Insights into DNA substrate selection by APOBEC3G from structural, biochemical, and functional studies

Samantha J. Ziegler[1☯]**, Chang Liu**[1☯]**, Mark Landau**[1,2]**, Olga Buzovetsky**[1]**, Belete A. Desimmie**[3]**, Qi Zhao**[1]**, Tomoaki Sasaki**[2]**, Ryan C. Burdick**[3]**, Vinay K. Pathak**[3]**, Karen S. Anderson**[1,2]**, Yong Xiong**[1] *

1 Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, United States of America, 2 Department of Pharmacology, Yale University School of Medicine, New Haven, Connecticut, United States of America, 3 Viral Mutation Section, HIV Dynamics and Replication Program, Center for Cancer Research, National Cancer Institute at Frederick, Frederick, Maryland, United States of America

☯ These authors contributed equally to this work.
* yong.xiong@yale.edu

Abstract

Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3 (A3) proteins are a family of cytidine deaminases that catalyze the conversion of deoxycytidine (dC) to deoxyuridine (dU) in single-stranded DNA (ssDNA). A3 proteins act in the innate immune response to viral infection by mutating the viral ssDNA. One of the most well-studied human A3 family members is A3G, which is a potent inhibitor of HIV-1. Each A3 protein prefers a specific substrate sequence for catalysis—for example, A3G deaminates the third dC in the CC**C**A sequence motif. However, the interaction between A3G and ssDNA is difficult to characterize due to poor solution behavior of the full-length protein and loss of DNA affinity of the truncated protein. Here, we present a novel DNA-anchoring fusion strategy using the protection of telomeres protein 1 (Pot1) which has nanomolar affinity for ssDNA, with which we captured an A3G-ssDNA interaction. We crystallized a non-preferred adenine in the -1 nucleotide-binding pocket of A3G. The structure reveals a unique conformation of the catalytic site loops that sheds light onto how the enzyme scans substrate in the -1 pocket. Furthermore, our biochemistry and virology studies provide evidence that the nucleotide-binding pockets on A3G influence each other in selecting the preferred DNA substrate. Together, the results provide insights into the mechanism by which A3G selects and deaminates its preferred substrates and help define how A3 proteins are tailored to recognize specific DNA sequences. This knowledge contributes to a better understanding of the mechanism of DNA substrate selection by A3G, as well as A3G antiviral activity against HIV-1.

Introduction

Apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G or A3G) is a human host restriction factor that inhibits human immunodeficiency virus type 1 (HIV-1), murine leukemia virus, and equine infectious anemia virus [1–5] primarily through its deoxycytidine deaminase activity on the viral minus-strand DNA (-ssDNA). A3G is one of the seven APOBEC3 (A3) proteins that inhibits replication of a diverse set of viruses [4, 6–8]. The human cytidine deaminase superfamily members have highly conserved protein sequences, tertiary structural folds, and catalytic mechanisms [4, 9]. All A3 proteins and the closely related activation-induced cytidine deaminase (AID) contain a single catalytically active cytidine deaminase (CDA) domain. However, some family members, such as A3B, A3D, A3F, and A3G, have a second pseudocatalytic domain that retains the same tertiary fold, but are not catalytically active. A3G inhibits viral replication by preferentially deaminating dC to dU in the viral -ssDNA during reverse transcription [7, 10–12]. Specific "hotspot" sequences of DNA are targeted for deamination, with the highest preference for the 5'-CC**C**A sequence (where the deaminated C is underlined) in the case of A3G [2, 13]. Deamination of the -ssDNA by A3G leads to extensive G-to-A hypermutation in the viral genome, ultimately eliminating viral infectivity [1, 5, 13]. Some studies suggest that mutations induced by A3G are occasionally sub-lethal, facilitating viral evolution and allowing the virus to develop drug resistance [3, 14]. However, a recent study concluded that hypermutation is almost always a lethal event and makes little or no contribution to viral genetic variation [15]. This underlines the importance of understanding the mechanism by which A3G induces mutations in the HIV-1 genome [3, 16]. While the A3s' deaminase activity is beneficial during viral restriction, this function has also been associated with multiple types of cancer [17]. For example, the deaminase activity of A3B has been linked to mutations found in breast cancer [18], and overactive AID can lead to non-Hodgkin's lymphoma [19]. In this regard, APOBECs are important contributors to cancer development, and thus, it is critical to uncover the mechanisms underlining the interactions between the A3 proteins and their DNA substrates [20].

Biochemical studies [21–24], as well as the apo structure of the C-terminal catalytic domain of A3G (A3G$_{CTD}$) [21, 22, 24–26], have established a framework for A3G hotspot selectivity. Extensive biochemical and mutagenesis studies have shown that a single loop, loop 7, of A3G is responsible for selecting the base at the -1 position (the nucleotide at the 5' side of the deamination site) of the DNA substrate hotspot [27–30]. Swapping this loop into the complementary site in AID altered AID hotspot preference to that of A3G [27, 29]. Although the crystal structures of A3A and an A3B/A3A chimera, where A3B loop 1 is replaced with that of A3A, in complex with DNA have recently been published [31, 32], a structure of a catalytically active A3G bound to its substrate has yet to be determined. The lack of structural insight into the A3G-DNA complex has limited our understanding of the molecular determinants that govern A3G specificity and selectivity for its substrate. Structural studies have been hampered by the inherent difficulty of purifying full-length A3G with sufficient solubility. The A3G N-terminus (A3G$_{NTD}$) also readily forms soluble aggregates partly due to its propensity to bind RNA and DNA [33]. Although the isolated A3G$_{CTD}$ is able to bind and deaminate ssDNA, the loss of the A3G$_{NTD}$ results in a drastic decrease in DNA binding affinity [22].

In this study, we used a novel protein fusion strategy to obtain structural insight into the A3G-ssDNA interaction. We captured a non-preferred adenine in the -1 nucleotide (the nucleotide at the 5' side of deaminated cytidine) binding pocket, providing insight into DNA substrate scanning at the -1 site. Our results reveal key interactions that govern substrate sequence specificity of A3G and related cytidine deaminases. Furthermore, our results reveal how these key interactions perpetuate changes throughout the A3G hotspot binding pocket.

Together, this study provides a deeper understanding of A3G anti-viral activity and expands our knowledge of the deamination mechanism of related cytidine deaminases.

Results

Novel fusion design captures A3G$_{CTD}$ bound to a nucleotide

Although A3G$_{CTD}$ overcomes the solubility problem of the full-length protein and retains activity, DNA binding is weak and does not form a stable DNA-protein complex required for structural analysis. To address the problem of weak DNA binding, we designed a novel A3G$_{CTD}$ fusion to protection of telomeres protein 1 (Pot1) from *Schizosaccharomyces pombe* that serves as an anchor for the DNA (experimental schematic shown in Fig 1A). *S.pombe* Pot1 is a DNA binding protein that binds its specific cognate sequence, GGTTAC, at 1 nM affinity [34]. We designed a ssDNA substrate that contains both the Pot1 cognate sequence, GGTTAC, and the A3G hotspot sequence, CC**C**A, separated by a 14-nucleotide linker sequence (Fig 1A). We confirmed that the Pot1A3G fusion binds to ssDNA using size exclusion chromatography (Fig 1B) and that the fusion construct retains its native activity towards the A3G DNA substrate (Fig 1C). Pot1A3G can deaminate both the intended target C as well as the -1 C in vitro resulting in the two bands shown in Fig 1C, which agrees with previously published data that shows the -1 C can be deaminated under similar experimental conditions [35, 36]. These results show that Pot1 fusion may serve as a general strategy to stably anchor otherwise weakly bound DNA substrates to proteins.

Using the Pot1A3G$_{CTD}$ fusion construct, we obtained a crystal structure of A3G$_{CTD}$ bound to an adenine nucleotide at 2.9 Å resolution (Fig 1D and Table 1). The structure was
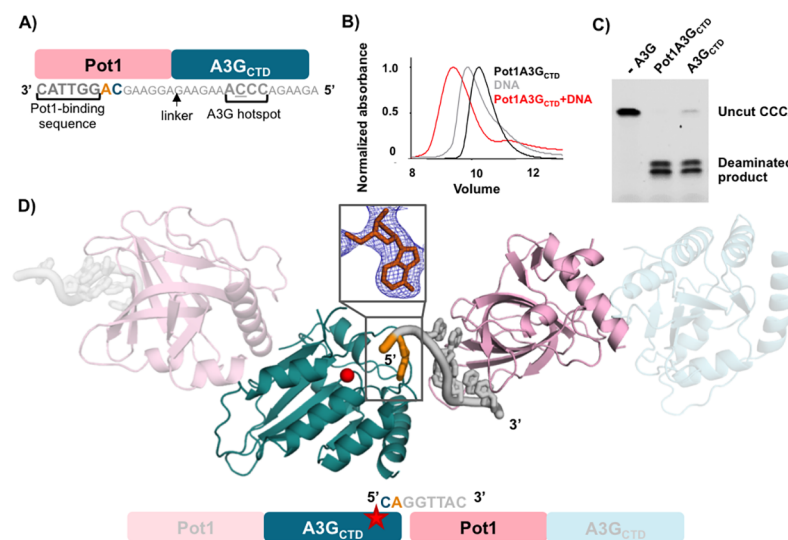
Fig 1. Structure of Pot1A3G$_{CTD}$ with ssDNA. A) Schematic of the Pot1A3G$_{CTD}$ fusion protein design. Pot1 (pink) is fused directly to the N-terminus of A3G$_{CTD}$ (blue). The ssDNA contains both Pot1 and A3G binding sites: the Pot1 site in dark gray and the A3G hotspot in light gray with the linker sequence in smaller font. The resolved adenine in the -1 pocket is colored orange and the expected deaminated cytidine is blue. B) Size exclusion binding test shows that Pot1A3G$_{CTD}$ binds to the ssDNA substrate. Pot1A3G$_{CTD}$ alone is in black, the ssDNA is in gray, and the mixture of the two is in red. C) Deamination activity using a UDG-dependent cleavage assay. The Pot1A3G$_{CTD}$ fusion protein has the same deamination activity as that of A3G$_{CTD}$. D) Schematic and structure of the Pot1A3G$_{CTD}$ in complex with DNA as observed in the crystal. The dA nucleotide bound to the -1 pocket is shown in orange. Two copies of the complex observed in the asymmetric unit are shown in blue (A3G), pink (Pot1), and grey/orange (DNA). The red star (schematic) and red sphere (structure) represent the zinc ion found in the catalytic site. The inset shows the 2Fo-Fc density (1σ contour level) observed for the adenine in the -1 nucleotide-binding pocket.

https://doi.org/10.1371/journal.pone.0195048.g001

Table 1. Data collection and refinement statistics.

Space group	P 4$_3$
Cell dimensions	
a, b, c (Å)	79.07, 79.07, 266.41
α, β, γ (˚)	90.00, 90.00, 90.00
Resolution (Å)	44–2.9 (3.0–2.9) *
No. reflections	40118
Redundancy	3.3 (3.3)
R$_{merge}$	0.105 (0.856)
R$_{pim}$	0.064 (0.539)
Mean I / σI	5.1 (1.1)
Completeness (%)	97.5
CC$_{1/2}$ last shell	0.70
Refinement	
Resolution (Å)	44–2.9
No. Reflections	32909
R$_{work}$ / R$_{free}$	0.24/0.29 (0.30/0.37)
No. atoms	11958
B-factors	88.2
RMSD in	
Bond lengths (Å)	0.019
Bond angles (˚)	1.9
Ramachandran	
Preferred	94.6
Allowed	5.0
Outliers	0.4

* Values in parentheses are for the highest-resolution shell

https://doi.org/10.1371/journal.pone.0195048.t001

determined in a P4$_3$ space group with four independent copies of the protein in the asymmetric unit of the crystal. Both A3G$_{CTD}$ and Pot1, together with the Pot1 cognate DNA, were clearly visible in the electron density. In the crystal, the ssDNA stretches from a Pot1 protein of one fusion complex to the A3G of the adjacent fusion complex. The rest of the ssDNA was disordered with clear electron density for only one nucleotide, adenine, located next to the catalytic site of A3G$_{CTD}$. This fusion design allowed us to capture the structure of A3G$_{CTD}$ bound to an adenine nucleotide in the -1 DNA binding pocket (Fig 1D inset).

A3G-adenine structure captures a non-preferred nucleotide in the -1 pocket

Our crystal structure of the Pot1A3G$_{CTD}$ fusion potentially captured an intermediate state of A3G during its search for a ssDNA hot spot sequence. A3G processively scans ssDNA [35] and prefers to deaminate dC in the context of the 5'-CC**C**A sequence, where the underlined C is deaminated. In this sequence, the nucleotide preceding the deaminated C is referred to as the -1 nucleotide and the adenine following the deaminated C is the +1 nucleotide [2, 13]. Although we intended to crystallize the A3G$_{CTD}$ bound to its entire hotspot region as described in Fig 1A, we instead captured the A3G$_{CTD}$ bound to a dA in the -1 pocket, which originates from the ssDNA bound to a neighboring Pot1 molecule (Fig 1D). We do not observe DNA binding at the catalytic site (0 position) or the +1 pocket (for the nucleotide at the 3'-side of the deaminated dC). Although the preferred nucleotide for A3G at the -1 position is dC, the binding pocket

appears to be flexible enough to accommodate an adenine at this position. During the processive scanning of ssDNA, the binding pocket likely allows binding of all nucleotides before encountering the preferred hotspot sequence for catalysis. The conformation we captured therefore possibly reflects how the enzyme encounters a non-preferred nucleotide at the -1 pocket, but not the catalytic state.

In our structure, the adenosine is stabilized by a network of hydrogen bonds and stacking interactions in the -1 pocket (Fig 2A and 2B). The backbone of residues P210 and I314 hydrogen bond with the amine group of the adenine, while Y315 hydrogen bonds with the phosphate group of the DNA backbone. The adenine base is further stabilized in the pocket through stacking interactions with W211, W285, and Y315 (Fig 2B). The overall architecture of the A3G$_{CTD}$ DNA binding pocket is conserved when compared to other structures of A3s bound to DNA. Our structure aligns well with the structures of A3A bound to thymidine at the -1 position (PDBIDs 5SWW, 5KEG) [31, 32] with an overall root mean square deviation (RMSD) of 0.8 Å and 0.7 Å, respectively. Comparison between the A3A and A3G$_{CTD}$ DNA bound structures reveals that highly conserved residues within the A3 family (A3G residues Y315 and W285) adopt similar orientations in the -1 pocket to bind to a nucleotide (Fig 2C). This highlights that some interactions between the -1 pocket and DNA are conserved between the proteins in the A3 family, even though different nucleotides are preferred at this position.

Analysis of the adenine-bound A3G$_{CTD}$ structure in comparison to other A3 crystal structures reveals that residues in the -1 pocket change conformations in response to the identity of the nucleotide that A3G encounters. Previous studies have shown that A3G$_{CTD}$ loop 7, specifically residue D317, is important for -1 nucleotide selectivity [30]. When comparing our structure of A3G$_{CTD}$ to the A3A-DNA structures (PDBID 5SWW, 5KEG) [31, 32], residues D316 and D317 change their conformations. When an A3A-preferred thymine nucleotide is bound to the -1 pocket, D131 of A3A (corresponding to D136 in A3G) forms a hydrogen bond with the Watson-Crick edge of the thymidine to orient the nucleotide in the pocket (Fig 2D). If A3G$_{CTD}$ D316 remained in a similar orientation as A3A D131, this residue would sterically clash with the adenine base. Instead, the A3G$_{CTD}$ -1 pocket accommodates the adenine base by flipping D316 180 degrees away from the nucleotide (Fig 2D), which opens the -1 pocket to allow for adenine to bind. To compensate for the loss of the hydrogen bond between the base and D316, A3G$_{CTD}$ interacts with the Watson-Crick edge of the adenine through backbone hydrogen bonds (Fig 2B). The adenine bound structure of A3G$_{CTD}$ aligns to the apo structures of A3G$_{CTD}$ (PDBIDs 3IR2 [24],4ROW/4ROV [37], 3IQS/3E1U [22]) with an RMSD of ~0.4 Å. Residues D316 and D317 in the adenine bound structure adopt a similar conformation to that of the apo structure (Fig 2D, green residues). This suggests that A3G samples this conformation in the absence of DNA and while scanning the DNA for its preferred hotspot sequence.

A3G loop 1 participates in -1 nucleotide recognition

In addition to the conformational changes of loop 7 residues, we also observed structural changes of A3G$_{CTD}$ loop 1. Comparison of our adenine bound structure to the A3G$_{CTD}$ apo structures (PDB 3IR2 [24],4ROW/4ROV [37], 3IQS/3E1U [22]) reveals that loop 1 of A3G$_{CTD}$ swings approximately 3Å towards the adenine (Fig 3A). The movement of loop 1 leads to the repositioning of W211 for stacking interactions with the adenine as well as bringing P210 closer to the base to allow for a hydrogen bond to form with the nucleotide (Fig 3B). Notably, while W211 on loop 1 flips inwards to close the binding pocket upon adenine binding, residues Y315 and W285 remain static (Fig 3B). The conformational changes that loop 1 undergoes in response to binding a non-preferred nucleotide in the -1 pocket suggests that this loop is also a
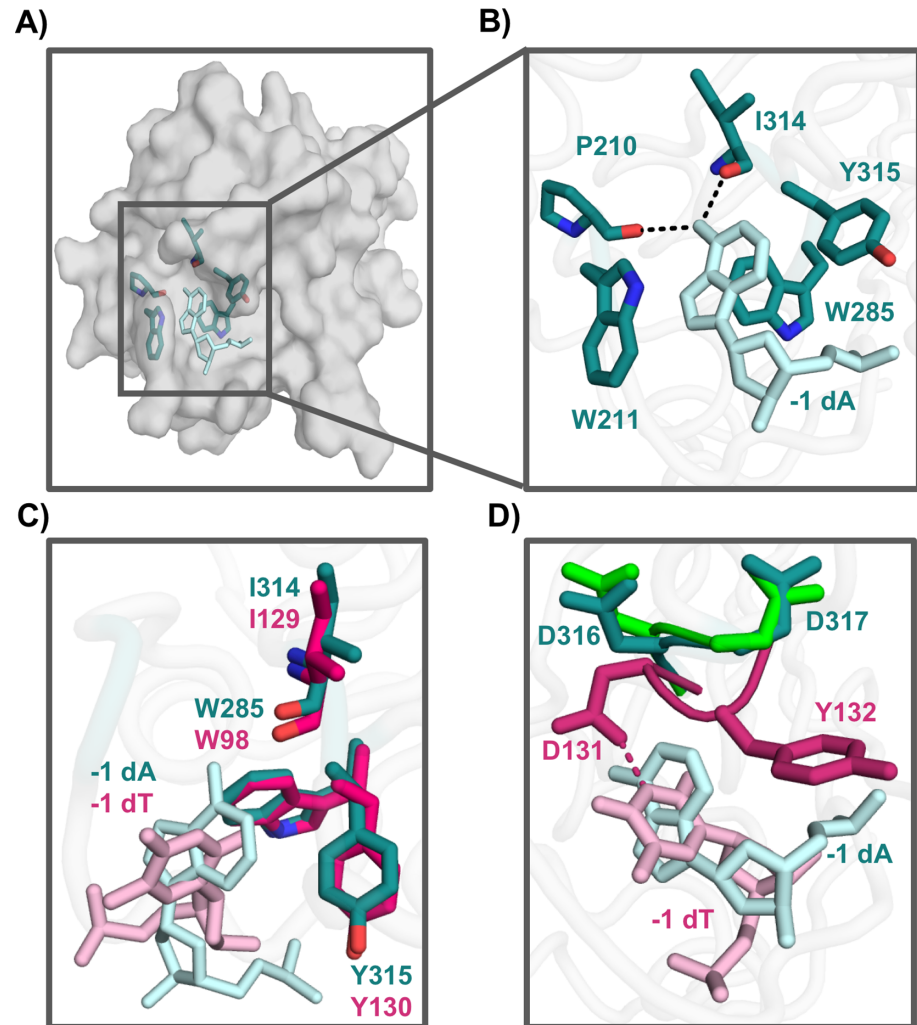
Fig 2. Structural analysis of loop 7 in A3G-ssDNA interaction. A) Overview of the A3G$_{CTD}$, shown in gray surface representation, bound to an adenine in the -1 nucleotide-binding pocket, shown in light blue sticks. Selected A3G$_{CTD}$ residues in the pocket are shown in teal sticks. B) The backbones of I314 and P210 form hydrogen bonds with the adenosine (shown in light blue) in the -1 nucleotide pocket. W211, W285, and Y315 stack with the nucleotide to stabilize it in the pocket. C) Structural alignment of the A3G$_{CTD}$ (teal) to A3A (magenta, PDBID 5SWW) [31], shows that conserved residues W285, I314, and Y315 of the -1 nucleotide-binding pocket are held in similar positions. D) Comparison of the A3G$_{CTD}$, teal, to the A3A-DNA complex, magenta (PDBID 5SWW) [31]. In the A3A structure, D131 forms a hydrogen bond with the Watson-Crick edge of the hotspot nucleotide thymidine (light pink). In the A3G$_{CTD}$ structure, D316 and D317 are flipped 180 degrees to avoid clashing with the non-preferred adenine (light blue). The flipped conformation is similar to that of the apo structure of the A3G$_{CTD}$, green (PDBID 3IR2) [24].

https://doi.org/10.1371/journal.pone.0195048.g002

flexible protein module, allowing for nucleotides that differ from the hotspot sequence to enter the nucleotide-binding pocket during scanning.

To further confirm the importance of loop 1 in catalysis, we mutated residues P210 and W211 and determined the enzymatic activity of the mutants. We found that mutating W211 to alanine abolished the catalytic activity of the enzyme (Fig 3C). This is consistent with the previous mutagenesis findings that the aromatic residues in the -1 nucleotide-binding pocket are critical for the catalytic and antiviral activity of A3G [21, 22, 24]. Furthermore, as proline residues introduce conformational constraints to the protein backbone, we examined the importance of rigidity of loop 1 by mutating P210 to a flexible glycine. This mutation caused a
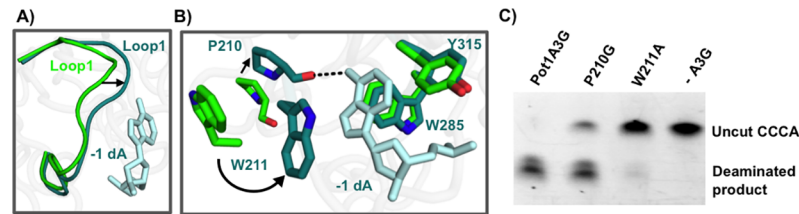
Fig 3. A3G$_{CTD}$ loop 1 is important for substrate recognition. A) Loop 1 in A3G$_{CTD}$-DNA complex (teal, coil representation) moves 3Å compared to apo A3G$_{CTD}$ (PDBID 3IR2, green) [24] to enclose the adenine in the -1 pocket. B) Comparison of the A3G$_{CTD}$ apo crystal structure (green; PDB ID 3IR2) [24] to the A3G$_{CTD}$-DNA structure (teal). Residue W211 flips in to stack with the nucleotide and residue P210 moves toward the nucleotide as compared to the apo structures, while W285 and Y315 remain static. C) Deaminase activity assays on A3G$_{CTD}$ mutants show that mutating residues on loop 1 can disrupt the deaminase activity of A3G completely in the case of W211A or partially in the case of P210G.

reduction in the deaminase activity (Fig 3C). These results suggest that loop1 is important not only for interacting with the nucleotides, but also affects A3G catalysis.

Mutating the -1 nucleotide-binding site of A3G affects hotspot preference at the +1 position

The recent A3A-DNA structures [31, 32] show that the ssDNA substrate is bound to the protein in a U-shaped conformation, with the +1 and the -1 nucleotide-binding pockets positioned adjacent to each other (Fig 4A). The close clustering of the nucleotide-binding pockets may allow the binding sites to influence each other. To examine if these sites are interconnected, we tested whether mutating -1 nucleotide-binding residues in loop 1 would affect the A3G hotspot preference at the +1 position. Some of the amino acid residues forming the -1 nucleotide-binding pocket (W285, I314, and Y315) are highly conserved among the human A3 family and AID (Fig 4A). In contrast, the A3 proteins and AID vary in sequence in loop 1 at residues P210 and W211 in the A3G$_{CTD}$; those with proline at the P210-equivalent positions favor adenine at the +1 site (A3G and A3B) [2], while those with arginine at this position have a higher propensity to favor a thymidine at this position (AID and A3F), although A3F still has a slight preference for an adenine [2, 38] (Fig 4B). This suggests that these -1 nucleotide-binding residues may also affect the selectivity for the +1 nucleotide in the ssDNA substrate. Specifically, among the A3 family homologs, AID shares the greatest degree of sequence conservation with A3G in the residues that interact with the +1 base, with only a single amino-acid variation at the A3G P210 (AID R19) position. This single residue may have substantial influence on the DNA preference of the two enzymes for the +1 position, where A3G favors adenosine while AID favors thymidine [39, 40].

We examined the extent to which the identity of the amino acid at the A3G P210 position dictates the +1 nucleotide preference. To mimic AID, we mutated P210 in A3G to arginine and measured deamination kinetics and substrate DNA preference using different ssDNA oligonucleotides containing a solitary target dC to be deaminated, where the underlined C in the preference motif CC**C**X is the target (Fig 4C and S1 Fig). We conducted single-turnover kinetics experiments to determine the K$_d$ and k$_{chem}$ for WT and P210R for different ssDNA oligonucleotides (Fig 4D and S1 Fig), and found that WT A3G$_{CTD}$ binds DNA with the canonical hotspot with a K$_d$ of 26 μM (Fig 4D). Notably, the P210R mutation decreased the affinity for CC**C**A, while increasing the affinity for both CC**C**T and CC**C**G substrates. The preference for the +1 nucleotide was calculated based on the ratio of efficiency for catalyzing DNA containing adenine compared to thymine or guanine at this position. A3G$_{CTD}$ preference for adenosine at
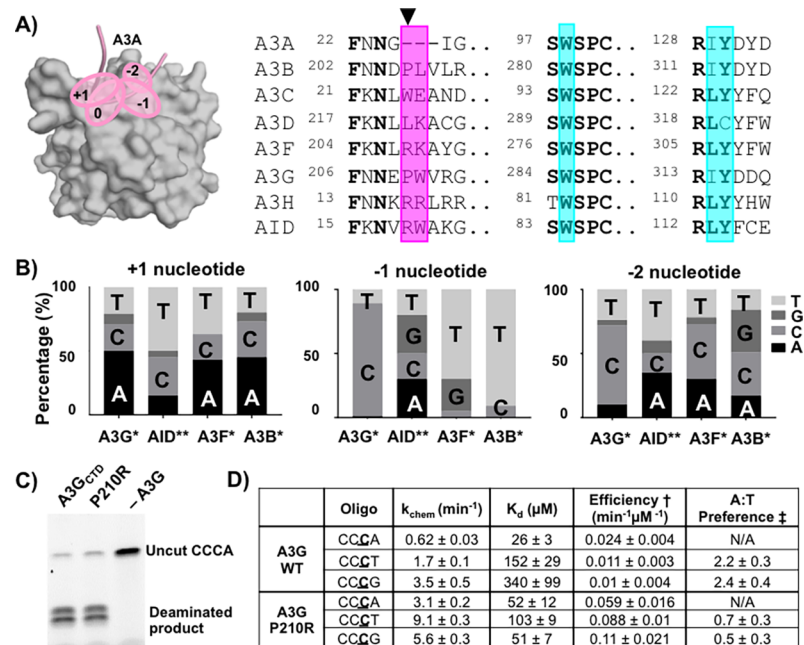
Fig 4. Mutating residues in loop 1 results in change of preference for +1 nucleotide. A) Schematic of the nucleotide binding sites of A3A (PDBID 5SWW) [31] that are spatially close to one another (marked by pink ovals). A3A is in surface presentation and the backbone of the bound ssDNA is shown as a pink coil. Right panel: sequence alignment of the proteins from the A3 superfamily. Conserved residues are in bold. Residues involved in hydrogen bonding with the nucleotide at the -1 position during scanning are shaded in magenta, and the other residues forming the -1 nucleotide pocket are shaded in teal. The arrow marks the A3G P210 corresponding position. B) Frequency (%) of the preference for each nucleotide at the -2, -1, and +1 positions for A3B, A3F, A3G (* data from [2]) and AID (** data from [38]). C) Deaminase activity assay shows that the P210R mutation does not disrupt deaminase function when compared to WT A3G$_{CTD}$. D) A3G$_{CTD}$ catalytic parameter measurements and sequence preference as determined by a UDG-dependent cleavage assay (graphs shown in S1 Fig). Error values are based on fits to the hyperbolic K$_d$ curve, $k_{obs} = (k_{chem}*[E])/(K_d+[E])$. The errors represent standard errors of the parameters. † Efficiency = k_{chem}/K_d, ‡ Preference = Efficiency (CC**C**A)/Efficiency(CC**C**T/G).

the +1 position was consistent with previous observations [2, 38]. However, the A3G$_{CTD}$ P210R mutant had a three-fold decrease in A:T preference (from 2.2 to 0.7) compared to WT, indicating a switch of preference from adenosine to thymidine at the +1 position. This is consistent with the fact that the AID, which prefers thymidine in the +1 position, has an arginine at the corresponding residue in loop 1. Together, these results show that mutations in the -1 nucleotide-binding pocket perpetuate changes in the +1 nucleotide preference, supporting that the two nucleotide-binding pockets are not independent, but are structurally connected and communicate with each other.

Mutating loop 1 of A3G has a broad effect on hotspot mutation rates *in vivo*

To confirm that mutating the -1 pocket affects the preference for the other nucleotide positions in the hotspot, we evaluated the full-length A3G P210R mutant on both its antiviral activity and the sequence context of the G-to-A hypermutation in human cell culture. We produced VSV-G pseudotyped viruses in HEK293T cells in the presence of A3G-WT or A3G-P210R and determined infectivity in TZM-bl cells in a single round of replication, as previously described [41] (Fig 5A). In line with previous studies [1, 5, 25, 42], WT and mutant A3G potently blocked replication of HIV-1 produced in their presence compared to viruses generated in
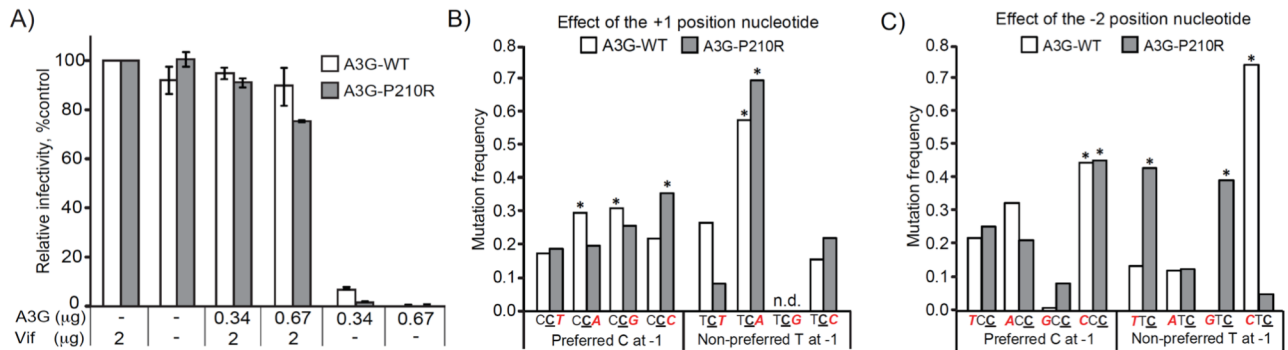
Fig 5. P210 is important for selection of the -2 and +1 nucleotides at the A3G deamination sites. A) Relative single-cycle infectivity of VSV-G-psuedotyped HIV-1Δ*vif* viruses produced in the presence or absence of A3G-WT or A3G-P210R. Mean of two independent experiments done in triplicates are shown relative to the "no A3G" control (set to 100%); error bars, standard deviation. B and C) Effect of A3G-P210R substitution on the relative mutation frequencies at the +1 position nucleotide (b) and the -2 position nucleotide (c) with the preferred C at the -1 position (5'C<u>C</u>) or the non-preferred T at the -1 (5'T<u>C</u>) position. B) In the + 1 position nucleotide, mutation frequencies for the preferred sites with a C at -1 position (5'C<u>C</u>), A3G-WT prefers C<u>C</u>A or C<u>C</u>G over C<u>C</u>T and C<u>C</u>C; A3G-P210R prefers C<u>C</u>C over C<u>C</u>A, C<u>C</u>G, and C<u>C</u>T, but has no significant difference in preference between C<u>C</u>A and C<u>C</u>T. For the non-preferred sites with a T at -1 position (5'T<u>C</u>), A3G-WT and A3G-P210R both prefer T<u>C</u>A over T<u>C</u>T, T<u>C</u>G, or T<u>C</u>C. C) In the -2 position nucleotide, mutation frequencies for the preferred sites with a C at the -1 position (5'C<u>C</u>), both A3G-WT and A3G-P210R prefer C<u>C</u>C over T<u>C</u>C, A<u>C</u>C, or G<u>C</u>C. For the non-preferred sites with a T at the -1 position (5'T<u>C</u>), A3G-WT prefers CT<u>C</u> over TT<u>C</u>, AT<u>C</u>, or GT<u>C</u>, whereas A3G-P210R prefers TT<u>C</u> or GT<u>C</u> over AT<u>C</u> and CT<u>C</u>. The significantly preferred nucleotide in the -2 or +1 positions are indicated (*P < 0.001). The number of sites, number of mutations, and relative mutation frequencies are shown in S1 Table.

their absence (Fig 5A). The presence of viral infectivity factor (Vif) completely averted both WT and P210R mutant A3G antiviral activity. These results demonstrated that the P210R substitution did not affect the overall antiviral activity of A3G or its susceptibility to Vif-mediated degradation.

Despite a comparable overall antiviral activity of WT A3G and the P210R mutant, our biochemical data suggest that each A3G construct exhibits different nucleotide preference at the +1 position, even though P210 is interacting with the -1 nucleotide (Fig 4D). We therefore infected CEM-SS T cells with virus produced in the presence of A3G-WT or A3G-P210R, and analyzed G-to-A hypermutation patterns in sequences of proviral DNAs (Fig 5B). The overall 5'GG-to-AG hypermutation frequency was similar for both WT and P210R A3G, indicating that the P210R mutation did not significantly alter the deamination activity of the enzyme; however, substrate selectivity was altered for the +1 position. WT A3G preferred adenosine at the +1 position (5'C<u>C</u>A:5'C<u>C</u>T ratio = 0.28/0.17; P < 0.0001), whereas the P210R mutant had a substantial decrease in its preference for an adenosine (5'C<u>C</u>A:5'C<u>C</u>T ratio = 0.11/0.10; P > 0.05) (Fig 5B and S1 Table). This change in 5'C<u>C</u>A:5'C<u>C</u>T preference is consistent with, but not as drastic as observed in our biochemical studies using the A3G_CTD (Fig 4D), likely because the A3G_NTD also interacts with the DNA [43] and thus, contributes to substrate specificity. Nonetheless, the substantial change in 5'C<u>C</u>A:5'C<u>C</u>T ratio in the cell-based assay using full-length A3G variants corroborates the notion that the nucleotide binding pockets are tightly entwined.

Mutating P210 also affects the nucleotide preference at the -2 position in the context of the less preferred thymidine at the -1 position (5'T<u>C</u>), but not in the context of the preferred cytidine at the -1 position (5'C<u>C</u>). Both WT A3G and P210R A3G preferred cytidine compared to thymidine at the -2 position when the -1 position was a cytidine (WT 5'C<u>C</u>C:5'T<u>C</u>C ratio = 0.37/0.18; P < 0.0001 and P210R 5'C<u>C</u>C:5'T<u>C</u>C ratio = 0.21/0.12; P < 0.0001) (Fig 5C and S1 Table). When the -1 nucleotide was the less preferred thymidine, WT A3G maintained a preference for cytidine at the -2 position (5'CT<u>C</u>:5'TT<u>C</u> ratio = 0.07/0.01; P < 0.0001) (Fig 5C and S1 Table). In contrast, P210R had a substantially increased preference for thymidine and guanidine compared to cytidine at the -2 position (5'TT<u>C</u>:5'CT<u>C</u> ratio = 0.05/0.01; P < 0.0001 and

5'GT$\underline{C}$:5'CT$\underline{C}$ ratio = 0.05/0.01; P < 0.0001) (Fig 5C and S1 Table). In summary, changes to the -1 nucleotide pocket, specifically the P210 residue, affect the nucleotide preference at the +1 position when the -1 nucleotide is the preferred cytidine (5'C$\underline{C}$) and at the -2 position when the -1 nucleotide is the non-preferred thymidine (5'T$\underline{C}$). Therefore, these results lend support to our biochemical and structural studies with Pot1A3G$_{CTD}$, demonstrating that changes to the -1 nucleotide pocket, specifically the P210 residue, affect the entirety of the hotspot preference of A3G.

Discussion

A3G is one of the most potent restriction factors of HIV-1, yet its mechanism of substrate selection is still poorly understood. A3G prefers to deaminate deoxycytidines in the hotspot sequence 5'-CC$\underline{C}$A (where the deaminated C is underlined) [2, 27]. Despite this hotspot preference, deamination can still occur to a lesser extent when other nucleotides are at the flanking positions. For example, A3G is capable of cytidine deamination with any nucleotide at the +1 position, albeit at different frequencies as shown in Figs 4 and 5B and 5C. In addition, it has been shown that the cytidine at the -1 position (the 5' side of the deaminated cytidine) can also be deaminated [35]. In this study, we used the novel fusion of Pot1 to the A3G$_{CTD}$ to capture the low affinity A3G$_{CTD}$-ssDNA interaction, identifying a non-preferred adenosine in the -1 nucleotide-binding pocket of A3G$_{CTD}$. This is the first structure of an A3 bound to a non-preferred hotspot substrate. Since A3G is a highly processive enzyme [35], it frequently encounters purines in the -1 nucleotide binding pocket while scanning the DNA for its hotspot. It was unknown how A3G allows binding but discriminates against deaminating such substrates.

Our comparative structural analysis, biochemistry, and virology studies provide insight into how encountering a purine in the A3G -1 pocket would not result in deamination of a deoxycytidine at the 0 position. We show that unlike the A3A-DNA interactions with the preferred substrate, residue D316 does not flip in to interact with the Watson-Crick edge of the base (Fig 2D) and there is no selectivity toward the nucleotide. A3G instead uses the backbone of P210 in loop 1 to interact with the -1 nucleotide along the Hoogsteen edge of the non-preferred adenine, which causes a structural change in the A3G -1 nucleotide-binding pocket (Fig 3A). This structural change further perturbs the conformation of other nucleotide-binding sites; as recent structures [31, 32] and our biochemical and cell-based data show, these sites are clustered and inter-connected. In fact, mutating residues ^{210}PW^{211} causes deformations in the -1 nucleotide pocket that perpetuate preference changes throughout the entire hotspot sequence (Figs 4 and 5). Thus, these perturbations from binding non-preferred nucleotides in the -1 nucleotide pocket may shift the residues in the catalytic pocket making the local environment non-ideal for the deamination reaction.

Our analysis advances the understanding toward the process that allows A3G to find its preferred deamination site in DNA. During the process of DNA scanning, when A3G encounters a non-preferred sequence, the rearrangements in the binding pockets necessary to accommodate the nucleotide result in a conformation that is not suitable for deamination at the catalytic site (Fig 6B). Only when the enzyme encounters the hotspot sequence, the collaborative interactions of the preferred nucleotides and their binding pockets on A3G result in a catalytically productive conformation for deamination (Fig 6A).

Materials and methods

Protein expression and purification

His$_6$-Staph Nuclease (SN) with SARS-CoV Mpro cleavage site [44, 45] was inserted into the NcoI–BamHI site of pRSF-Duet-1 plasmid (Novagen), followed by the insertion of Pot1-
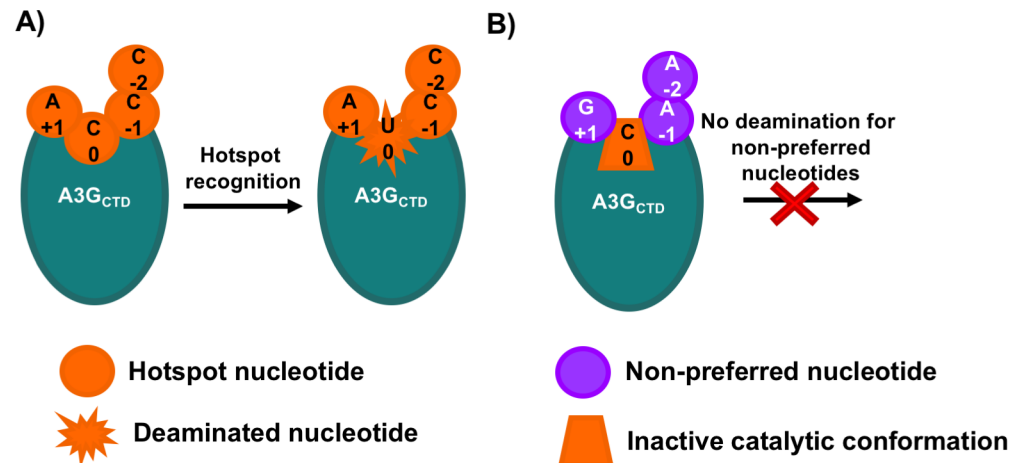
Fig 6. Schematic for DNA selection and nucleotide pocket communication. A) When A3G (represented by a teal oval) encounters a hotspot (preferred nucleotides represented by orange circles), the A3G is active and the cytidine in the 0 position is deaminated, resulting in a uridine at the 0 position (orange star). B) When A3G (teal oval) encounters non-preferred nucleotides flanking a cytidine (purple circles), it adapts an unfavorable conformation (orange trapezoid) at the catalytic site and no deamination occurs.

https://doi.org/10.1371/journal.pone.0195048.g006

A3G195-384-2K3A fusion gene into the EcoRI–XhoI site [23, 34, 46]. The wild-type A3G$_{CTD}$ was constructed from A3G191-384-2K3A gene inserted into NcoI–XhoI site of the expression vector pMAT9s, containing an N-terminal His$_6$-tag followed by maltose binding protein (MBP) and a SARS-CoV Mpro cleavage site [44]. All A3G mutants, including P210R and W211A were generated using this construct as a template. The vectors encoding A3G$_{CTD}$ mutants were made by QuickChange mutagenesis. All constructs were transformed and expressed in BL21 (DE3) *Escherichia coli* cells (Lucigen) grown in TB media to an OD600 of 0.6 and induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight at 16˚C. The cells were then harvested by centrifugation (5000 rpm, 10 min, 4˚C) and resuspended in lysis buffer [50 mM Tris (pH 7.5), 500mM NaCl, and 0.1mM Tris(2-carboxyethyl) phosphine (TCEP)]. Resuspension was followed by lysis with a microfluidizer. The lysate was centrifuged [13,000 rpm, 40 min, 4˚C] and proteins were purified by nickel affinity column (Qiagen) on FPLC (GE Healthcare). SN or MBP tag was removed by digestion with SARS-CoV Mpro protease overnight at 4˚C. The target protein was separated using a HiTrapQ anion exchange column (GE Heathcare) in 20 mM Tris (pH 8.0) using a 0–1000 mM NaCl (with 0.1mM TECP) gradient elution, followed by Superdex-200 gel-filtration column (GE Heath-care) in corresponding buffer [50 mM Tris (pH 7.0) (for Pot1-A3G$_{CTD}$) or 50 mM NaH$_2$PO$_4$ (pH 8.0) (for all A3G$_{CTD}$ and mutants), 100 mM NaCl and 0.1 mM TCEP]. The protein purity was examined by SDS-PAGE. Pot1-A3G$_{CTD}$ was mixed with a DNA oligonucleotide [30nt substrate: AGA AGA CC**C** AAA GAA GAG GAA GCA GGT TAC] at 1:1 molar ratio, and further purified using a Superdex-75 size-exclusion column. The protein/DNA complex was then concentrated to 3 mg/mL for crystallization.

Crystallization and data collection

Pot1-A3G$_{CTD}$/DNA complex crystals were grown at 20˚C using the microbatch-under-oil method by mixing equal amounts of sample [in buffer of 50 mM Tris (pH 7.0), 100 mM NaCl and 0.1 mM TCEP] and crystallization buffer [100 mM HEPES pH7, 200 mM LiCl, and 20% (w/v) Polyethylene glycol (PEG) 6000]. Crystals were cryo-protected by the crystallization buffer with 30% (v/v) PEG 400 and frozen in liquid nitrogen. Diffraction data were collected at

the National Synchrotron Light Source beamline X29A to the resolution of 2.9 Å. Data were processed using HKL2000 [47]. Analysis of the data showed that the crystal has close to perfect twinning. The data statistics are summarized in Table 1. Coordinates and structural factors have been deposited in the Protein Data Bank under the accession code 6BWY.

Structure determination and refinement

There were four Pot1-A3G$_{CTD}$ molecules in the asymmetric unit of the crystal. The structure was solved by molecular replacement using PHASER [48] with the A3G$_{CTD}$ structure (PDBID 3IR2 [24]) and the Pot1 structure (PDBID 1QZH [34]) as search models. Clear electron density of Pot1-A3G$_{CTD}$, including that for the Pot1 cognate DNA, was evident in the electron density map. Additional electron density was observed for the adenine at the 5' side of the Pot1 cognate DNA sequence. Furthermore, weak electron density 5' to the density was also observed but the quality was not sufficient for model building. Rigid body and iterative rounds of restrained refinement (including amplitude-based twin refinement) were carried out using Refmac5 [49], followed by rebuilding the model to the 2Fo-Fc and the Fo-Fc maps using Coot [50]. Non-crystallographic symmetry restraints were applied in the refinement cycles. The final model has an R$_{work}$/R$_{free}$ of 23.3%/28.9%. The refinement statistics are summarized in Table 1. The structure was analyzed and illustrated with Coot and PyMOL [51].

Deaminase assay using fluorescent-tagged ssDNA substrates

The substrate oligos containing CC**C**A, CC**C**T, and CC**C**G (IDT, Coralville, IA) were labeled with 6-carboxyfluorescein (6-FAM) at the 5' terminus. 2 nM 6-FAM-labeled oligos were incubated at 37°C for different time lengths, with 30 μg A3G protein samples and 5 units of Uracil-DNA Glycosylase (UDG) (New England BioLabs, Ipswich, MA). The abasic sites were then hydrolyzed by a 30-minute incubation with 0.25 M NaOH, which was followed by the addition of 20 μL of 1M Tris-HCl, pH 8.0. The reaction products were separated on a TBE-Urea PAGE gel (Life Technologies, Carlsbad, California). Gel bands were imaged with a CCD imager.

Radiolabeling of primers for *in vitro* kinetics

The substrate 30-mer oligos 5'—TAG AAA GGG AGA CC**C** AAA GAG GAA AGG TGA—3', 5'—TAG AAA GGG AGA CC**C** GAA GAG GAA AGG TGA—3', and 5'—TAG AAA GGG AGA CC**C** TAA GAG GAA AGG TGA—3' (IDT, Coralville, IA) were radiolabeled at the 5' terminus with [γ-^{32}P] ATP (Perkin Elmer, Waltham, MA) using T4 polynucleotide kinase (New England Biolabs, Ipswich, MA), as described previously [35, 52]. Radiolabeled oligos were desalted using a Bio-Spin 6 column (Bio-Rad Laboratories, Hercules, CA).

A3G *in vitro* single-turnover kinetics

WT or P210R A3G$_{CTD}$ was buffer exchanged into Reaction Buffer (20 mM Tris pH 8.0, 1 mM DTT). Enzyme at varying concentrations (1 μM– 50 μM) was incubated at 37°C for 5 minutes and reactions were induced with 40 nM of ^{32}P-labeled DNA oligomer. At given time points, 8 μL of the enzyme-oligo mixture was removed from the mix and quenched by the addition of 12 μL of quench buffer (50 mM EDTA, pH 8.0, final concentration) preheated to 95°C. After a 5-minute incubation at 95°C, the quenched mixture was incubated at 37°C for 5 minutes. Five units of UDG (New England BioLabs, Ipswich, MA) were incubated with the A3G-quenched mixture for two hours to cleave free uracil from any uracil-containing oligomers formed by A3G catalysis. The abasic sites were then hydrolyzed by a 30-minute incubation with 0.25 M NaOH, which was followed by the addition of 20 μL of denaturing PAGE dye.

The reaction products were separated on a 20% denaturing PAGE gel. Gel band intensities were measured by Bio-Rad Phosphorimager (Bio-Rad Laboratories) and analyzed by Quantity One software (Bio-Rad Laboratories). The ratio of the intensities of cleaved to uncleaved oligomer at each time point were plotted using Kaleidagraph (version 4.03, Synergy Software) and the rate at a given concentration of enzyme was fit to a single exponential curve, *Percent converted = A(1-e$^{-kobs*time}$)*, where A is maximum conversion (~100%) and k_{obs} is the single-turnover rate. The resultant rates at varying concentrations of enzyme were plotted using Kaleidagraph and fit to a hyperbolic K_d curve, $k_{obs} = (k_{chem}*[E])/(K_d+[E])$, where k_{chem} is the rate of chemistry and K_d is the dissociation constant. Errors given are standard errors of parameters.

Cell culture, plasmids, transfections, and virus production

HEK293T, TZM-bl, and CEM-SS cell lines were obtained from the American Type Culture Collection. HEK293T and TZM-bl cell lines were maintained in Dulbecco's modified Eagle's medium and CEM-SS cell line was maintained in RPMI 1640 medium (Corning Cellgro). Both media were supplemented to contain 10% fetal calf serum (Hyclone), 100 IU/ml penicillin and 100 µg/ml streptomycin (GIBCO).

A3G-P210R mutant was generated by site-directed mutagenesis (QuickChange Lightening site-directed mutagenesis kit, Agilent Technologies) using the following primers: P210R_sense: 5'ATTCACTTTCAACTTTAACAATGAACGGTGGGTCAGAGGAC3' P210R_antisense: 5'GTCCTCTGACCCACCGTTCATTGTTAAAGTTGAAAGTGAAT3'.

All viruses were prepared using a previously described HIV-1 vector pHDV-eGFP pseudotyped by co-transfecting with phCMV-G plasmid, which expresses vesicular stomatitis virus glycoprotein (VSV-G) [42, 53–58]. Briefly, we co-transfected pHDV-eGFP (1.0 µg), pHCMV-G (0.25 µg), and either 0.34 µg or 0.67 µg of pFlag-WT-A3G or pFlag-P210R-A3G expression plasmids in the presence or absence of pcDNA-hVif using polyethylenimine (PEI) as previously described [42, 53–55]. Virus-containing supernatant was clarified by filtering through a 0.45-µm filter and kept at -80˚C until use.

Virus infectivity and hypermutation analysis

Virus p24 CA amounts were determined using enzyme-linked immunosorbent assay (Xpress-Bio). TZM-bl indicator cells [59] were infected using equivalent p24 CA amounts of viruses, and infectivity was determined by measuring luciferase enzyme activity using Britelite luciferase solution (PerkinElmer) and a LUMIstar Galaxy luminometer (PerkinElmer).

For the hypermutation pattern analysis, CEM-SS cells (300,000 cells/well in 24-well plate) were infected with HIV-1Δ*vif* virus produced in the presence of 0.67 µg A3G. Cell pellets were collected 72 h post-infection and total DNA was extracted using QIAamp DNA blood minikit (Qiagen Inc). The reverse transcriptase (RT) coding region of HIV was PCR amplified using primers (HIV-10 FW: GGACAGCTGGACTGTCAATGACATAC and HIV-11 rev: GTTCATTTCCTC CAATTCCTTTGTGTG) and cloned into the pCR2.1-TOPO backbone vector using TOPO TA cloning kit (Invitrogen). To avoid bias in our analysis, we selected an 893-nt fragment that has a comparable number of 5'-C**C**A (19 editing sites) and 5'-C**C**T (20 editing sites) trinucleotide hotspots; the fragment also has 61 preferred 5'C**C** editing sites and 76 less preferred 5'T**C** editing sites. Individual clones were purified, sequenced, and analyzed for evidence of A3G-mediated G-to-A hypermutation using a previously described custom-written MatLab program [41].

Statistical analysis

All comparisons were made based on the WT-A3G for each mutant using Fisher's Exact test. Probability values (*P*) < 0.05 were considered significant.

Supporting information

S1 Fig. Kinetics of the Pot1A3G$_{CTD}$ deamination reaction. A) Representative kinetic curve to determine the rate of deamination of WT A3G$_{CTD}$. Shown here is the result for 5μM A3G$_{CTD}$. B) The kinetics plot results for the WT A3G$_{CTD}$ and P210R A3G$_{CTD}$ reaction on CC**C**A, CC**C**G, and CC**C**G substrates over a range of A3G concentrations. All data were analyzed and summarized in Fig 4D.
(TIF)

S1 Table. Total number of G-to-A mutations was determined for 93 proviruses produced in the presence of A3G-WT and 53 proviruses produced in the presence of A3G-P210R. Hypermutation is defined as $\geq$ 2 G-to-A mutations per clone (the no A3G control, on average, had <1 G-to-A mutations per clone). Each clone contained 61 5'C**C** and 76 5'T**C** target sites. Mutations/site = total mutations/[sites/clone × no. of clones]. For all conditions, the mutation frequencies for each nucleotide are shown relative to the total mutations/site as determined by the +1 or -2 position nucleotides. Relative preference of nucleotides at the +1 or -2 position in both the 5'-C**C** and 5'-T**C** edited sites are plotted for virions produced in the presence of A3G-WT or A3G-P210R.
(TIF)

Acknowledgments

Author Contributions

Conceptualization: Samantha J. Ziegler, Chang Liu, Vinay K. Pathak, Karen S. Anderson, Yong Xiong.

Data curation: Chang Liu, Mark Landau, Belete A. Desimmie, Ryan C. Burdick.

Formal analysis: Samantha J. Ziegler, Chang Liu, Mark Landau, Olga Buzovetsky, Belete A. Desimmie, Qi Zhao, Ryan C. Burdick.

Funding acquisition: Vinay K. Pathak, Karen S. Anderson, Yong Xiong.

Investigation: Samantha J. Ziegler, Chang Liu, Mark Landau, Olga Buzovetsky, Belete A. Desimmie, Qi Zhao.

Project administration: Vinay K. Pathak, Karen S. Anderson, Yong Xiong.

Resources: Vinay K. Pathak, Karen S. Anderson, Yong Xiong.

Supervision: Vinay K. Pathak, Yong Xiong.

Validation: Chang Liu, Mark Landau, Tomoaki Sasaki.

Visualization: Samantha J. Ziegler.

Writing – original draft: Samantha J. Ziegler, Olga Buzovetsky.

Writing – review & editing: Samantha J. Ziegler, Olga Buzovetsky, Belete A. Desimmie, Tomoaki Sasaki, Ryan C. Burdick, Karen S. Anderson, Yong Xiong.

References

1. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature. 2002; 418(6898):646–50. https://doi.org/10.1038/nature00939 PMID: 12167863.

2. Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol. 2004; 14(15):1392–6. https://doi.org/10.1016/j.cub.2004.06.057 PMID: 15296758.

3. Chiu YL, Greene WC. APOBEC3G: an intracellular centurion. Philos Trans R Soc Lond B Biol Sci. 2009; 364(1517):689–703. https://doi.org/10.1098/rstb.2008.0193 PMID: 19008196; PubMed Central PMCID: PMCPMC2660915.

4. Malim MH. APOBEC proteins and intrinsic resistance to HIV-1 infection. Philos Trans R Soc Lond B Biol Sci. 2009; 364(1517):675–87. https://doi.org/10.1098/rstb.2008.0185 PMID: 19038776; PubMed Central PMCID: PMCPMC2660912.

5. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature. 2003; 424(6944):99–103. https://doi.org/10.1038/nature01709 PMID: 12808466.

6. Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. Nat Struct Mol Biol. 2010; 17(2):222–9. https://doi.org/10.1038/nsmb.1744 PMID: 20062055; PubMed Central PMCID: PMCPMC2921484.

7. Harris RS, Hultquist JF, Evans DT. The restriction factors of human immunodeficiency virus. J Biol Chem. 2012; 287(49):40875–83. https://doi.org/10.1074/jbc.R112.416925 PMID: 23043100; PubMed Central PMCID: PMCPMC3510791.

8. Siriwardena SU, Chen K, Bhagwat AS. Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases. Chem Rev. 2016; 116(20):12688–710. https://doi.org/10.1021/acs.chemrev.6b00296 PMID: 27585283; PubMed Central PMCID: PMCPMC5528147.

9. Conticello S. The AID/APOBEC family of nucleic acid mutators. Genome Biology. 2008; 9(6):229. https://doi.org/10.1186/gb-2008-9-6-229 PMID: 18598372

10. Harris RS, Liddament MT. Retroviral restriction by APOBEC proteins. Nat Rev Immunol. 2004; 4 (11):868–77. https://doi.org/10.1038/nri1489 PMID: 15516966.

11. Conticello SG, Langlois MA, Yang Z, Neuberger MS. DNA deamination in immunity: AID in the context of its APOBEC relatives. Adv Immunol. 2007; 94:37–73. https://doi.org/10.1016/S0065-2776(06)94002-4 PMID: 17560271.

12. Suspene R, Sommer P, Henry M, Ferris S, Guetard D, Pochet S, et al. APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. Nucleic Acids Res. 2004; 32(8):2421–9. https://doi.org/10.1093/nar/gkh554 PMID: 15121899; PubMed Central PMCID: PMCPMC419444.

13. Yu Q, Konig R, Pillai S, Chiles K, Kearney M, Palmer S, et al. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol. 2004; 11(5):435–42. https://doi.org/10.1038/nsmb758 PMID: 15098018.

14. Sadler HA, Stenglein MD, Harris RS, Mansky LM. APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. J Virol. 2010; 84(14):7396–404. https://doi.org/10.1128/JVI.00056-10 PMID: 20463080; PubMed Central PMCID: PMCPMC2898230.

15. Delviks-Frankenberry KA, Nikolaitchik OA, Burdick RC, Gorelick RJ, Keele BF, Hu WS, et al. Minimal Contribution of APOBEC3-Induced G-to-A Hypermutation to HIV-1 Recombination and Genetic Variation. PLoS Pathog. 2016; 12(5):e1005646. https://doi.org/10.1371/journal.ppat.1005646 PMID: 27186986; PubMed Central PMCID: PMCPMC4871359.

16. Jern P, Russell RA, Pathak VK, Coffin JM. Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. PLoS Pathog. 2009; 5(4):e1000367. https://doi.org/10.1371/journal.ppat.1000367 PMID: 19343218; PubMed Central PMCID: PMCPMC2659435.

17. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet. 2013; 45(9):970–6. https://doi.org/10.1038/ng.2702 PMID: 23852170; PubMed Central PMCID: PMCPMC3789062.

18. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 2013; 494(7437):366–70. https://doi.org/10.1038/nature11881 PMID: 23389445; PubMed Central PMCID: PMCPMC3907282.

19. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, et al. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell. 2000; 102(5):565–75. PMID: 11007475.

20. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500(7463):415–21. https://doi.org/10.1038/nature12477 PMID: 23945592; PubMed Central PMCID: PMCPMC3776390.

21. Chen KM, Harjes E, Gross PJ, Fahmy A, Lu Y, Shindo K, et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature. 2008; 452(7183):116–9. https://doi.org/10.1038/nature06638 PMID: 18288108.

22. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, et al. Crystal structure of the antiviral APOBEC3G catalytic domain and functional implications. Nature. 2008; 456(7218):121–4. https://doi.org/10.1038/nature07357 PMID: 18849968; PubMed Central PMCID: PMCPMC2714533.

23. Harjes E, Gross PJ, Chen KM, Lu Y, Shindo K, Nowarski R, et al. An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. J Mol Biol. 2009; 389(5):819–32. https://doi.org/10.1016/j.jmb.2009.04.031 PMID: 19389408; PubMed Central PMCID: PMCPMC2700007.

24. Shandilya SM, Nalam MN, Nalivaika EA, Gross PJ, Valesano JC, Shindo K, et al. Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. Structure. 2010; 18 (1):28–38. https://doi.org/10.1016/j.str.2009.10.016 PMID: 20152150; PubMed Central PMCID: PMCPMC2913127.

25. Desimmie BA, Delviks-Frankenberrry KA, Burdick RC, Qi D, Izumi T, Pathak VK. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. J Mol Biol. 2014; 426(6):1220–45. https://doi.org/10.1016/j.jmb.2013.10.033 PMID: 24189052; PubMed Central PMCID: PMCPMC3943811.

26. Furukawa A, Nagata T, Matsugami A, Habu Y, Sugiyama R, Hayashi F, et al. Structure and real-time monitoring of the enzymatic reaction of APOBEC3G which is involved in anti-HIV activity. Nucleic Acids Symp Ser (Oxf). 2009;(53):87–8. https://doi.org/10.1093/nass/nrp044 PMID: 19749273.

27. Kohli RM, Abrams SR, Gajula KS, Maul RW, Gearhart PJ, Stivers JT. A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. J Biol Chem. 2009; 284(34):22898–904. https://doi.org/10.1074/jbc.M109.025536 PMID: 19561087; PubMed Central PMCID: PMCPMC2755697.

28. Carpenter MA, Rajagurubandara E, Wijesinghe P, Bhagwat AS. Determinants of sequence-specificity within human AID and APOBEC3G. DNA Repair (Amst). 2010; 9(5):579–87. https://doi.org/10.1016/j.dnarep.2010.02.010 PMID: 20338830; PubMed Central PMCID: PMCPMC2878719.

29. Wang M, Rada C, Neuberger MS. Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. J Exp Med. 2010; 207(1):141–53. https://doi.org/10.1084/jem.20092238 PMID: 20048284; PubMed Central PMCID: PMCPMC2812546.

30. Rathore A, Carpenter MA, Demir O, Ikeda T, Li M, Shaban NM, et al. The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. J Mol Biol. 2013; 425(22):4442–54. https://doi.org/10.1016/j.jmb.2013.07.040 PMID: 23938202; PubMed Central PMCID: PMCPMC3812309.

31. Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nat Struct Mol Biol. 2017; 24(2):131–9. https://doi.org/10.1038/nsmb.3344 PMID: 27991903; PubMed Central PMCID: PMCPMC5296220.

32. Kouno T, Silvas TV, Hilbert BJ, Shandilya SMD, Bohn MF, Kelch BA, et al. Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. Nat Commun. 2017; 8:15024. https://doi.org/10.1038/ncomms15024 PMID: 28452355; PubMed Central PMCID: PMCPMC5414352.

33. Wedekind JE, Gillilan R, Janda A, Krucinska J, Salter JD, Bennett RP, et al. Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits. J Biol Chem. 2006; 281(50):38122–6. https://doi.org/10.1074/jbc.C600253200 PMID: 17079235; PubMed Central PMCID: PMCPMC1847398.

34. Lei M, Podell ER, Baumann P, Cech TR. DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. Nature. 2003; 426(6963):198–203. https://doi.org/10.1038/nature02092 PMID: 14614509.

35. Chelico L, Pham P, Calabrese P, Goodman MF. APOBEC3G DNA deaminase acts processively 3' —> 5' on single-stranded DNA. Nat Struct Mol Biol. 2006; 13(5):392–9. https://doi.org/10.1038/nsmb1086 PMID: 16622407.

36. Shindo K, Li M, Gross PJ, Brown WL, Harjes E, Lu Y, et al. A Comparison of Two Single-Stranded DNA Binding Models by Mutational Analysis of APOBEC3G. Biology (Basel). 2012; 1(2):260–76. https://doi.org/10.3390/biology1020260 PMID: 24832226; PubMed Central PMCID: PMCPMC4009770.

37. Lu X, Zhang T, Xu Z, Liu S, Zhao B, Lan W, et al. Crystal structure of DNA cytidine deaminase ABOBEC3G catalytic deamination domain suggests a binding mode of full-length enzyme to single-stranded

DNA. J Biol Chem. 2015; 290(7):4010–21. https://doi.org/10.1074/jbc.M114.624262 PMID: 25542899; PubMed Central PMCID: PMCPMC4326812.

38.   Kohli RM, Maul RW, Guminski AF, McClure RL, Gajula KS, Saribasak H, et al. Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. J Biol Chem. 2010; 285(52):40956–64. https://doi.org/10.1074/jbc.M110.177402 PMID: 20929867; PubMed Central PMCID: PMCPMC3003395.

39.   Yu K, Huang F-T, Lieber MR. DNA Substrate Length and Surrounding Sequence Affect the Activation-induced Deaminase Activity at Cytidine. Journal of Biological Chemistry. 2004; 279(8):6496–500. https://doi.org/10.1074/jbc.M311616200 PMID: 14645244

40.   Chen Z, Viboolsittiseri SS, O'Connor BP, Wang JH. Target DNA Sequence Directly Regulates the Frequency of Activation-Induced Deaminase-Dependent Mutations. The Journal of Immunology. 2012; 189(8):3970–82. https://doi.org/10.4049/jimmunol.1200416 PMID: 22962683

41.   Desimmie BA, Burdick RC, Izumi T, Doi H, Shao W, Alvord WG, et al. APOBEC3 proteins can copackage and comutate HIV-1 genomes. Nucleic Acids Res. 2016; 44(16):7848–65. https://doi.org/10.1093/nar/gkw653 PMID: 27439715; PubMed Central PMCID: PMCPMC5027510.

42.   Russell RA, Pathak VK. Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F. J Virol. 2007; 81(15):8201–10. https://doi.org/10.1128/JVI.00395-07 PMID: 17522216; PubMed Central PMCID: PMCPMC1951317.

43.   Xiao X, Li S-X, Yang H, Chen XS. Crystal structures of APOBEC3G N-domain alone and its complex with DNA. Nature Communications. 2016; 7:12193. https://doi.org/10.1038/ncomms12193 http://www.nature.com/articles/ncomms12193#supplementary-information. PMID: 27480941

44.   Xue X, Yang H, Shen W, Zhao Q, Li J, Yang K, et al. Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. J Mol Biol. 2007; 366(3):965–75. https://doi.org/10.1016/j.jmb.2006.11.073 PMID: 17189639.

45.   Truckses DM, Prehoda KE, Miller SC, Markley JL, Somoza JR. Coupling between trans/cis proline isomerization and protein stability in staphylococcal nuclease. Protein Science. 1996; 5(9):1907–16. https://doi.org/10.1002/pro.5560050917 PMID: 8880915

46.   Chen KM, Martemyanova N, Lu Y, Shindo K, Matsuo H, Harris RS. Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G. FEBS Lett. 2007; 581 (24):4761–6. https://doi.org/10.1016/j.febslet.2007.08.076 PMID: 17869248; PubMed Central PMCID: PMCPMC2014798.

47.   Otwinowski ZM, W. Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology. 1997; 276:307–26.

48.   McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40(Pt 4):658–74. https://doi.org/10.1107/S0021889807021206 PMID: 19461840; PubMed Central PMCID: PMCPMC2483472.

49.   Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr D Biol Crystallogr. 2004; 60(Pt 12 Pt 1):2184–95. https://doi.org/10.1107/S0907444904023510 PMID: 15572771.

50.   Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60(Pt 12 Pt 1):2126–32. https://doi.org/10.1107/S0907444904019158 PMID: 15572765.

51.   Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010.

52.   Ray AS, Basavapathruni A, Anderson KS. Mechanistic studies to understand the progressive development of resistance in human immunodeficiency virus type 1 reverse transcriptase to abacavir. J Biol Chem. 2002; 277(43):40479–90. https://doi.org/10.1074/jbc.M205303200 PMID: 12176989.

53.   Russell RA, Smith J, Barr R, Bhattacharyya D, Pathak VK. Distinct domains within APOBEC3G and APOBEC3F interact with separate regions of human immunodeficiency virus type 1 Vif. J Virol. 2009; 83(4):1992–2003. https://doi.org/10.1128/JVI.01621-08 PMID: 19036809; PubMed Central PMCID: PMCPMC2643761.

54.   Smith JL, Pathak VK. 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 (24):12599–608. https://doi.org/10.1128/JVI.01437-10 PMID: 20943965; PubMed Central PMCID: PMCPMC3004357.

55.   Smith JL, Izumi T, Borbet TC, Hagedorn AN, Pathak VK. HIV-1 and HIV-2 Vif interact with human APOBEC3 proteins using completely different determinants. J Virol. 2014; 88(17):9893–908. https://doi.org/10.1128/JVI.01318-14 PMID: 24942576; PubMed Central PMCID: PMCPMC4136346.

56.   Nguyen KL, llano M, Akari H, Miyagi E, Poeschla EM, Strebel K, et al. Codon optimization of the HIV-1 vpu and vif genes stabilizes their mRNA and allows for highly efficient Rev-independent expression. Virology. 2004; 319(2):163–75. PMID: 15015498.

57. Yee JK, Friedmann T, Burns JC. Generation of high-titer pseudotyped retroviral vectors with very broad host range. Methods Cell Biol. 1994; 43 Pt A:99–112. PMID: 7823872.

58. Unutmaz D, KewalRamani VN, Marmon S, Littman DR. Cytokine signals are sufficient for HIV-1 infection of resting human T lymphocytes. J Exp Med. 1999; 189(11):1735–46. PMID: 10359577; PubMed Central PMCID: PMCPMC2193071.

59. Wei X, Decker JM, Liu H, Zhang Z, Arani RB, Kilby JM, et al. Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. Antimicrob Agents Chemother. 2002; 46(6):1896–905. https://doi.org/10.1128/AAC.46.6.1896-1905.2002 PMID: 12019106; PubMed Central PMCID: PMCPMC127242.