HIV-1 mutations in HIV-1 Gag, protease, RT p66 and when they appear: Insights from an *in vitro* BSL2 assay on mutation rates and types.

Joshua Yi Yeo[a], Darius Wen-Shuo Koh[a], Ping Yap[a], Ghin-Ray Goh[a], & Samuel Ken-En Gan[a,b,c,#]

[a]*Antibody & Product Development Lab, BII, Agency for Science, Technology and Research (A*STAR), Singapore 138672.*

[b]*Experimental Drug Development Centre, Agency for Science, Technology and Research (A*STAR), Singapore 138670*

[c]*p53 Laboratory, Agency for Science, Technology and Research (A*STAR), Singapore 138648.*

Running Title: HIV-1 Gag, protease, RT p66: Mutation rates and types

#Corresponding author: Samuel_Gan@eddc.a-star.edu.sg; samuelg@bii.a-star.edu.sg

Antibody & Product Development Lab, BII, A*STAR

60 Biopolis Street, #B2 Genome

Singapore 138672

Tel: ,65 64070584

Keywords: HIV, mutation rate, Gag, protease, reverse transcriptase, drug resistance

Abstract

Drug resistant mutants of HIV-1 is largely credited to its error prone HIV-1 RT, making it an important aspect for investigation. Previous HIV-1 RT studies rely on reporter genes (LacZ) as template, leaving the effects on HIV genes still uncharacterized. To address this, we studied HIV-1 RT mutation rates and bias on the Gag, protease, and RT p66 in an *in-vitro* selection pressure free assay and found clinical mutations with a general avoidance of key sites and detrimental mutations in the backdrop of mutations rates: 4.71×10^{-5}, 6.03×10^{-5}, and 7.09×10^{-5} mutations/bp for Gag, protease and RT p66 respectively. Gag and p66 had significant 'A to G' hypermutations we attributed to cellular adenosine deaminases and from comparisons with silently mutated p66 sequences, we observed an increase in mutation rates (1.88×10^{-4} mutations/bp) and 'A to G' mutations in regions reminiscent of ADAR recognition motifs. Analysis of change in mutational free energies of the 'A to G' mutations revealed a general tendency to avoid destabilizing effects with the natural p66 gene codon usage providing barriers to effects of ADAR. Our study demonstrates the importance of studying mutation emergence in HIV genes to understand how fast drug resistance can emerge, and in this, provide potential transferable understanding to other viruses to how new viral disease and drug resistance can emerge.

Importance

Majority of previous studies on the mutation rate of HIV-1 RT utilises reporter genes as templates, and not on HIV-1 genes. Our study characterised the mutations of HIV-1 RT in HIV-1 Gag, protease and RT p66 in an *in vitro* selection pressure free assay, demonstrating that known drug resistance mutations can occur within a single replication cycle with avoidance of mutating key sites for functionality. The findings of this study on mutation emergence provides potential horizontal understanding in drug resistance emergence, as well as the emergence of zoonotic viruses.

Introduction

Emerging viral infections from RNA viruses have been shown to have a higher likelihood of undergoing genetic changes, leading to species jump (1) and efficient spread among humans (2). Among the RNA viruses, the simian immunodeficiency virus (SIV) lineages are believed to be ancestral to the human immunodeficiency viruses such asHIV-1 and HIV-2 (3, 4).

HIV-1 Gag and protease play vital roles in viral assembly and maturation for the production of infectious virions (5) where HIV-1 protease cleaves Gag and pol polyproteins into functional subunits (6). To prevent maturation, protease inhibitors (PI) compete with Gag for the active site of protease (7, 8). In PI resistance, drug resistance mutations on viral protease reduce affinity to PIs (8, 9), and the accumulation of such resistance mutations (10-12) induced by HIV-1 RT (13, 14), render PIs ineffective and confer cross-resistance, limiting clinical drug selection (8). In HIV, mutations have been reported on the substrate Gag to compensate for reduced viral fitness (8, 9, 14-16), with paired-mutations that work synergistically to that of protease (17-21).

The generation of mutations in HIV genes is mainly contributed by the error prone HIV-1 reverse transcriptase (RT), an asymmetric heterodimer consisting of the p66 and p51 subunits (22). The p66 subunit performs the key enzymatic functions, catalysing DNA polymerisation and cleaving the RNA of the RNA/DNA duplex (23, 24), while the p51 subunit plays a more supportive role to p66 (25).

Since most of the previous studies on HIV mutation rate were based on reporter genes (LacZ) and not HIV genes, there is a gap in the understanding of HIV RT mutation rates on HIV genes and how early drug resistance mutations emerge in the infection cycle (26). To fill this gap, we developed an *in-vitro* cell-based assay without translational, immune and drug selection pressures to characterize the basal HIV-1 RT mutations and biases on HIV-1 Gag,

protease, RT p66 genes, and a sequence variant of RT p66 which was also silently mutated codons to investigate the effect of codon usage on mutation rates. This *in vitro* assay allowed us to establish the native error rate of HIV-1 RT to establish the rate, position and bias in HIV-1 Gag, protease, RT p66.

Results

Characterization of HIV-1 Gag mutant variants

In our study, we identified and classified 'A to G' mutations as hypermutations when multiple mutations were found on the same sequence variant given that mutations are should statistically be consistently spread out.

We generated a total of 801 HIV-1 Gag sequences and found a mutation rate of 3.36 x 10^{-5} mutations/bp when excluding 'A to G' hypermutations and a slightly higher rate of 4.71 x 10^{-5} mutations/bp when including them, based on the assumption that the experiment setup successfully mimicked one biological replication cycle (see Table 1).

There were more transitions (n = 24 excluding 'A to G' hypermutations, n = 40 when including 'A to G' hypermutations) as compared to transversions (n = 12) shown in Fig. 1A. For nucleotide substitutions, 'A to G' substitutions were found to be the highest (n = 20, 37.7%) when including 'A to G' hypermutations. When excluded, 'T to C' substitutions were the highest (n = 9, 25.0%), followed by 'G to A' and 'G to T' (n = 6, 16.7%), 'C to A' and 'C to T' (n = 5, 13.9%), 'A to G' (n = 4, 11.1%), and 'T to G' (n = 1, 2.8%), with several substitutions not observed (see Fig. 1B and 1C).

Analysis at the amino acid level showed missense mutations (n = 30, 55.6%) to occur at approximately 2 times the frequency of silent mutations (n = 15, 27.8%, see Fig. 2) while nonsense mutations (n = 5, 9.3%) and frameshift mutations (n = 4, 7.5%) had lower occurrences (see Table S1 for full list of mutations).

Characterization of HIV-1 protease mutant variants

For HIV-1 protease, we generated a total of 640 sequences, finding a mutation rate of 6.03×10^{-5} mutations/bp (see Table 1), noting that no 'A to G' hypermutations were found for protease.

As with Gag, there were more transition mutations (n = 6) than transversions (n = 2) as shown in Fig. 1A. 'G to A' substitutions were found to be the highest (n = 3, 37.5%), followed by 'A to G' (n = 2, 25%), 'A to C', 'T to C' and 'C to A' (n = 1, 12.5%), with several substitutions not observed (see Fig. 1B and 1C).

At the amino acid level, missense mutations (n = 4, 36.4%) occurred at approximately 1.3 times the frequency of silent mutations (n = 3, 27.3%) and frameshift mutations (n = 3, 27.3%), whereas nonsense mutations (n = 1, 9.1%) were of lower occurrences (see Fig. 2, Table S2).

Characterization of HIV-1 RT p66 mutant variants

From 571 HIV-1 RT p66 subunit sequences, we calculated a mutation rate of 4.27×10^{-5} mutations/bp when excluding 'A to G' hypermutations and 7.09×10^{-5} mutations/bp when including them (see Table 1).

As with Gag and protease, there were more transition mutations (n = 18 excluding 'A to G' hypermutations, 45 including 'A to G' hypermutations) as compared to transversions (n = 11) as shown in Fig. 1A. 'A to G' substitutions were found to be the highest (n = 29, 51.8%) only when including 'A to G' hypermutations. When 'A to G' hypermutations were excluded, 'T to C' and 'G to T' substitutions were instead the highest occurrences (n = 8, 27.6%), followed by 'G to A' (n = 6, 20.7%), 'C to A' (n = 4, 12.5%), and 'A to G' and 'C to T' (n = 2, 6.9%). There were several substitutions not observed (see Fig. 1B and 1C).

Analysis after translation found relative consistency between Gag and protease where missense mutations (n = 33, 49.3%) occurred at 1.7 times the frequency of silent mutations (n = 20, 29.9%) as shown in Fig. 2. Nonsense mutations (n = 2, 3.0%) and frameshift mutations (n = 12, 17.9%) were of lower occurrences (Table S3).

Characterization of HIV-1 Codon Mutated RT p66 mutant variants

A total of 700 codon mutated p66 gave a calculated mutation rate of 6.53 x 10^{-5} mutations/bp when excluding 'A to G' hypermutations and 1.88 x 10^{-4} mutations/bp when including them (see Table 1).

Consistent with Gag and protease, there were more transition mutations (n = 37 excluding 'A to G' hypermutations, 176 including 'A to G' hypermutations) as compared to transversion (n = 29) shown in Fig. 1A. As with the wild-type p66, 'A to G' substitutions were found to be the highest (n = 139, 71.2%) only when including 'A to G' hypermutations. Excluding 'A to G' hypermutations, 'G to T' substitutions were found to be the highest (n = 17, 25.8%), followed by 'G to A' (n = 13, 19.7%) and 'T to C' (n = 13, 19.7%), 'A to G' (n = 7, 10.6%), 'C to A' (n = 8, 12.1%), 'C to T' (n = 4, 6.1%), 'A to T' (n = 2, 3.0%), and 'T to A' and 'T to G' substitutions (n = 1, 1.5%), with the absence of the following substitutions: 'A to C', 'C to G' and 'G to C' (see Fig. 1B and 1C).

At the amino acid level, relative consistency with the other HIV-1 genes was also observed with missense mutations (n = 109, 56.2%) occurring at approximately 1.6 times the frequency of silent mutations (n = 70, 36.1%, see Fig. 2). Nonsense mutations (n = 7, 3.6%) and frameshift mutations (n = 8, 4.1%) were at lower occurrences (Table S4).

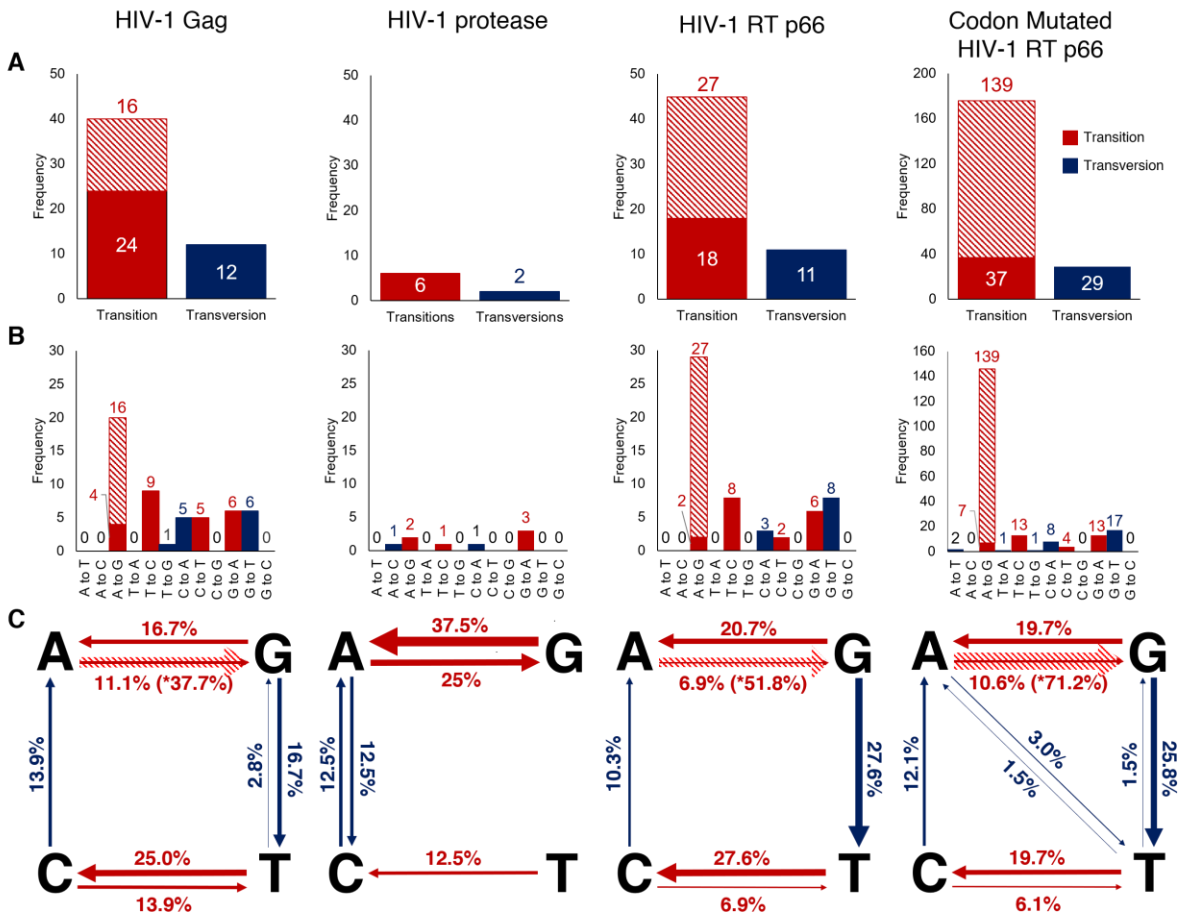

Fig. 1: Generated HIV-1 Gag, protease, RT p66 and Codon Mutated RT p66 nucleotide substitution mutations using HIV-1 RT.

(A) Bar chart of the transition and transversion mutations in HIV-1 Gag (n = 52), protease (n = 8), RT p66 (n = 56) and codon mutated p66 (n = 205). Transition and transversion mutations are shown in red and blue respectively. 'A to G' Hypermutations are shown separately in a diagonal striped pattern. (B) Bar chart showing substitution mutations in HIV-1 Gag, protease, RT p66 and codon mutated p66 sequences. (C**)** Schematic diagram of nucleotide substitutions occurring on the respective HIV-1 genes (expressed as a percentage).

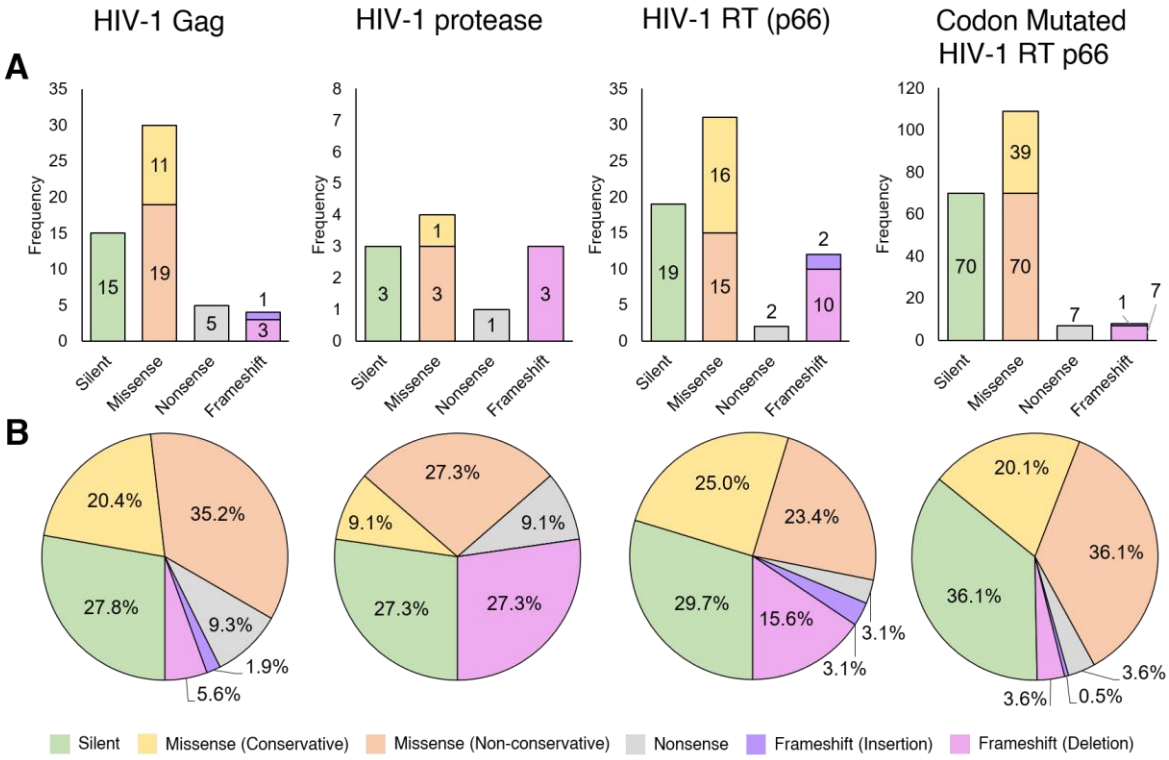

Fig. 2: Generated amino acid mutations using HIV-1 RT.

(A) Bar chart of amino acid mutations observed in HIV-1 Gag (n = 54), protease (n = 11), RT p66 (n = 64) and codon mutated RT p66 (n = 194). (B) Pie chart of the percentage proportions of the type of translation effects resulting from the mutations. For the codon mutated RT p66, two variants (20 and 80 a,b) containing deletions without frameshifts are not shown. Silent, missense (conservative), missense (non-conservative), nonsense, frameshift (insertion) and frameshift (deletion) mutations are shown in green, yellow, orange, grey, purple and pink respectively.

Table 1: Calculated error rates of HIV-1 RT on the respective HIV-1 genes

HIV-1 Gene	No. of clones	Nucleotide Length	Total no. of bases	Excluding 'A to-G' Hypermutations		Including 'A to G' Hypermutations	
				No. of mutations	Mutation rate (mutations/bp)	No. of mutations	Mutation rate (mutations/bp)
Gag	801	1485	1,189,485	40	3.36×10^{-5}	56	4.71×10^{-5}
protease	640	285	182,400	11	6.03×10^{-5}	-	-
RT p66	571	1680	959,280	41	4.27×10^{-5}	68	7.09×10^{-5}
Codon mutated RT p66	700	1617	1,131,900	74	6.53×10^{-5}	213	1.88×10^{-4}

Mutation rates were calculated as the ratio of the total number of mutations and the total number of nucleotide bases, comparable to the reported range of 1.8×10^{-5} – 6.67×10^{-4} as previously reviewed (26, 27).

Mutation Rates of the Domains of HIV-1 Gag, protease and RT p66

The HIV-1 genes were sorted into their respective domains and the mutation rates calculated (see Fig. 3). Within Gag, the overall mutation rates throughout the domains were within a fairly narrow range of 2.29 to 4.46 x 10^{-5} mutations/bp when excluding 'A to G' hypermutations, but increased to 8.16 x 10^{-5} when including them. 'A to G' hypermutations were found in the capsid (CA), nucleocapsid (NC), and p6 subunits, but not the matrix (MA), p2 and p1 domains. Mutations were not observed in the MA/CA ^{129}SQNY/PIV^{135}, CA/p2 ^{360}ARVL/AEA^{366}, p2/NC ^{374}ATIM/IQK^{380}, P1/p6 ^{445}PGNF/LQS^{451} cleavage sites and the ^{277}YSPTSIL^{283} capsid linker. However, mutations (A431D and F433L) were observed in the NC/p1 ^{429}RQAN/FLG^{435} cleavage site.

For HIV-1 protease, there were no 'A to G' hypermutations found throughout the gene and no mutations observed in the active site ^{25}DTG^{27}. For RT p66, when excluding hypermutations, there was a range of no mutations to 9.12 x 10^{-5} mutations/bp, which increased to 1.35 x 10^{-4} mutations/bp when including them. Despite the 2nd finger domain being of similar length as the 1st palm segment, only 'A to G' hypermutations were observed without other mutations. In the RT p51 and RNase H domain (p51-RNH) cleavage site AETF^{440}↓Y^{441}VD and catalytic triad D110, D185 and D186, there were also no mutations observed. In the codon mutated p66, the mutation rate was in a range of 2.98 x 10^{-5} – 1.06 x 10^{-4} mutations/bp when excluding hypermutations which increased the rate to 3.38 x 10^{-4} mutations/bp. Contrary to the wild-type p66, there were mutations observed in the 2nd finger domain, suggesting that codon usage of the wild-type 2nd finger domain may provide some barrier to mutations. Similarly, no mutations were observed in the p51-RNH cleavage site AETF^{440} ↓ Y^{441}VD despite the different codon usage. Interestingly, 'A to G' hypermutations were not observed in the 1st palm region of the codon mutated p66.

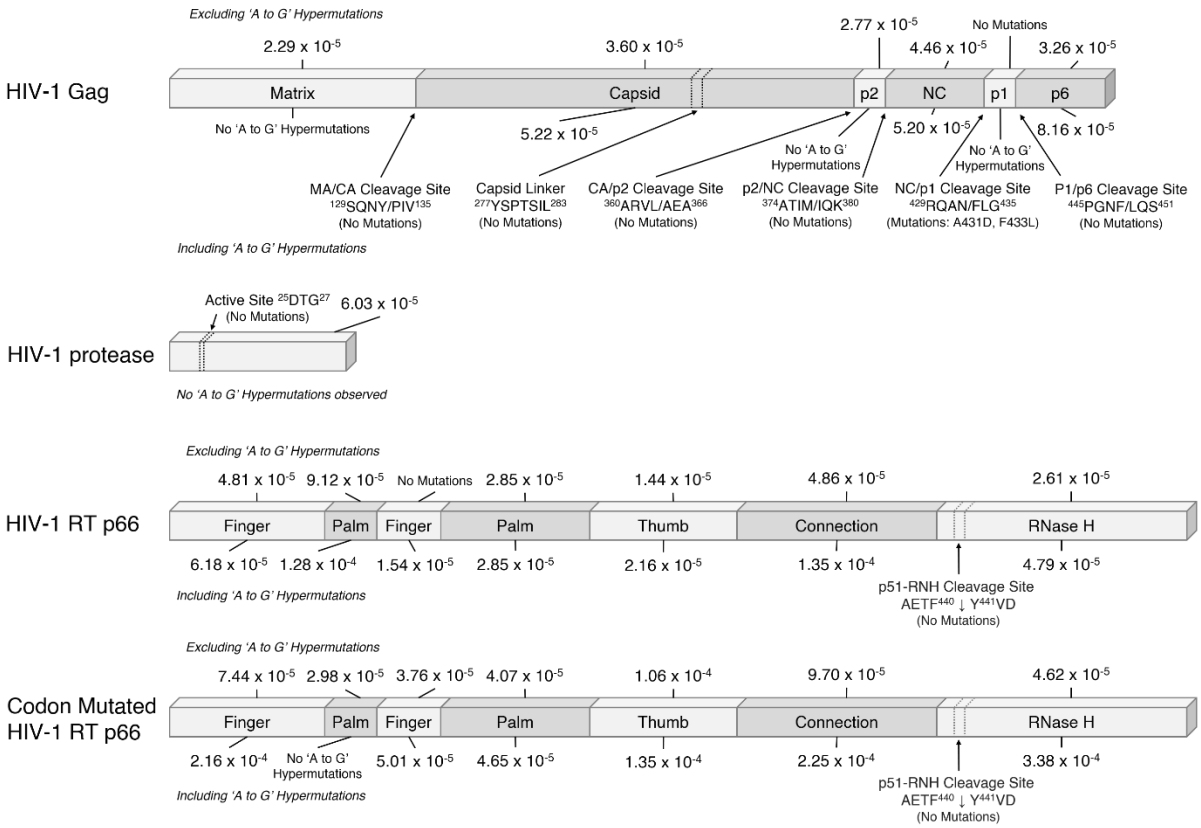

Fig. 3: Mutation rates (mutations/bp) calculated in the domains of HIV-1 Gag, protease, RT p66 and Codon Mutated RT p66.

Mutation rates were calculated over the total number of mutations and the total number of nucleotide bases, and counted separately excluding and including A to G hypermutations. Truncations that span across multiple domains were excluded in the calculations.

Sequence contexts of 'A to G' mutations

To determine the influence of host defence mechanism, especially ADAR in the overrepresentation of 'A to G' mutations in our assay, we analysed underlying sequence contexts neighbouring all multiple and single mutation events by comparing the sequence contexts +/-4 nucleotides (nt) from 'A to G' mutations with the sequence contexts of all Adenosines using two-sample logo analysis (Fig. 4A-C).

In our Gag analysis, 'A' and 'G' bases downstream of 'A to G' mutation sites at -1 and -4 nt positions on the 5' end were significantly depleted ($p < 0.05$). Using the Two Sample Logo of multiple and single substitution mutations, 'C' at the +1, +4 ($p < 0.05$) and -4 nt ($p < 0.01$) positions were enriched (see Fig. S1), with the former two being in agreement with ADAR preferences (see Fig. S1, Fig. 4B). On protease, only the 'T' at the -2 nt of the 3' end was found congruent with previously reported ADAR preferences (Fig. 4B).

Sequence analysis of RT p66 'A to G' mutations also showed significant depletion of both 'G' and 'C' bases at the -1 position, enrichments of 'T' bases at the -1 and -2 positions (that agree with reported usage sites of ADAR, see Fig. 4B-C, (28). 'A' bases were also enriched at the +4 position ($p < 0.05$). Such 'G**,** C and T' signature depletions and enrichments were also observed with significant differences for RT Codon Mutated p66 ($p<0.01$), where enrichments of 'T' were found at the -2 nt, 'A' bases at the +3 nt , 'G' at the +1 and +4 nt positions ($p<0.01$). On the 5'end, depletion of 'G' bases at the -4 position and 'C' and 'T' bases at the -3 nt position ($p < 0.05$), and depletion of 'A' bases at the -2 nt position ($p < 0.01$) were observed. On the 5' end, 'T' bases depletion at the -1 ($p<0.05$) and -4 nt positions ($p<0.01$) were found.

In both p66 wildtype and codon mutated variants, the large representation of multiple mutation variants appear to have masked the underlying sequence contexts of the single

substitution variants (Fig. S1) with identified signature ADAR preference sites with enrichments of ‘T’ at the 3’ -1 and -2 nt positions.

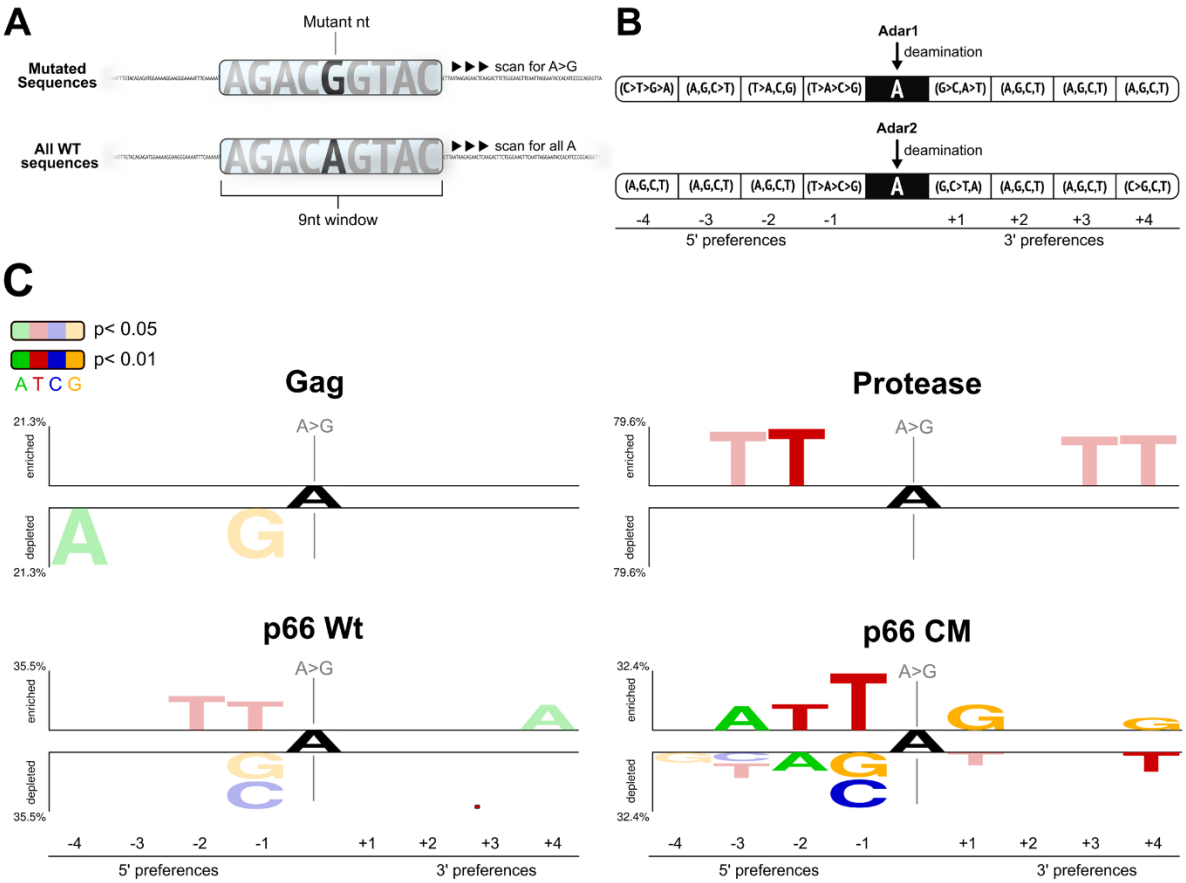

Fig. 4: Two Sample logos of HIV-1 Gag, protease, RT p66 (p66 Wt) and p66 Codon Mutated RT (p66 CM)

(A) Underlying sequence signatures 4 nt upstream and downstream of all 'A to G' mutations were determine by subtracting all 'A to G' mutation contexts with all possible background mutations from a 9 nt sliding window. (B) Previously reported sequence preferences of ADAR1 and 2 isoforms. Base preferences are shown with inequality signs, where the identity of the base did not significantly contribute to the editing of the 4 bases are delimited by commas (28). (C) Two Sample Logos showed enrichment and depletion of neighbouring nucleotides flanking A to G mutations. Bases were coloured opaque when $p < 0.01$, and translucent when $p < 0.05$. Statistics: two sample t-test without Bonferroni correction.

Effects of 'A to G' Mutations on Protein Thermostability

The distinct observed robustness of viruses to adapt to mutations rapidly in the face of drugs, immune pressures and the error prone nature of RTs may be attributed to the ability of viral protein structures to buffer the compromise of fitness by minimizing the change in mutational free energies ($\Delta\Delta G_{Mt}$) (29). When large enough to destabilize or stabilized (4kcal/mol), the protein is likely to have adopted a substantially different fold or be misfolded (30). Taking advantage of relatively large numbers of pooled 'A to G' mutations and other transition and transversion mutations, their distributions in $\Delta\Delta G_{Mt}$ were determined using protein design tools: Rosetta (31) and FoldX (32), to assay the effect of 'A to G' transitions on free energy distributions. . We pooled all the single substitution mutations and individual substitutions from generated multiple mutations to calculate the change in $\Delta\Delta G_{Mt}$ of our experimentally generated mutations, putting the threshold of significance for different distributions of $\Delta\Delta G_{Mt}$ at >1 kcal/mol for Rosetta (31) and $\Delta\Delta G_{Mt}$ >0.46 kcal/mol for FoldX (32).

From the distribution of the $\Delta\Delta G_{Mt}$ of all the mutations in Gag-p66 genes (from pooled $\Delta\Delta G_{Mt}$ taken from pooled $\Delta\Delta G_{Mt}$ values of Gag and RT p66 wildtype. Protease was excluded as it had no detected hypermutations), both protein design tools showed that most 'A to G' mutations exerted little or negative change in $\Delta\Delta G_{Mt}$ distributions, while other transversion mutations typically destabilized (Fig. 5A). There was also a secondary peak observed in the Rosetta results showing a notable distribution of 'A to G' transitions mutants with $\Delta\Delta G_{Mt}$ around a stabilizing -3.99 kcal/mol we attributed to Gag (Fig. 5A and 5B). From the FoldX result, there were fewer destabilizing mutations with $\Delta\Delta G_{Mt}$ at +5.17 kcal/mol and +7.52 kcal/mol for 'A to G' and other transition and transversion mutations respectively (Fig. 5A). This enforces the trend that 'A to G' mutations tend to experience less $\Delta\Delta G_{Mt}$, especially since respective secondary peaks were also seen in Gag and the wildtype RT p66 (Fig 5B and 5C).

Looking at the effect of mutations generated from mutations in individual Gag and RT p66 genes at the protein level, 'A to G' mutations generally had lower $\Delta\Delta G_{Mt}$ compared to other mutations from calculations by both protein design tools (Fig. 5A-C). Looking at the FoldX results, there were significant differences in distributions of 'A to G' $\Delta\Delta G_{Mt}$ compared to other mutations in both Gag and RT p66. Rosetta did not show any significant $\Delta\Delta G_{Mt}$ differences in this respect.

The significant differences and trends suggest that the HIV-1 genes, particularly RT p66 wildtype, were more robust to 'A to G' mutations at the protein level. We saw that 'A to G' mutations in codon mutated genes were more destabilizing compared to all other mutations; with significant differences in calculations generated by FoldX.

When the $\Delta\Delta G_{Mt}$ of multiple mutation variants arising from hypermutation events in Gag, p66 wildtype and codon mutated were modelled (Table S6), the average of all hypermutation calculations in Gag were: 0.37 kcal/mol, 1.75 kcal/mol; and in p66 wildtype: 2.53 kcal/mol, 3.25 kcal/mol as determined by Rosetta and FoldX respectively. This was in contrast to p66 codon mutated with an average doubling of destabilizing $\Delta\Delta G_{Mt}$ mutants to that of p66 wild-type at 6.48 kcal/mol, 7.00 kcal/mol for Rosetta and FoldX modelled values.

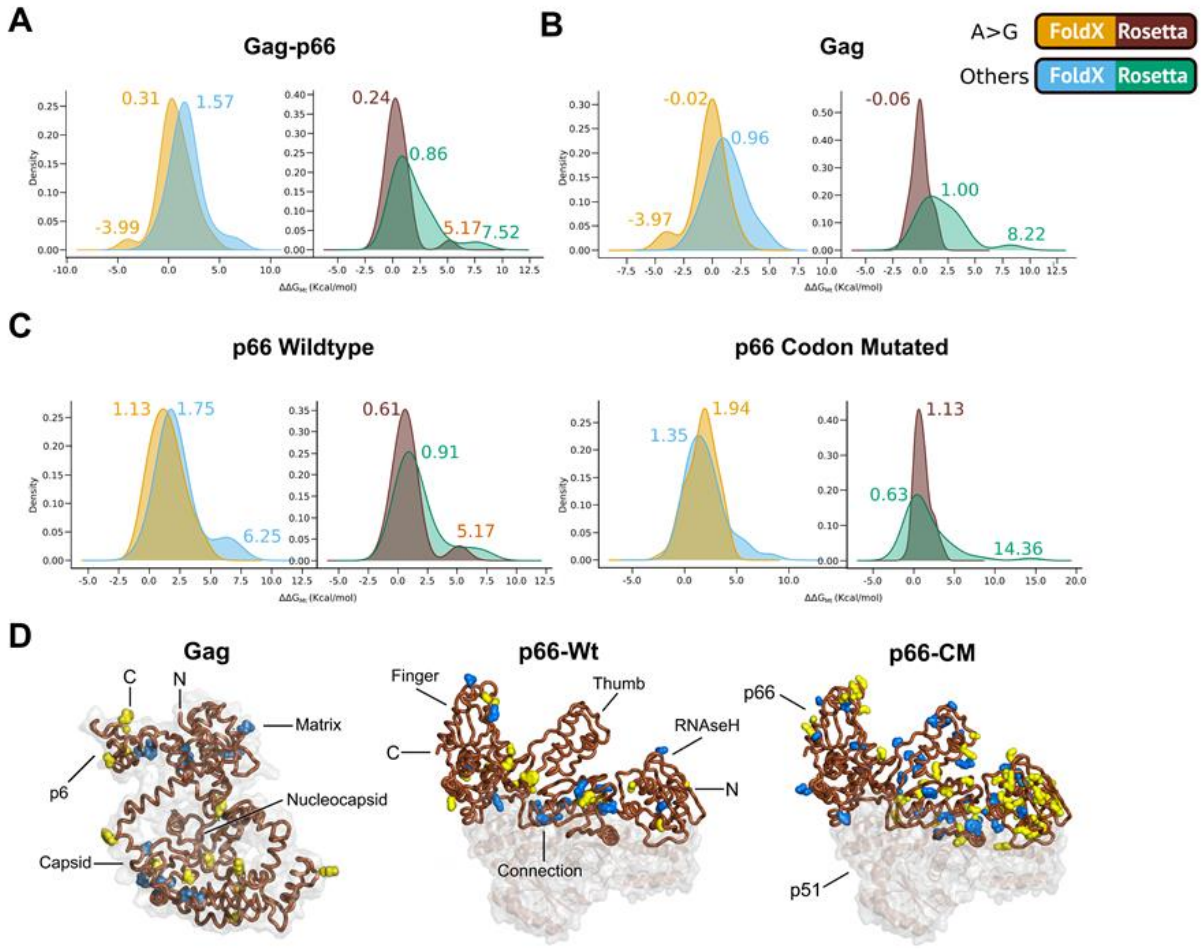

Fig. 5: Distributions of change in mutational free energies ($\Delta\Delta G_{Mt}$) of pooled amino acid in-vitro generated mutations.

Pooled $\Delta\Delta G_{Mt}$ of point mutations from (A) Gag-p66, (B) Gag (C) RT p66 wildtype (p66-Wt) and RT p66 codon mutated (p66-CM) genes . (D) Positions of 'A to G' mutations in the protein structures of Gag and p66 are shown in yellow while all other mutations are in blue. There were both 'A to G' and 'Other' mutations at position 462 of p66-CM which was labelled blue. The C and N terminals and locations of major domains and functional protein regions were labelled. Peaks are numbered with corresponding $\Delta\Delta G_{Mt}$ values. Change in mutational free enrgies were modelled with the Rosetta Cartesian_ddg and FoldX BuildPDB protocols. $\Delta\Delta G_{Mt}$ values or $\Delta\Delta G_{Mt}$ differences between different distributions of >1 kcal for Rosetta (29) and $\Delta\Delta$GMt >0.46 kcal/mol for FoldX (30) were considered to be significant. Note that Gag exists

in both the Compact and Extended states and since the former preceeds the later during viral assembly, the former was used.

Discussion

We set out to study the mutation rate of HIV-1 RT on HIV-1 genes: Gag, protease and RT p66 subunit in a single replication cycle using our *in vitro* BSL2 assay that is devoid of obvious selection pressures to establish the native mutation rate of HIV genes. Given the absence of protein translational, immune, fitness, and drug selection pressures in our system, the observed biases are presumed to be intrinsic to HIV-RT and also host cell factors. Although mutations can be contributed from plasmid hosts (DH5α and/or HEK293F), Q5 polymerase and sequencing artefacts, these were unlikely given that the observed phenotypic mutations/bp/replication for *E. coli* and mammalian cells were estimated to be 5.4×10^{-10} and 5.0×10^{-11} respectively (33), and that Q5 polymerase is among the most accurate commercial high-fidelity polymerase (34). Admittedly, there may be a role for primer selection bias when priming the first few bases of the genes, and as such, these bases were excluded from analysis and calculation of mutation rates.

While numerous *in vitro* studies have attempted to establish the error rate of HIV-1 RT (as previously reviewed in (26, 27)), we found only two that utilised HIV-1 genes as their template (35, 36). In our assay, we found the mutation rates of HIV-1 Gag, protease, RT p66 and codon mutated p66 to be 4.71×10^{-5}, 6.03×10^{-5}, 7.09×10^{-5} and 1.88×10^{-4} mutations/bp (inclusive of hypermutations) respectively, within the range of 1.8×10^{-5} – 6.67×10^{-4} mutations/bp previously reported (26, 27). Across the HIV-1 genes, we observed a predominance of transition mutations that is consistent with the characterization across most phyla, previously hypothesized for conservation of protein functions (37-39). Comparing the types of amino acid introduced, we observed a general trend with missense mutations having the highest frequency, followed by silent mutations, frameshifts and nonsense. This is notable because even in the absence of selection pressures to maintain protein translatability, frameshift

and nonsense mutations (with more detrimental effects) were occurring at lower frequencies, possibly constrained by in-built genetic code probabilities (40)

When changing the HIV-1 gene templates, a change in mutation rates, type of mutations and mutational biases was also observed. In general, there were no 'A – T' or 'C – G' mutations in the HIV genes studied, which were only detected in the codon mutated p66 (see Fig. 1), suggesting influence of the sequence and codon usage. Comparing the HIV-1 codon mutated RT p66 to wild-type HIV-1 RT p66, the mutation rate (1.88×10^{-4} mutations/bp) increased by 2.7 folds as compared to 7.09×10^{-5} mutations/bp. Interestingly, the RNase H domain of the codon mutated HIV-1 RT p66 was also found to have a 7.1 folds increase (3.38×10^{-4} mutations/bp), as compared to the wild-type (4.79×10^{-5} mutations/bp) suggesting RNA secondary structural protection/susceptibility to host adenosine deaminases that might be due to sequence secondary structure characteristics(28, 41, 42).

Of the genes, 'A to G' hypermutations were observed in HIV-1 Gag, RT p66 and codon mutated RT p66, but not in protease, possibly due to its short nucleotide length. In this, there is an interesting possibility that smaller genes avoided the effects of 'A to G' mutational effects that are likely caused by host adenosine deaminases such as double-stranded RNA-specific adenosine deaminase (ADAR). ADAR editing (by either ADAR1 and ADAR2) have been suggested to influence such cell based in-vitro RT-Fidelity assays (43). ADAR deaminates 'A' residues to inosine (I) residues (read as 'G' by ribosomes due to structural homology), leading to 'C' misincorporations in the plus strand DNA after reverse transcription, and thus resulting in 'G' residues being incorporated during second (minus) stand synthesis (44, 45). ADAR deaminases are host defence mechanism against viral infections and has great influence on the genomes of such RNA based lifeforms (46-48). As such, ADAR has been shown to display both pro-viral and anti-viral activity (49-52).

Given the lack of apparent ADAR effects on protease, partially explained by the lower numbers of ADAR preferred sites (Figure 4), protease is likely the better target than RT or Gag, which may explain the lower occurrences of PI resistance in clinical studies (53), alongside better potency than reverse transcriptase inhibitors (RTIs) (54).

With 'A to G' hypermutations, there would be a general accumulation towards G residues mutations for HIV-1 Gag, RT p66 and codon mutated RT p66, except for protease. Such 'G' biasness would result in increased glycine, arginine, valine and alanine occurrence by the probability of codons (40). On the contrary, this probability will result in reduced phenylalanine, isoleucine, tyrosine, histidine and asparagine presence in these genes. For HIV-1 protease, the mutational bias is towards 'A' and 'C', which is also the case when 'A to G' hypermutations are not considered for Gag and both p66 genes. The accumulation of 'A' would lead to a bias towards threonine and lysine, isoleucine, asparagine, arginine and stop codons, while reducing phenylalanine, tryptophan and cysteine occurrences. Similarly, the accumulation of 'C' would lead to proline, serine, leucine, threonine, alanine and arginine while reducing methionine, lysine, glutamic acid, tryptophan, and more importantly, stop codons. Analysing the 'A to G' hypermutations, signatures for ADAR editing were found at the sites, with the generated two sample logo motif demonstrating varying degrees of agreement, and with multiple mutation variants (hypothesized to be due to ADAR effects) agreeing to a high degree.

There was a tendency for Gag and RT p66 wildtype to have less destabilization effects in $\Delta\Delta G_{Mt}$ due to 'A to G' transitions compared to other mutations. Interestingly, Gag seemed more robust compared to p66 with its $\Delta\Delta G_{Mt}$ being generally neutral (Rosetta showed distributions of significantly stabilizing mutations that could have acted as epistatic buffers against destabilizing ones), although this may require confirmation by the study of cleaved Gag models (55). On the other hand, p66 codon mutated appeared to be less robust to changes in

‘A to G’ compared to p66 wildtype. This suggests that proteins of RNA viruses do not just minimize $\Delta\Delta G_{Mt}$ at the structural level as suggested by Tawfik, Berezovsky and colleagues (29), but also occurred at the sequence level. With respects to ‘A to G’ multiple substitution mutations as generated by ADAR hyper-editing, we see that the average $\Delta\Delta G_{Mt}$ of all p66 mutations (and also Gag) were half that of the p66 codon mutated gene, with lower rates of ‘A to G’ transition in the wildtype p66 compared to its codon mutated counterpart. Such observations suggest the natural wildtype p66 (and possibly Gag) gene had codon usage to be to be robust to ADAR effects.

Upon mapping the mutations observed, we did not find mutations in the majority of the active sites of the enzymes (protease and RT p66) nor in the cleavage sites of Gag, with the exception of A431D and F433L in NC/p1 cleavage site ^{429}RQAN/FLG^{435} of Gag, suggesting that cleavage compensation mutations coud have emerged early in the infection process even in the absence of drug seletion. Despite mimiking a single replication cycle, our assay also detected known clinical drug resistance mutations (see Table S5 for Gag, protease, and p66). For instance, in HIV-1 Gag, the observed E17K mutation (in the matrix) was previously reported as a rare transient mutation implicated in cytotoxic T lymphocyte (CTL) immune evasion resistance (56, 57). In protease, the observed K70T, previously reported to be a minority mutation associated with resistance to PIs (58) was also found in our assay. In RT p66, a known polymorphic mutation K103R (in the palm domain) was observed, which when in combination with V179D (not found in our assay), reduced susceptibility to non-nucleoside reverse transcriptase inhibitors (NNRTIs): nevirapine (NVP) and efavirenz (EFV) by about 15 folds (59). Interestingly the observed K103R originated from an ‘A to G’ hypermutation event, specifically from “AAA” to “AGA” suggesting that drug resistance mutations can be induced by host cell ADAR.

Mutations were also found in crucial functional sites within the genes, albeit in lower frequency. In HIV-1 Gag, we found P222L in the cyclophilin A (CyPA) binding site (along with G221) that might potentially affect binding of the capsid to CyPA as previously reported for P222A (60). Given that the probability of P (CCT) to L (CTT) is 0.33 in having a T substitution in the 2nd codon position, and that P (CCT) to A (GCT) is also 0.33 in having the first C mutated to a G, there were equal probabilities. However, when considering that our Gag mutation analysis did not show a transition mutation link between C to G (shown in Fig. 1), it is likely that the clinical mutation is a rare-occurrence (under additional selection pressures like drug) or a two or more mutation steps. Mutations A431D in the nucleocapsid and P453T in the p6 domain were also observed, with reported clinical drug resistance counterparts as A431V and a L449F/P453T pair (61). In HIV-1 RT p66, the substitution F61S was observed, in which the F61 residue was previously reported to play a vital role in strand synthesis, with F61Y/L/W altering activity (62). Similarly, we observed P95L, in which P95 was reported to be at the dimerization interface for the formation of the bottom of NNRTI pocket (63), and proposed to be a target amino acid in NNRTI design together with N137 and P140 (64). Although P95 was reported to be in a highly conserved location (63), it was observed in our assay with a limited sample size.

The occurrence of previously reported *in vitro* and clinical mutations in our assay suggest that these mutations occurred without immune surveillance and drug selection pressure (such as ART), occurring by probability within HIV gene specific biases of RT (Fig. 1). The general avoidance of inducing mutations at key regions/sites even at the nucleotide stage and tendency for the natural p66 nucleotide codon usage (compared to p66 codon mutated) to minimize changes in protein $\Delta\Delta G_{Mt}$, seems to suggest in-built barriers to detrimental mutations. Most relevant to the clinic is that these mutations occurred within our mimic of a single cycle of replication. Considering that HIV-1 generates approximately 10^9 virions per day

in an infected individual (65) which is magnitudes higher than our hundreds of sequence variants per gene, there is reason to believe that drug resistance isolate can occur even within the first day of infection.

In the absence of selection pressures, mutations that are selected against can be better detected, allowing for a more comprehensive analysis of the HIV-1 RT bias. Using current methods, including the high-throughput ones like next-generation sequencing, low frequency mutations can be missed to only emerge during intense selection. Given that many HIV proteins can function in intense drug/immune selection environments with significant reduced activity (66, 67), an all-out campaign against HIV could involve inhibitors against low-frequency occurring drug resistance mutations. This pushes the HIV infection towards clearance by Muller's ratchet (68-70) or by increasing viral mutation rate beyond pass a critical threshold towards lethal mutagenesis, leading to the loss in genetic information and viability, error catastrophe (71, 72). Since HIV-1 replicates at the edge of this critical threshold (73), a slight increase in the mutation rate may be sufficient to bring about error catastrophe, which have been successfully induced by mutagenic nucleoside analogues (74-76).

Coupling with *in silico* tools, such as those incorporating probability mutational change, it may be possible to identify mutational hotspots. As it was previously shown that protease and RT drug cross resistance have a structural basis governed by drug resistance mutations (77, 78), the bias of restricting specific amino acid changes by the absence of A-T, C-G mutation occurrences can be exploited to constrain the emergence of drug resistance. However, such an approach will require an in-depth understanding of HIV-RT mutations that do not proceed to a protein level under the most basic biological selection pressures of protein function.

Through the understanding of the mutational biases and mutation rate of HIV-1 RT, it is possible to calculate the mutational events leading to the speciation jump from an early SIV

to the current HIV, making it important for surveillance for emerging viral threats (2). Although species jump is unpredictable, RNA viruses are most likely to species jump after undergoing genetic changes (1).

In conclusion, we have established an assay to generate and characterise novel HIV-1 RT mutations on HIV-1 Gag, protease and RT p66 in a safe BSL2 manner, allowing insights to the mutational bias and mutation rate of HIV in the absence of biological selection pressures. In our analysis, we also characterized the likely effect of host factors e.g. ADAR which increased the genetic diversity beyond the already error-prone RT. In this genetic diversity, we showed that we were able to generate known clinical drug resistance mutations while mimicking a single replication cycle, showing that these mutations would be present very early in infection. The assay can provide deeper insights relevant for drug and vaccine development and be applied for horizontal understanding to other viruses.

Methods

Transfection of HIV-1 Gag, PR and RT plasmids

HIV-1 Gag and PR genes were PCR amplified from plasmid p8.91 (79). HIV-1 RT p66 (GenBank: K03455.1) and a codon mutated version of RT p66 was gene synthesized (BioBasic). Codon mutated p66 was generated by reverse translation from the amino acid sequence that would give rise to silent mutations with differences at the nucleotide sequence level (sequence available in Table S4) to potentially disrupt RNA secondary structures. The genes were cloned separately into pTT5 plasmid vector (YouBio), transformed into in-house competent DH5α *E. coli* cells (80), and transfected into HEK293F cells as previously performed (81-83).

RNA extraction and cDNA synthesis

Total RNA was extracted using TRIzol reagents according to manufacturer's instructions (Invitrogen) from transfected cells two days after transfection with the gene of interest. cDNA synthesis was performed using commercial HIV-1 RT subunits: (i) p51 (0.2475 µg) and p66 (0.2125 µg) from Sino Biological Inc (catalogue: 40244-V07E and 40244-V07E1, respectively), (ii) DNase-treated RNA, (iii) 5X RT buffer (25 mM Dithiothreitol, 375 mM KCl, 15 mM $MgCl_2$, 250 mM Tris-HCl [pH 8.3]), (iv) Oligo$(dT)_{18}$ (Thermo Scientific), (v) dNTPs, and (vi) RiboLock RNase inhibitor (ThermoFisher Scientific), in a single cycle of 25°C for 18 min, 37°C for 1 hr and 85°C for 5 min. RT negative controls were prepared without the addition of HIV-1 RT p51 and p66 subunits.

Amplification of cDNA and TOPO cloning

High-fidelity Q5 Polymerase PCR (New England Biolabs) was used to amplify cDNA templates using the following primers for the respective genes: HIV-1 Gag - forward (5 '– TAT TAG GAA TTC ATG GGT GCG AGA GCG– 3') and reverse (5' – CTG GTA AAG CTT CTA GTG GTG GTG GTG – 3'); for protease - forward (5' – GCG GCC GAA TTC ATG CCT CAA ATC AC – 3') and reverse (5' – TAT AAT AAG CTT CTA GTG GTG GTG GTG –3'); for RT p66 - forward (5' – ATG GCC TTG ACC TTT GCT TTA CTG – 3') and reverse (5'– CTT GTC GTC ATC GTC TTT GTA GTC – 3'); for codon mutated p66 - forward (5'– GCG GTG ATG GAT GGA CCA AAA GTA AA – 3') and reverse (5'– CTG CGC CTA ATG ATG ATG ATG AT – 3') as per manufacturer's recommendations. PCR products were analysed by gel electrophoresis using GelApp (84) and purified using Gel extraction and PCR purification kits as previously described (85). Purified Q5 Polymerase PCR products were cloned using Zero Blunt TOPO PCR cloning kit (Invitrogen) as per manufacturer's protocol and transformed into in-house competent DH5α cells as previously described (80). Transformed DH5α cells were plated and grown overnight at 37°C on LB agar plates

supplemented with kanamycin (50 μg/ml). Transformants were screened using GoTaq PCR (Promega) with universal M13F (-20) forward and reverse primers prior to sequencing.

Sequence analysis

Sequence assembly and alignment of HIV-1 sequences were performed using YAQAAT Webserver (yaqaat.apdskeg.com). DNA2App (86) was used to translate nucleotide sequences to their corresponding amino acid sequences. Previously characterised HIV-1 sequences were retrieved from the Los Alamos sequence database (87). Mutations were identified through multiple sequence alignments with the transfected sequence. To rule out sequencing artefacts, sequence chromatograms were analysed, and ambiguous peaks and mutated sequences were sent for re-sequencing. Mutation rates were calculated as mutations/bp where the total number of mutations and the total number of nucleotide bases of the respective HIV-1 genes were compared.

A Two Sample logo was generated to determine the underlying sequence contexts of 'A to G' transitions generated in our in-vitro assay for comparisons against previously reported sequence context preferences of ADAR (28) . A custom script was written to locate, pool and align windows of 4 nt upstream and downstream of either all the adenosine mutants; or all naturally adenosines which would be targets of host ADAR deaminases to generate an ersatz set of sequences as the background control in the Gag, Protease, wild-type p66 and p66 codon mutated sequences (see Fig. 4a for schematic representation). The mutant sequences were tested against the respective ersatz background sequences using a locally installed version of the Two Sample logo software (88) with two sample t-test without Bonferroni correction to test for significantly enriched and depleted bases within the 9 nt window.

In-silico assessment on protein thermostability using FoldX and Rosetta Cartesian_ddg

Mutagenesis using PyMOL was performed on previously modelled compact Gag structures (13) to match the Gag sequence used and were minimized using GROMOS96 implementation in Swiss-PdbViewer. For p66, the 3T19 crystal structure was used.

To predict and model the effects of the mutations on protein thermostability, free energy changes of the mutations $\Delta\Delta G_{mt}$ ($\Delta G_{wildtype}$ - ΔG_{mutant}) were modelled using two software approaches: FoldX5 and Rosetta Cartesian_ddg (Version : 2017.52.58848) (31). Structures were first relaxed and minimized using either the FoldX RepairPDB protocol or with the cartesian-space refinement relax protocol and ref2015_cart score function prior to mutagenesis and free energy modelling with the Cartesian_ddg protocol (As described in https://www.rosettacommons.org/docs/latest/cartesian-ddG, with modifications). To calculate mutational effects on protein thermostability with FoldX, BuildModel (using default settings) was run with the numberOfRuns parameter set to 10. For all Rosetta procedures, the lowest scoring structure out of 1000 cartesian-space minimization iterations were assessed to be free of serious clashes by the MolProbity server (89). The Cartesian_ddg protocol and talaris_cart score function were used to generate 15 models for each mutation and the average of the lowest 3 scoring mutations were averaged and multiplied by the scaling factor (31) to calculate the $\Delta\Delta G_{mt}$ in kcal/mols.

Competing Interests

There are no conflicts of interest

Acknowledgments

We thank Chinh Tran-To Su, Wai-Heng Lua and Wei-Li Ling for useful comments and discussion.

Funding

This work was supported by the A*STAR Industry Alignment Fund (IAF) grant IAF111149 and by EDDC, A*STAR.

Author Contributions

JYY, PY, and GRG performed the *in vitro* experiments. JYY, PY, GRG, and DWSK analysed the results. JYY, PY, GRG, DWSK, and SKEG discussed and drafted the manuscript. SKEG conceived the idea and supervised the whole project. All authors read and approved the manuscript.

References

1. Woolhouse MEJ, Haydon DT, Antia R. 2005. Emerging pathogens: the epidemiology and evolution of species jumps. Trends in Ecology & Evolution 20:238-244.
2. Flanagan ML, Parrish CR, Cobey S, Glass GE, Bush RM, Leighton TJ. 2012. Anticipating the Species Jump: Surveillance for Emerging Viral Threats. Zoonoses and Public Health 59:155-163.
3. Wertheim JO, Worobey M. 2009. Dating the Age of the SIV Lineages That Gave Rise to HIV-1 and HIV-2. PLOS Computational Biology 5:e1000377.
4. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD. 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 391:594-597.
5. Freed EO. 2015. HIV-1 assembly, release and maturation. Nature Reviews 13:484-496.

6. Pettit SC, Everitt LE, Choudhury S, Dunn BM, Kaplan AH. 2004. Initial Cleavage of the Human Immunodeficiency Virus Type 1 GagPol Precursor by Its Activated Protease Occurs by an Intramolecular Mechanism. Journal of Virology 78:8477-8485.
7. Fun A, van Maarseveen NM, Pokorná J, Maas REM, Schipper PJ, Konvalinka J, Nijhuis M. 2011. HIV-1 protease inhibitor mutations affect the development of HIV-1 resistance to the maturation inhibitor bevirimat. Retrovirology 8:70.
8. Fun A, Wensing AMJ, Verheyen J, Nijhuis M. 2012. Human Immunodeficiency Virus gag and protease: partners in resistance. Retrovirology 9:63.
9. Clavel F, Mammano F. 2010. Role of Gag in HIV Resistance to Protease Inhibitors. Viruses 2:1411-1426.
10. Wensing AM, Calvez V, Günthard HF, Johnson VA, Paredes R, Pillay D, Shafer RW, Richman DD. 2017. 2017 Update of the Drug Resistance Mutations in HIV-1. Topics in Antiviral Medicine 24.
11. Park J, Sayer J, Aniana A, Yu X, Weber I, Harrison R, Louis J. 2016. Binding of clinical inhibitors to a model precursor of a rationally selected multidrug resistant HIV-1 protease is significantly weaker than that to the released mature enzyme. Biochemistry 55:2390-2400.
12. Yu Y, Wang J, Shao Q, Shi J, Zhu W. 2015. Effects of drug-resistant mutations on the dynamic properties of HIV-1 protease and inhibition by Amprenavir and Darunavir. Sci Rep 5:10517.
13. Su CT-T, Kwoh C-K, Verma CS, Gan SK-E. 2017. Modeling the full length HIV-1 Gag polyprotein reveals the role of its p6 subunit in viral maturation and the effect of non-cleavage site mutations in protease drug resistance. Journal of Biomolecular Structure and Dynamics doi:10.1080/07391102.2017.1417160:1-12.
14. Clutter DS, Jordan MR, Bertagnolio S, Shafer RW. 2016. HIV-1 drug resistance and resistance testing. Infection, Genetics and Evolution 46:292-307.
15. Gupta RK, Kohli A, McCormick AL, Towers GJ, Pillay D, Parry CM. 2010. Full-length HIV-1 Gag determines protease inhibitor susceptibility within in-vitro assays. AIDS 24:1651-1655.
16. Dam E, Quercia R, Glass B, Descamps D, Launay O, Duval X, Kräusslich H-G, Hance AJ, Clavel F, Group AS. 2009. Gag Mutations Strongly Contribute to HIV-1 Resistance to Protease Inhibitors in Highly Drug-Experienced Patients besides Compensating for Fitness Loss. PLoS Pathogens 5:e1000345.
17. Parry CM, Kohli A, Boinett CJ, Towers GJ, McCormick AL, Pillay D. 2009. Gag Determinants of Fitness and Drug Susceptibility in Protease Inhibitor-Resistant Human Immunodeficiency Virus Type 1. Journal of Virology 83:9094-9101.
18. Dam E, Quercia R, Glass B, Descamps D, Launay O, Duval X, Krausslich H-G, Hance AJ, Clavel F, Group AS. 2009. Gag mutations strongly contribute to HIV-1 resistance to Protease Inhibitors in highly drug-experienced patients besides compensating for fitness loss. Plos Pathogen 5:e1000345.
19. Tamiya S, Mardy S, Kavlick MF, Yoshimura K, Mitsuya H. 2004. Amino acid insertions near Gag cleavage sites restore the otherwise compromised replication of human immunodeficiency virus type 1 variants resistant to Protease Inhibitors. J Virol 78:12030-12040.
20. Maguire MF, Guinea R, Griffin P, Macmanus S, Elston RC, Wolfram J, Richards N, Hanlon MH, Porter DJ, Wrin T, Parkin N, Tisdale M, Furfine E, Petropoulos C, Snowden BW, Kleim J-P. 2002. Changes in Human Immunodeficiency Virus Type 1 Gag at Positions L449 and P453 Are Linked to I50V Protease Mutants In Vivo and Cause Reduction of Sensitivity to Amprenavir and Improved Viral Fitness In Vitro. Journal of Virology 76:7398-7406.

21. Gatanaga H, Suzuki Y, Tsang H, Yoshimura K, Kavlick MK, Nagashima K, Gorelick RJ, Mardy S, Tang C, Summers MF, Mitsuya H. 2002. Amino Acid Substitutions in Gag Protein at Non-cleavage sites Are Indispensable for the Development of a high Multitude of HIV-1 Resistance against Protease Inhibitors. The journal of Biological Chemistry 277:5952-5961.
22. Lightfoote MM, Coligan JE, Folks TM, Fauci AS, Martin MA, Venkatesan S. 1986. Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus. Journal of virology 60:771-775.
23. Hostomsky Z, Hostomska Z, Fu TB, Taylor J. 1992. Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis. J Virol 66:3179-82.
24. Sarafianos SG, Marchand B, Das K, Himmel DM, Parniak MA, Hughes SH, Arnold E. 2009. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. Journal of molecular biology 385:693-713.
25. Le Grice SF, Naas T, Wohlgensinger B, Schatz O. 1991. Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase. The EMBO journal 10:3905-3911.
26. Yeo JY, Goh GR, Su CT-T, Gan SK-E. 2020. The Determination of HIV-1 RT Mutation Rate, Its Possible Allosteric Effects, and Its Implications on Drug Resistance. Preprints doi:doi: 10.20944/preprints202002.0099.v1:2020020099.
27. Menéndez-Arias L. 2009. Mutation Rates and Intrinsic Fidelity of Retroviral Reverse Transcriptases. Viruses 1:1137-1165.
28. Eggington JM, Greene T, Bass BL. 2011. Predicting sites of ADAR editing in double-stranded RNA. Nature Communications 2:319.
29. Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS. 2009. Do viral proteins possess unique biophysical features? Trends in Biochemical Sciences 34:53-59.
30. Strokach A, Corbi-Verge C, Kim PM. 2019. Predicting changes in protein stability caused by mutation using sequence-and structure-based methods in a CAGI5 blind challenge. Human Mutation 40:1414-1423.
31. Park H, Bradley P, Greisen P, Jr., Liu Y, Mulligan VK, Kim DE, Baker D, DiMaio F. 2016. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. Journal of chemical theory and computation 12:6201-6212.
32. Delgado J, Radusky LG, Cianferoni D, Serrano L. 2019. FoldX 5.0: working with RNA, small molecules and a new graphical interface. Bioinformatics 35:4168-4169.
33. Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics 148:1667-1686.
34. Potapov V, Ong JL. 2017. Examining sources of error in PCR by single-molecule sequencing. PloS One 12:e0169774.
35. Geller R, Domingo-Calap P, Cuevas JM, Rossolillo P, Negroni M, Sanjuán R. 2015. The external domains of the HIV-1 envelope are a mutational cold spot. Nature Communications 6:8571.
36. Ji J, Loeb LA. 1994. Fidelity of HIV-1 Reverse Transcriptase Copying a Hypervariable Region of the HIV-1 env Gene. Virology 199:323-330.
37. Lyons DM, Lauring AS. 2017. Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses. Molecular Biology and Evolution 34:3205-3215.

38. Stoltzfus A, Norris RW. 2016. On the Causes of Evolutionary Transition:Transversion Bias. Molecular Biology and Evolution 33:595-602.
39. Rosenberg MS. 2003. Patterns of Transitional Mutation Biases Within and Among Mammalian Genomes. Molecular Biology and Evolution 20:988-993.
40. Chan K-F, Koukouravas S, Yeo JY, Koh DW-S, Gan SK-E. 2020. Probability of change in life: Amino acid changes in single nucleotide substitutions. Biosystems 193-194:104135.
41. Tian N, Yang Y, Sachsenmaier N, Muggenhumer D, Bi J, Waldsich C, Jantsch MF, Jin Y. 2011. A structural determinant required for RNA editing. Nucleic Acids Research 39:5669-5681.
42. Dawson TR, Sansam CL, Emeson RB. 2004. Structure and Sequence Determinants Required for the RNA Editing of ADAR2 Substrates. Journal of Biological Chemistry 279:4941-4951.
43. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. 2010. Nature, Position, and Frequency of Mutations Made in a Single Cycle of HIV-1 Replication. Journal of Virology 84:9864-9878.
44. Bass BL, Weintraub H. 1988. An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell 55:1089-1098.
45. Nishikura K. 2010. Functions and Regulation of RNA Editing by ADAR Deaminases. Annual Review of Biochemistry 79:321-349.
46. Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG. 2020. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. Science Advances 6:eabb5813.
47. Piontkivska H, Frederick M, Miyamoto MM, Wayne ML. 2017. RNA editing by the host ADAR system affects the molecular evolution of the Zika virus. Ecology and Evolution 7:4475-4485.
48. Piontkivska H, Matos LF, Paul S, Scharfenberg B, Farmerie WG, Miyamoto MM, Wayne ML. 2016. Role of Host-Driven Mutagenesis in Determining Genome Evolution of Sigma Virus (DMelSV; Rhabdoviridae) in Drosophila melanogaster. Genome Biology and Evolution 8:2952-2963.
49. Clerzius G, Gélinas J-F, Daher A, Bonnet M, Meurs EF, Gatignol A. 2009. ADAR1 Interacts with PKR during Human Immunodeficiency Virus Infection of Lymphocytes and Contributes to Viral Replication. Journal of Virology 83:10119-10128.
50. Doria M, Neri F, Gallo A, Farace MG, Michienzi A. 2009. Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. Nucleic Acids Research 37:5848-5858.
51. Gélinas J-F, Clerzius G, Shaw E, Gatignol A. 2011. Enhancement of Replication of RNA Viruses by ADAR1 via RNA Editing and Inhibition of RNA-Activated Protein Kinase. Journal of Virology 85:8460-8466.
52. Weiden MD, Hoshino S, Levy DN, Li Y, Kumar R, Burke SA, Dawson R, Hioe CE, Borkowsky W, Rom WN, Hoshino Y. 2014. Adenosine Deaminase Acting on RNA-1 (ADAR1) Inhibits HIV-1 Replication in Human Alveolar Macrophages. PLOS ONE 9:e108476.
53. Lathouwers E, Wong EY, Luo D, Seyedkazemi S, De Meyer S, Brown K. 2017. HIV-1 resistance rarely observed in subjects using darunavir once-daily regimens across clinical studies. HIV Clinical Trials 18:196-204.
54. Titanji BK, Pillay D, Jolly C. 2017. Combination antiretroviral therapy and cell–cell spread of wild-type and drug-resistant human immunodeficiency virus-1. Journal of General Virology 98:821-834.

55. Samsudin F, Gan SK-E, Bond PJ. 2020. The Structural Basis for Gag Non-Cleavage Site Mutations in Determining HIV-1 Viral Fitness. bioRxiv doi:10.1101/2020.07.05.188326:2020.07.05.188326.
56. Sanchez-Merino V, Farrow MA, Brewster F, Somasundaran M, Luzuriaga K. 2008. Identification and Characterization of HIV-1 CD8+ T Cell Escape Variants with Impaired Fitness. The Journal of Infectious Diseases 197:300-308.
57. Sanchez-Merino V, Nie S, Luzuriaga K. 2005. HIV-1-Specific CD8+ T Cell Responses and Viral Evolution in Women and Infants. The Journal of Immunology 175:6976-6986.
58. Mohamed S, Penaranda G, Gonzalez D, Camus C, Khiri H, Boulmé R, Sayada C, Philibert P, Olive D, Halfon P. 2014. Comparison of ultra-deep versus Sanger sequencing detection of minority mutations on the HIV-1 drug resistance interpretations after virological failure. AIDS 28:1315-1324.
59. Parkin NT, Gupta S, Chappey C, Petropoulos CJ. 2006. The K101P and K103R/V179D Mutations in Human Immunodeficiency Virus Type 1 Reverse Transcriptase Confer Resistance to Nonnucleoside Reverse Transcriptase Inhibitors. Antimicrobial Agents and Chemotherapy 50:351-354.
60. De Iaco A, Luban J. 2014. Cyclophilin A promotes HIV-1 reverse transcription but its effect on transduction correlates best with its effect on nuclear entry of viral cDNA. Retrovirology 11:11.
61. Banke S, Lillemark MR, Gerstoft J, Obel N, Jørgensen LB. 2009. Positive Selection Pressure Introduces Secondary Mutations at Gag Cleavage Sites in Human Immunodeficiency Virus Type 1 Harboring Major Protease Resistance Mutations. Journal of Virology 83:8916-8924.
62. Fisher TS, Darden T, Prasad VR. 2003. Substitutions at Phe61 in the β3-β4 Hairpin of HIV-1 Reverse Transcriptase Reveal a Role for the Fingers Subdomain in Strand Displacement DNA Synthesis. Journal of Molecular Biology 325:443-459.
63. Ceccherini-Silberstein F, Gago F, Santoro M, Gori C, Svicher V, Rodríguez-Barrios F, d'Arrigo R, Ciccozzi M, Bertoli A, Monforte AdA, Balzarini J, Antinori A, Perno C-F. 2005. High Sequence Conservation of Human Immunodeficiency Virus Type 1 Reverse Transcriptase under Drug Pressure despite the Continuous Appearance of Mutations. Journal of Virology 79:10718-10729.
64. Auwerx J, Van Nieuwenhove J, Rodríguez-Barrios F, de Castro S, Velázquez S, Ceccherini-Silberstein F, De Clercq E, Camarasa M-J, Perno C-F, Gago F, Balzarini J. 2005. The N137 and P140 amino acids in the p51 and the P95 amino acid in the p66 subunit of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase are instrumental to maintain catalytic activity and to design new classes of anti-HIV-1 drugs. FEBS Letters 579:2294-2300.
65. Santoro MM, Perno CF. 2013. HIV-1 Genetic Variability and Clinical Implications. ISRN microbiology 2013:481314-481314.
66. Chung S, Miller JT, Lapkouski M, Tian L, Yang W, Le Grice SF. 2013. Examining the role of the HIV-1 reverse transcriptase p51 subunit in positioning and hydrolysis of RNA/DNA hybrids. Journal of Biological Chemistry 288:16177-16184.
67. Betancor G, Alvarez M, Marcelli B, Andrés C, Martínez MA, Menéndez-Arias L. 2015. Effects of HIV-1 reverse transcriptase connection subdomain mutations on polypurine tract removal and initiation of (+)-strand DNA synthesis. Nucleic Acids Research 43:2259-2270.
68. Hu W-S, Temin HM. 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proceedings of the National Academy of Sciences 87:1556-1560.

69. Yuste E, Sánchez-Palomino S, Casado C, Domingo E, López-Galíndez C. 1999. Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. Journal of Virology 73:2745-2751.
70. Ormond L, Liu P, Matuszewski S, Renzette N, Bank C, Zeldovich K, Bolon DN, Kowalik TF, Finberg RW, Jensen JD. 2017. The combined effect of oseltamivir and favipiravir on influenza A virus evolution. Genome biology and evolution 9:1913-1924.
71. Summers J, Litwin S. 2006. Examining The Theory of Error Catastrophe. Journal of Virology 80:20-26.
72. Bull JJ, Sanjuán R, Wilke CO. 2007. Theory of Lethal Mutagenesis for Viruses. Journal of Virology 81:2930-2939.
73. Drake JW, Holland JJ. 1999. Mutation rates among RNA viruses. Proceedings of the National Academy of Sciences 96:13910-13913.
74. Dapp MJ, Clouser CL, Patterson S, Mansky LM. 2009. 5-Azacytidine Can Induce Lethal Mutagenesis in Human Immunodeficiency Virus Type 1. Journal of Virology 83:11950-11958.
75. Harris KS, Brabant W, Styrchak S, Gall A, Daifuku R. 2005. KP-1212/1461, a nucleoside designed for the treatment of HIV by viral mutagenesis. Antiviral Research 67:1-9.
76. Loeb LA, Essigmann JM, Kazazi F, Zhang J, Rose KD, Mullins JI. 1999. Lethal mutagenesis of HIV with mutagenic nucleoside analogs. Proceedings of the National Academy of Sciences 96:1492-1497.
77. Chiang Ron Z-H, Gan Samuel K-E, Su Chinh T-T. 2018. A computational study for rational HIV-1 non-nucleoside reverse transcriptase inhibitor selection and the discovery of novel allosteric pockets for inhibitor design. Bioscience Reports 38:BSR20171113.
78. Su CT-T, Ling W-L, Lua W-H, Haw Y-X, Gan SK-E. 2016. Structural analyses of 2015-updated drug-resistant mutations in HIV-1 protease: an implication of protease inhibitor cross-resistance. BMC Bioinformatics 17:500.
79. Naldini L, Blömer U, Gallay P, Ory D, Mulligan R, Gage FH, Verma IM, Trono D. 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science 272:263-267.
80. Chan WT, Verma Chandra S, Lane David P, Gan Samuel KE. 2013. A comparison and optimization of methods and factors affecting the transformation of Escherichia coli. Bioscience Reports 33:931-937.
81. Ling W-L, Lua W-H, Poh J-J, Yeo JY, Lane DP, Gan SK-E. 2018. Effect of VH–VL Families in Pertuzumab and Trastuzumab Recombinant Production, Her2 and FcγIIA Binding. Frontiers in Immunology 9.
82. Lua W-H, Ling W-L, Yeo JY, Poh J-J, Lane DPS, Gan SK-E. 2018. The effects of Antibody Engineering CH and CL in Trastuzumab and Pertuzumab Recombinant Models: Impact on antibody production and antigen-binding. Scientific Reports 8:718.
83. Lua W-H, Su CT-T, Yeo JY, Poh J-J, Ling W-L, Phua S-X, Gan SK-E. 2019. Role of the IgE variable heavy chain in FcεRIα and superantigen binding in allergy and immunotherapy. Journal of Allergy and Clinical Immunology.
84. Sim J-Z, Nguyen P-V, Lee H-K, Gan SK-E. 2015. GelApp: Mobile gel electrophoresis analyser. Nature Methods Application Notes.
85. Poh JJ, Gan SKE. 2014. The Determination of Factors involved in Column-Based Nucleic Acid Extraction and Purification. Journal of Bioprocessing & Biotechniques 04.

86. Sim JZ, Nguyen PV, Zang Y, Gan SKE. 2016. DNA2App: Mobile sequence analyser. Scientific Phone Apps and Mobile Devices 2:2.
87. LosAlamos. HIV Sequence Database.
88. Vacic V, Iakoucheva LM, Radivojac P. 2006. Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. Bioinformatics 22:1536-1537.
89. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall III WB, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC. 2018. MolProbity: More and better reference data for improved all-atom structure validation. Protein Science 27:293-315.

Supplemental Legends

Table S1 HIV-1 Gag nucleotide and amino acid mutations. Free energy values $\Delta\Delta G_{Mt}$ (kcal / mol) were rounded to 2 significant figures.

Table S2 HIV-1 Protease nucleotide and amino acid mutations.

Table S3 HIV-1 RT p66 nucleotide and amino acid mutations. Free energy values $\Delta\Delta G_{Mt}$ (kcal / mol) were rounded to 2 significant figures.

Table S4 Codon mutated HIV-1 RT p66 nucleotide and amino acid mutations. Free energy values $\Delta\Delta G_{Mt}$ (kcal / mol) were rounded to 2 significant figures.

Table S5 Clinical, drug resistant and mutations with reported functions of HIV-1 Gag, Protease and RT p66.

Table S6 Change in mutational free energies ($\Delta\Delta G_{Mt}$, kcal/mol) of experimentally generated multiple mutation variants.

Fig S1 Two Sample logos of HIV-1 Gag, protease, RT p66 (p66 Wt) and p66 Codon Mutated RT (p66 CM) illustrating the underlying sequence contexts of all adenosine mutations. Bases were coloured opaque when $p < 0.01$, and translucent when $p < 0.05$.