z.; Skarina, T.; Wang, Y.; Grimshaw, S.; Papazisi, L.; Savchenko, A.; Anderson, W.F.; Center for Structural Genomics of Infectious Diseases (CSGID). 2.2 Angstrom Crystal Structure of Cytidine deaminase from *Vibrio cholerae* in Complex with Zinc and Uridine. *Protein Data Bank* **2012**. [[CrossRef](#)]
192. Teh, A.H.; Kimura, M.; Yamamoto, M.; Tanaka, N.; Yamaguchi, I.; Kumasaka, T. The 1.48 Å resolution crystal structure of the homotetrameric cytidine deaminase from mouse. *Biochemistry* **2006**, *45*, 7825–7833. [[CrossRef](#)]
193. Marx, A.; Alian, A. The first crystal structure of a dTTP-bound deoxycytidylate deaminase validates and details the allosteric-inhibitor binding site. *J. Biol. Chem.* **2015**, *290*, 682–690. [[CrossRef](#)] [[PubMed](#)]
194. Kouno, T.; Silvas, T.V.; Hilbert, B.J.; Shandilya, S.M.D.; Bohn, M.F.; Kelch, B.A.; Royer, W.E.; Somasundaran, M.; Kurt Yilmaz, N.; Matsuo, H.; et al. Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nat. Commun.* **2017**, *8*, 15024. [[CrossRef](#)] [[PubMed](#)]
195. Shi, K.; Carpenter, M.A.; Banerjee, S.; Shaban, N.M.; Kurahashi, K.; Salamango, D.J.; McCann, J.L.; Starrett, G.J.; Duffy, J.V.; Demir, O.; et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.* **2017**, *24*, 131–139. [[CrossRef](#)] [[PubMed](#)]
196. Hu, Y.; Delviks-Frankenberry, K.A.; Wu, C.; Arizaga, F.; Pathak, V.K.; Xiong, Y. Structural insights into PPP2R5A degradation by HIV-1 Vif. *Nat. Struct. Mol. Biol.* **2024**, *31*, 1492–1501. [[CrossRef](#)] [[PubMed](#)]
197. Richards, C.; Albin, J.S.; Demir, O.; Shaban, N.M.; Luengas, E.M.; Land, A.M.; Anderson, B.D.; Holten, J.R.; Anderson, J.S.; Harki, D.A.; et al. The Binding Interface between Human APOBEC3F and HIV-1 Vif Elucidated by Genetic and Computational Approaches. *Cell Rep.* **2015**, *13*, 1781–1788. [[CrossRef](#)]
198. Nakashima, M.; Tsuzuki, S.; Awazu, H.; Hamano, A.; Okada, A.; Ode, H.; Maejima, M.; Hachiya, A.; Yokomaku, Y.; Watanabe, N.; et al. Mapping Region of Human Restriction Factor APOBEC3H Critical for Interaction with HIV-1 Vif. *J. Mol. Biol.* **2017**, *429*, 1262–1276. [[CrossRef](#)]
199. Ooms, M.; Letko, M.; Simon, V. The Structural Interface between HIV-1 Vif and Human APOBEC3H. *J. Virol.* **2017**, *91*, 10–1128. [[CrossRef](#)]
200. Smith, J.L.; Pathak, V.K. Identification of specific determinants of human APOBEC3F, APOBEC3C, and APOBEC3DE and African green monkey APOBEC3F that interact with HIV-1 Vif. *J. Virol.* **2010**, *84*, 12599–12608. [[CrossRef](#)]
201. Modenini, G.; Abondio, P.; Boattini, A. The coevolution between APOBEC3 and retrotransposons in primates. *Mob. DNA* **2022**, *13*, 27. [[CrossRef](#)]

202. Shaban, N.M.; Yan, R.; Shi, K.; Moraes, S.N.; Cheng, A.Z.; Carpenter, M.A.; McLellan, J.S.; Yu, Z.; Harris, R.S. Cryo-EM structure of the EBV ribonucleotide reductase BORF2 and mechanism of APOBEC3B inhibition. *Sci. Adv.* **2022**, *8*, eabm2827. [[CrossRef](#)]
203. 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. *Sci. Rep.* **2018**, *8*, 3763. [[CrossRef](#)]
204. Shaban, N.M.; Shi, K.; Lauer, K.V.; Carpenter, M.A.; Richards, C.M.; Salamango, D.; Wang, J.; Lopresti, M.W.; Banerjee, S.; Levin-Klein, R.; et al. The Antiviral and Cancer Genomic DNA Deaminase APOBEC3H Is Regulated by an RNA-Mediated Dimerization Mechanism. *Mol. Cell* **2018**, *69*, 75–86 e79. [[CrossRef](#)] [[PubMed](#)]
205. Shi, K.; Demir, O.; Carpenter, M.A.; Wagner, J.; Kurahashi, K.; Harris, R.S.; Amaro, R.E.; Aihara, H. Conformational Switch Regulates the DNA Cytosine Deaminase Activity of Human APOBEC3B. *Sci. Rep.* **2017**, *7*, 17415. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.