Predicting mutations in HIV-1 Gag: Insights from *in silico* and an *in vitro* BSL2 platform on thermostability and allosteric effects

Ping Yap[a], Darius Wen-Shuo Koh[a], Chinh Tran-To Su[a], Kwok-Fong Chan[a] & Samuel Ken-En Gan[a,b,#]

[a]*Antibody & Product Development Lab, Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore 138671.*

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

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

Bioinformatics Institute, A*STAR

30 Biopolis Street, #07-01 Matrix

Singapore 138671

Tel: +65 64788301

Fax: +65 6478 9047

Predicting mutations in HIV-1 Gag: Insights from *in silico* and an *in vitro* BSL2 platform on thermostability and allosteric effects

HIV-1 Gag and protease are both involved in protease inhibitor resistance, where Gag mutations alone were reported to be able to compensate for protease inhibition. These resistance mutations arise as a result of the error-prone HIV-1 reverse transcriptase. To study if such mutations were incorporated randomly, we investigated the mutations generated in our HIV-RT cDNA synthesis PCR assay that is devoid of selection pressures involving protein fitness and immune surveillance. Studying a total of 269 one-generation sequences using HIV-1 RT, we calculated an error rate of about 8.7 x10^{-5}/bp with a bias towards transition mutations. We found previously reported mutations as well as unreported novel mutations in our system. Computational structural analysis showed that the novel mutations had varying effects on the thermostability of the capsid linker and the first Gag cleavage site. In this work, we showed that our platform can be used for mutation generation that could aid in the design of pre-emptive interventions, and give us an insight to the effects of emerging mutations in HIV-1 Gag.

Keywords: Mutation prediction, HIV-1 drug resistance, allostery, thermostability, resistance mutations.

Introduction

HIV has infected more than 70 million individuals and is responsible for a total of 35 million deaths worldwide [1]. To date, Antiretroviral therapy (ART) remains the predominant treatment method [2], yet it is hampered by emerging drug-resistance mutations introduced during the reverse transcription process by HIV-1 reverse-transcriptase (HIV-RT) [3,4]. One main drug class used in ART is protease inhibitors (PIs) [2,3]. PIs bind to the active site of the viral protease, inhibiting its interaction with the substrate Gag, and thereby viral maturation [5,6]. To overcome PIs, mutations on viral protease reduce affinity to PIs [6,7] and the accumulation of such resistant mutations can render PIs ineffective as well as confer cross-resistance between PIs [6]

limiting drug selection. Recently, Gag has been reported to confer PI resistance [4,6-8] in the absence of mutations on Protease. Gag is a polyprotein comprising of several domains – matrix (MA), capsid (CA), sp1, nucleocapsid (NC), sp2, and p6 (these domains are reviewed extensively in [9,10]) and are cleaved before viral assembly [9,10]. Expectedly, mutations at the cleavage sites could confer PI-resistance [4,6-8]. Nonetheless, most drug-resistance mutations in protease comes at a fitness cost that is phenotypically manifested as decreased affinity to Gag or reduced catalytic capabilities [7,11], and thus, PI-resistant variants typically exhibit slower replication when compared to wild-type HIV variants [5,8]. To overcome this, Gag mutations compensated for such fitness cost effects [6-8], and these mutations may arise on both Gag cleavage or non-cleavage sites [6-8], making the understanding and characterization of drug-resistance Gag mutations imperative for improving patient prognosis, drug-prescription and design [4].

The capsid domain in Gag is a potential drug target [3,12]. Several small molecules targeting the capsid domain (capsid inhibitors; Bevirimat, PF74, etc) have shown promise, but only one -Bevirimat- entered phase II clinical trials [13]. Bevirimat inhibited proteolytic cleavage *in-vitro*, but natural Gag polymorphisms in some patients led to significantly reduced drug efficacy [13].

On the whole, Gag has also shown great promise as a potential vaccine candidate as observed from HIV-1 patients with normal $CD4^+$ T-lymphocytes levels and below detectable viral loads without anti-retroviral therapy (known as HIV-1 elite supressors) [14-16]. These elite suppressors showed $CD8^+$ T-lymphocyte (CTL) responses that naturally suppress HIV-1 viral loads [14,15,17], thus HIV-1 disease progression is mediated by strong and broad CTL responses to HIV-1 Gag protein. Two main HLA molecules (HLA-B57 & HLA-B27) that target Gag epitopes were found to

be overrepresented in elite suppressors [14,15]. By incorporating Gag as a target, the outcomes in the HIV-1 RV144 vaccine trial (RV144; 26.4%) [18,19] were improved.

We developed an *in-vitro* selection-free assay to characterize HIV-1 RT mutations on HIV-1 Gag. The absence of selection pressures e.g. (anti-retroviral drugs and immune surveillance) allows for an unbiased generation both novel and known HIV-1 Gag mutations. Further leveraging on our previously modelled Gag structures, we studied the effects of the mutations on thermostability and allosteric communications.

Materials & Methods

Generation of HIV-1 Gag mRNA

The HIV Gag gene was PCR amplified from plasmid p8.91 [20] was cloned into pTT5 plasmid vector (YouBio) and transformed into in-house competent *E. coli* cells [21]. HIV-1 Gag mRNA were expressed using transfected HEK293 EXPI maintained in Dulbecco modified Eagle medium supplemented with 10% foetal bovine serum, penicillin and streptomycin. pTT5-Gag plasmids (10 µg) were extracted (BioBasic; #BS614) and transfected into HEK293 EXPI cells (4 x 10^5 cells/ml) as previously described [22-25]. Two days after transfection, total RNA was extracted using TRIzol reagents (Invitrogen; #15596018) and subjected to DNase treatment (Sigma Aldrich; #716728001) as per manufacturer's protocols.

cDNA Synthesis using HIV-1 RT

cDNA synthesis reactions were performed using commercial HIV-1 RT subunits: p51 (0.2475 µg) and p66 (0.2125 µg) from Sino Biological Inc (catalogue: 40244-V07E and 40244-V07E1, respectively), (ii) 3 µg DNase-treated Gag RNA, (iii) 4 µl 5X RT

buffer (25 mM Dithiothreitol, 375 mM KCl, 15 mM $MgCl_2$, 250 mM Tris-HCl [pH 8.3]), (iv) 100 µM Oligo$(dT)_{18}$ (Thermo Scientific; #SO131), (v) 10 mM dNTPs, (vi) 40 units RiboLock RNase inhibitor (ThermoFisher Scientific; #EO0381) topped up to 20 µl of DEPC treated H_2O. No-RT negative controls were prepared without the addition of HIV-1 p51 and p66 HIV-1 RT subunits. cDNA synthesis reactions were performed using an Applied Biosystems Proflex thermal-cycle. cDNA synthesis was performed using 1 cycle of 25°C for 18 min, 37°C for 1 hr and 85°C for 5 min. Gag cDNA was detected from the cDNA reactions using GoTaq PCR (Promega; #M7823) and HIV-1 Gag specific primers (F: 5'CTG GTA AAG CTT CTA GTG GTG GTG GTG -3' and R:'5- TAT TAG GAA TTC ATG GGT GCG AGA GCG-3'). Each GoTaq Master Mix PCR reaction had 300 ng of cDNA as per manufacturers protocol (Promega). HIV-1 Gag cDNA templates were amplified using 35 cycles of 95°C for 30 secs, 58°C for 30 secs and 72°C for 1 min 40 secs. PCR products were subjected to 1% agarose gel electrophoresis (100V, 60 min) and visualized alongside a 10k bp DNA ladder. Gel analysis was performed using GelApp [26]. The cDNA reactions that had detected HIV-1 Gag cDNA were then subjected to high-fidelity Q5 Polymerase PCR (NEB; #M0491L).

High-fidelity amplification of HIV-1 gag cDNA and TOPO Cloning

Q5 Polymerase PCR (50 µl) was performed using 75 ng of HIV-1 Gag cDNA with the Gag primers listed above and assembled as per manufacturers protocol. HIV-1 Gag cDNA templates were amplified using 38 cycles of 98°C for 10 secs, 61°C for 10 secs and 72°C for 1 min 10 secs. Q5 Polymerase PCR products were analysed as performed above and purified using Gel extraction and PCR purification kits as previously described [27]. Purified Q5 Polymerase PCR products were cloned using TOPO PCR blunt end cloning kit (Invitrogen; #K280020) as per manufacturer's protocol. TOPO

vectors were transformed into in-house competent DH5α cells previously described [21]. Transformed DH5α cells were plated and grown overnight at 37°C on LB agar plates supplemented with kanamycin (100 mg/ml). Transformants were screened using colony PCR prior to sequencing.

Colony PCR & Sequencing

Transformants from TOPO reactions were screened using colony PCR screening method. Each colony PCR reaction comprised of 5 µl 2X GoTaq master mix, 1 µM each of universal M13F(-20) forward and M13R(-20) reverse primers, 3 µl nuclease free water and 1 TOPO colony. Negative controls were prepared without the addition of any colonies. Colony PCR reactions were performed using 35 cycles of 95°C for 30 secs, 55°C for 15 secs and 72°C for 2 min 15 secs, and the reactions were analysed by gel electrophoresis using GelApp [26]. Colonies that were positive for Gag inserts were grown overnight at 37°C, 250 rpm in LB broth supplemented with kanamycin. The resulting plasmids were extracted and sequenced using universal M13F(-20) and M13R(-20) primers. Mutations were identified from multiple sequence alignment to the original Gag sequence used for transfection. To rule out sequencing artefacts, sequence chromatograms were analysed, and ambiguous peaks and mutated Gag sequences were re-sequenced for verification

Computational analysis

Sequence assembly and multiple sequence alignment (MSA) of HIV-1 Gag sequences were performed using MAFFT [28] and Clustal Omega [29], respectively. ExPASy translate tool [30] and DNA2App [31] were used to convert HIV-1 Gag nucleotide sequences to its corresponding amino acid sequence. Previously characterized HIV-1 sequences were retrieved from the Los Alamos sequence database [32]. Mutagenesis

using PyMOL (version 2.2; [33]) was performed on both compact and extended Gag structures (previously modelled by Su *et al.* [3]) to match the Gag sequence used. The Gag structures were minimized using the GROMOS96 implementation of Swiss-PdbViewer (version 4.1.0; [34,35]) Possible effects of the mutations on Gag protein dynamics and thermostability were investigated using Elastic Network Contact Model (ENCoM) [36,37], FoldX4 [38] and MODELLER v9.17 [39,40] on an Ubuntu 16.04 machine.

Allosteric effects of mutations (Figure 1) were studied using the AlloSigMA server [41] to quantify the distal effect of the mutations on the first Gag cleavage site (ΔG_{site}) and the capsid linker (ΔG_{linker}). We used "Up mutation" (for residue positions E17, G192, S241, H124, K290, K418, K202, & S462) and "Down mutation" (for residue position N271, F433, F484, T239, T204, R214, I223) to simulate the residues that mutated accordingly to larger or smaller residues, respectively. The resulting residue-wise allosteric free energies (Δg, with negative indicating stabilizing and positive indicating destabilizing) due to the mutation effects were used to estimate the ΔG_{site} and ΔG_{linker} by averaging Δg of the involved residues.

Results

Characterization of HIV-1 Gag mutant variants

Translated Gag sequences showed that non-synonymous nucleotide mutations (n = 21) occurred at a higher frequency than synonymous mutations (n = 11, see Figure 1A-B). Nonsense mutations (n = 3) resulting in truncated Gag proteins were also found but excluded from further analysis. We were successful in characterizing Gag variants that possessed single (variants 1 to 6) and multiple (variants 7 to 9) missense mutations. With reference to the Los Alamos HIV sequence and mutation database [32,42], we

found that seven of our mutations were previously unreported alongside eight known mutations (underlined in Figure 1C). Some of these known mutations were previously reported to contribute to drug and immune evasion resistance (Table 1).

Further analysis did not show any particular mutational hotspots in our study. Gag variants 4, 9 and 11 were repeatedly found in their respective cDNA synthesis reactions and were not identified in other reactions. This is a likely result of having multiple cDNA copies. The multiple sequence alignment of the 12 unique Gag variants shown in Supplementary Figure S1 and S2, indicated that non-cleavage site mutations (n = 19) occurred at a higher frequency than those at the Gag cleavage sites (n = 2; F433L, at the fourth cleavage site [7]). We did not find mutations on the first Gag cleavage site (^{374}ATIMIQK^{380}) and there were more missense mutations (13/21) within the capsid domain, which was statistically expected for the largest domain in the Gag polyprotein.

Thermostability of Gag extended and compact structure

Gag protein was first formed as a compact structure and extends during the viral maturation process [43]. We used the extended and compact Gag structures [3] to investigate the changes in thermostability and protein dynamics caused by the mutations between the two Gag extreme states using ENCoM [36,37]. We excluded the truncated Gag variants (variants 10 and 11) in this analysis due to possible biases caused by the distant regions.

It was shown that when extended, the whole Gag structure generally became more thermostable (the effect is particularly pronounced and most varying on the capsid and p6) in the presence of the mutations. These trends were also observed when Gag was in the compact states.

The gain in thermostability at the p6 domain on the extended Gag was found for all the single mutation variants (except for variant 4, shown in Figure 2B and Supplementary Figure S3). On the other hand, the mutation F433L in variant 4 caused the loss of thermostability at the p6 domain of Gag. In variants 1 to 6, p6 was less thermostable in the compact Gag.

Among the variants with multiple mutations, variants 7 (H124R/K290R) and 8 (T239A/K418R) become less thermostable whereas variant 9 became more thermostable (Figure 2B). However, we observed varying effect on the four major domains matrix, capsid, nucleocapsid, and p6 regarding to different combinations of mutations in these variants.

To further characterize the possible compensatory or cumulative effects on both extended and compact Gag structures, we performed leave-one-out analysis (where we systematically reverted mutations one by one back to wild-type in the multiple mutation variants) for variants 7, 8, and 9 (with multiple mutations).

In variant 7 (H124R/K290R), the H124R mutation alone was responsible for rendering the matrix, N-terminal capsid, and p6 domains less thermostable but increased the thermostability of the C-terminal capsid (particularly at the linker ^{277}YSPTSIL^{283} that is involved in the core stability [44]), and nucleocapsid (Supplementary Figure S3B). The K290R alone increased thermostability, yet its effect was masked by the effects of H124R mutation in decreasing thermostability.

For variant 8 (T239A/K418R), the decrease of thermostability (on capsid and nucleocapsid) were elicited by the T239A mutation. However, the K418R mutation caused a decrease in thermostability of the p6 domain for the complete (in extended) or partial (in compact) structures (Supplementary Figure S3C).

In the case of compact variant 9 (K202R/T204A/R214G/I223V/S462R), the mutations decreased the thermostability of the nucleocapsid (Figure 2C), which became more thermostable when Gag extended (Figure 2B). This suggests a complementary neutralizing effect when Gag extends. Given that these mutations occurred at the capsid region (Figure 2D), they may be an indirect communication (allosteric effects) between these two regions in the various Gag conformational states.

Gag mutations and their allosteric effects

The thermostability profiles of both extended and compact Gag revealed distal effects between the various Gag domains. All the mutations (except for T239A in variant 8) resulted in the increased thermostability of both the capsid linker (^{277}YSPTSIL^{283}) and the first Gag cleavage site (^{374}ATIMIQK^{380}) in the extended Gag. The mutations T239A caused loss of thermostability, particularly at the cleavage site implicated in PI resistance [7,8]. On the other hand, some mutations amplified or neutralized the effects of the mutations when Gag extended. For example, the loss of thermostability by H124R (in variant 7) was only observed in extended Gag, while R214G and F484S effects on the nucleocapsid and p6 domains abated when Gag extended. Therefore, distal allosteric effects were caused by the mutations (either individually or accumulatively).

We used AlloSigMA server [41] to quantify the propagated signals of the mutations (Figure 3), especially onto the two regions of interest: the capsid linker (277YSPTSIL283) and the first Gag cleavage site (374ATIM/IQK380). We initiated perturbations using “Up mutation” and “Down mutation” to simulate mutations to bulkier or smaller amino acid residues, respectively (Table 3).

Only few mutations displayed allosteric effect on the capsid linker (ΔG_{capsid_linker}). Obvious differences in allosteric communications between the two Gag

conformational states were seen for the E17K (variant 1), G192R (variant 2), K418R (variant 8), K202R or S462R (variant 9, shown in Table 3) mutations. These mutations (except for G192R and K202R) destabilized the capsid linker in extended Gag. The stabilizing effect ($\Delta G_{capsid\_linker} < 0$) on the capsid by the G192R mutation was neutralized ($\Delta G_{capsid\_linker} = 0$) when Gag extended. The K202R mutation (in variant 9) stabilized (individually or predominantly in the presence of other mutations) the capsid linker whereas the individual S462R (in variant 9) destabilized the capsid linker when Gag extended.

Only the K418R (variant 8), and S462R (variant 9) had notable allosteric effect (destabilizing, with $\Delta G_{site} > 0$) on the first Gag cleavage site (Table 3). These two mutations displayed dominating effects over the other mutations present in variants 8 and 9.

Discussion

We set out to generate Gag mutations by HIV-1 RT in a single replication cycle using our virus-free *in vitro* BSL2 platform (without viral fitness selection pressure) to study the native error rate of HIV-RT. We found novel Gag mutations and investigated their effects on Gag thermostability and allosteric communication onto the first Gag cleavage site and capsid linker that could aid in rational drug and vaccine design.

The predominance of transition mutations in our selection-free conditions (n = 31 as opposed to transversions at n =3) is well characterized across most phyla [45-47], and is hypothesized to conserve protein functions. Given that our system was selection free, the bias seemed to be intrinsic to HIV-RT. While mutations may arise from the (i) DH5α and/or HEK293 EXPI host, (ii) Q5 polymerase and (iii) sequencing artefacts. These were controlled for given that the observed phenotypic mutation rate per base pair per replication for both *E. coli* and mammalian cells were estimated to be

approximately 5.4×10^{-10} and 5.0×10^{-11} respectively [48], and that Q5 polymerase is the most accurate commercially available high-fidelity polymerase[49].

Novel Gag mutations

Of our seven novel Gag mutations (G192R, S241I, K290R, T204A, R214G, S462R, and F433L), variants 2 (G192R), 3 (S241I), and 4 (F433L) were found as single mutation variants. Mutant variants 2 and 3 increased thermostability similar to variants 1 (E17K), 5 (N271S), and 6 (F484S). Whereas variant 4 (F433L) gave a general unique pattern of thermostabilizing the rest of Gag, while decreasing the thermostability of p6 domain (Figure 2). K290R, T204A, R214G and S462R novel mutations were found together with other known mutations. Given that these were unreported in clinical settings, they could be present in viral populations as a minority due to the compromise of viral fitness, or having other deleterious effects. G192 and F433 positions are highly conserved with no naturally occurring polymorphisms reported in the Los Alamos mutation database [32]. One reported G192W variant was generated using low-fidelity PCR mutagenesis and resulted in non-viable virions[50]. At the same time, there are no known F433 mutations reported. On the other hand, the S241 residue position is less conserved, with mutations S241A and S241D being reported to confer deleterious effects [51,52]. Since many of our novel mutations yielded similar thermostability profiles with those of the known mutation variants, they are likely to confer the same fitness effects and play a role in drug/immune resistance. The exception being variant 4, which exhibited effects on p6, potentially affecting conformational changes of Gag during maturation [3].

With the exception of F433L on the fourth cleavage site, the remaining novel mutations were non-cleavage site mutations. Non-cleavage site mutations may act to confer drug-resistance to PI through compensatory roles [3,4,53,54]. Five of the capsid

novel mutations (G192R, T204A, R214G, S241I, and K290R) were not previously reported, suggesting that they may have detrimental effects on virion formation [9], or predispose the virus to increased immune surveillance. Given that the G192R, T204A, R214G and S241I mutations are located at the N-terminal domain (NTD), virions possessing such mutations may suffer from reduced core formation, resulting in poorer viral replication and less infectivity [9,10]. The K290R mutation is on the C-terminal Domain (CTD), within a highly conserved motif known as the major homology region (MHR) [9,10,55], which Mammano *et al.* demonstrated that point mutations substantially impaired viral particle formation [55]. In addition, von Schwedler *et al.* found K290R mutants had aberrant capsid structures by artificially performing site directed mutagenesis to generate K290R mutants [56]. While it would be ideal to study the mutation effects at the protein level, Gag protein expression is notoriously difficult [57]. Nonetheless, given that a number of the novel mutations appear on the capsid, there could be potential vaccine relevance to target such mutations, especially if they result in notable capsid structural changes.

The novel mutation S462R on p6 is likely to affect the role of p6 on Gag conformational changes [3]. Given its absence in reports, with the fact that the mutation was found together with other mutations to display a thermostability overall profile (variant 9 bearing the S462R) similar to the single mutation variants (1, 2, 3, 5, and 6), S462R may play a compensatory role that requires further detailed investigation. In our study, we also found the NC/sp1 cleavage site mutation (F433L in ^{429}RQAN/FLG^{435}) in a site previously linked to PI resistance e.g. Q430R (BILA-2185BS) and A431V (KNI-272, Indinavir, Ritonavir and Saquinavir) [3,4].

Overall, none of the novel mutations were found on the restricted Gag epitopes involving HLA-B27 (KK10, ^{263}KRWIILGLNK^{272}) and HLA-B57 (TW10,

^{240}TSTLQEQIAW^{249}; KF11, ^{162}KAFSPEVIPMF^{172}), perhaps requiring a repertoire to find Gag mutations relevant to CTL responses and vaccine development.

Our analysis of allosteric effects showed that most of the Gag single mutations did not have significant allosteric effects onto the capsid linker or the first Gag cleavage site. We were not able to study beyond the first cleavage site which would involve modelling the structure dynamics of the cleaved Gag units. Of the mutations with effects, G192R stabilized the capsid linker and destabilized the first cleavage site. Yet these effects were abated in the extended Gag. Generally, the single mutations (variants 2 and 5) did not affect the cleavage site at the extended form, suggesting that the mutations were not random as expected. This observation was supported by the fact that Gag variants with multiple mutations also had only weak allosteric effects towards the capsid linker and the first cleavage site showing that the spontaneous mutations avoided drastic effects on critical sequences/regions. Even with effects in the compact conformation, the effects were diminished when Gag extended with very few exceptions (such as K418R in variant 8, which had the highest allosteric effect on the capsid linker when Gag was extended, but on the first cleavage site when Gag was compact; and S462R in variant 9, which destabilized the first cleavage site for extended conformations, see Table 3).

Potential for pre-emptive drug design and vaccine development

In the absence of selection pressures, low frequency occurring mutations may be picked up, allowing for a more comprehensive analysis of the HIV-1 RT bias. These low frequency mutations may be easily missed using current methods, but may emerge in the presence of selection pressures. Thus, by combining such a platform together with computational aided design methods (e.g. machine learning -based predictive models), it may be possible to pre-emptively predict emerging mutations and design next

generation drugs. Given that many HIV proteins can function in intense drug/immune selection environments even with significant reduced activity [58,59], it may be possible for our novel Gag mutations to be detected in the right conditions. Coupled with pre-emptive drugs/vaccine development against such variants, we can subject HIV infection towards Muller's ratchet [60-62] for viral clearance.

The identification of mutational hotspots can be complemented by providing huge datasets of hotspots that arise in the absence of selection pressures and coupled with computational structural analysis, we may be able to understand the constraints of HIV-RT mutations that do not proceed to a protein level under the most basic biological selection pressures i.e. at protein functional levels, prior to the presence of immune/drug selection.

Conclusion

We have established a proof-of-concept to generate unbiased HIV-1 RT mutations on HIV-1 Gag and investigated their effects on Gag thermostability and allosteric communications *in silico*. Our platform can be used for the rapid generation of novel HIV mutations in a safe BSL2 manner, and when coupled with computational analysis using our Gag models, we provide more insights to the drug and vaccine development against HIV-1.

Disclosure of potential conflicts of interest

There are no conflicts of interest

Acknowledgments

We thank Wai-Heng Lua for useful comments and discussion for the experimental component.

Funding

This work was supported by the A*STAR Industry Alignment Fund (IAF) grant IAF111149 and institute core funds.

Author contributions

PY performed all the *in vitro* and computational experiments. DWSK performed repeats of computational experiments and generated the final figures. PY, CTTS, and SKEG discussed, analysed the results, and drafted the manuscript. KFC provided assistance with the making and running of the computational analysis as well as created the AR used in this article. SKEG conceived the idea and supervised the whole project. All authors read and approved the manuscript.

References

1. WHO. HIV/AIDS: WHO; [cited 2019 Apr 9]. Available from: http://www.who.int/gho/hiv/en/
2. Castelli JC, Levy JA. HIV (Human Immunodeficiency Virus). In: Bertino JR, editor. Encyclopedia of Cancer (Second Edition). New York: Academic Press; 2002. p. 407-416.
3. Su CT-T, Kwoh C-K, Verma CS, et al. 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. 2017:1-12.
4. Clutter DS, Jordan MR, Bertagnolio S, et al. HIV-1 drug resistance and resistance testing. Infection, Genetics and Evolution. 2016;46:292-307.
5. Fun A, van Maarseveen NM, Pokorná J, et al. HIV-1 protease inhibitor mutations affect the development of HIV-1 resistance to the maturation inhibitor bevirimat. Retrovirology. 2011;8(1):70.
6. Fun A, Wensing AMJ, Verheyen J, et al. Human Immunodeficiency Virus gag and protease: partners in resistance. Retrovirology. 2012;9(1):63.
7. Clavel F, Mammano F. Role of Gag in HIV Resistance to Protease Inhibitors. Viruses. 2010;2(7):1411-1426.
8. Dam E, Quercia R, Glass B, et al. Gag Mutations Strongly Contribute to HIV-1 Resistance to Protease Inhibitors in Highly Drug-Experienced Patients besides Compensating for Fitness Loss. PLoS Pathogens. 2009;5(3):e1000345.
9. Bell NM, Lever AML. HIV Gag polyprotein: processing and early viral particle assembly. Trends in Microbiology. 2013;21(3):136-144.
10. Göttlinger HG. HIV-1 Gag: a Molecular Machine Driving Viral Particle Assembly and Release.27.
11. Shi J, Zhou J, Halambage UD, et al. Compensatory Substitutions in the HIV-1 Capsid Reduce the Fitness Cost Associated with Resistance to a Capsid-Targeting Small-Molecule Inhibitor. Journal of Virology. 2015;89(1):208-219.

12. Bhattacharya A, Alam SL, Fricke T, et al. Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. Proceedings of the National Academy of Sciences. 2014;111(52):18625-18630.
13. Thenin-Houssier S, T. Valente S. HIV-1 Capsid Inhibitors as Antiretroviral Agents. Current HIV Research. 2016;14(3):270-282.
14. Goulder Philip JR, Walker Bruce D. HIV and HLA Class I: An Evolving Relationship. Immunity. 2012;37(3):426-440.
15. Payne RP, Kløverpris H, Sacha JB, et al. Efficacious Early Antiviral Activity of HIV Gag- and Pol-Specific HLA-B*2705-Restricted CD8+ T Cells. Journal of Virology. 2010;84(20):10543-10557.
16. Brockman MA, Schneidewind A, Lahaie M, et al. Escape and Compensation from Early HLA-B57-Mediated Cytotoxic T-Lymphocyte Pressure on Human Immunodeficiency Virus Type 1 Gag Alter Capsid Interactions with Cyclophilin A. Journal of Virology. 2007;81(22):12608-12618.
17. McElrath MJ. Immune Responses to HIV Vaccines and Potential Impact on Control of Acute HIV-1 Infection. The Journal of Infectious Diseases. 2010;202(S2):S323-S326.
18. Williamson A-L, Rybicki EP. Justification for the inclusion of Gag in HIV vaccine candidates. Expert Review of Vaccines. 2016;15(5):585-598.
19. Gao Y, McKay P, Mann J. Advances in HIV-1 Vaccine Development. Viruses. 2018;10(4):167.
20. Naldini L, Blömer U, Gallay P, et al. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science. 1996;272(5259):263-267.
21. Chan WT, Verma Chandra S, Lane David P, et al. A comparison and optimization of methods and factors affecting the transformation of Escherichia coli. Bioscience Reports. 2013;33(6):931-937.
22. Ling W-L, Lua W-H, Poh J-J, et al. Effect of VH–VL Families in Pertuzumab and Trastuzumab Recombinant Production, Her2 and FcγIIA Binding. Frontiers in Immunology. 2018;9.
23. Lua W-H, Ling W-L, Yeo JY, et al. The effects of Antibody Engineering CH and CL in Trastuzumab and Pertuzumab Recombinant Models: Impact on antibody production and antigen-binding. Scientific Reports. 2018;8(1):718.
24. Phua S-X, Chan K-F, Su Chinh T-T, et al. Perspective: The promises of a holistic view of proteins—impact on antibody engineering and drug discovery. Bioscience Reports. 2019;39(1):BSR20181958.
25. Lua W-H, Su CT-T, Yeo JY, et al. Role of the IgE variable heavy chain in FcεRIα and superantigen binding in allergy and immunotherapy. Journal of Allergy and Clinical Immunology. 2019.
26. Sim J-Z, Nguyen P-V, Lee H-K, et al. GelApp: Mobile gel electrophoresis analyser. Nature Methods Application Notes. 2015.
27. Poh JJ, Gan SKE. The Determination of Factors involved in Column-Based Nucleic Acid Extraction and Purification. Journal of Bioprocessing & Biotechniques. 2014;04(03).
28. Katoh K, Misawa K, Kuma Ki, et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research. 2002;30(14):3059-3066.
29. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology. 2011;7(1):539.
30. Artimo P, Jonnalagedda M, Arnold K, et al. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Research. 2012;40(W1):W597-W603.
31. Sim JZ, Nguyen PV, Zang Y, et al. DNA2App: Mobile sequence analyser. Scientific Phone Apps and Mobile Devices. 2016;2(1):2.
32. HIV sequence database main page [cited 2019 May 19]. Available from: https://www.hiv.lanl.gov
33. The PyMOL Molecular Graphics System. Version 2.2. Schrödinger, LLC.

34. van Gunsteren WF, Billeter S, Eising A, et al. Biomolecular simulation: the {GROMOS96} manual and user guide. 1996.
35. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis. 1997;18(15):2714-2723.
36. Frappier V, Chartier M, Najmanovich R. Applications of Normal Mode Analysis Methods in Computational Protein Design. In: Samish I, editor. Computational Protein Design. Vol. 1529. New York, NY: Springer New York; 2017. p. 203-214.
37. Frappier V, Chartier M, Najmanovich RJ. ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability. Nucleic Acids Research. 2015;43(W1):W395-W400.
38. Schymkowitz J, Borg J, Stricher F, et al. The FoldX web server: an online force field. Nucleic Acids Research. 2005;33:W382-W388.
39. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology. 1993;234(3):779-815.
40. Webb B, Sali A. Comparative protein structure modeling using MODELLER. Current protocols in bioinformatics. 2014;47(1):5.6. 1-5.6. 32.
41. Guarnera E, Berezovsky IN. Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. PLOS Computational Biology. 2016;12(3):e1004678.
42. Davey NE, Satagopam VP, Santiago-Mozos S, et al. The HIV Mutation Browser: A Resource for Human Immunodeficiency Virus Mutagenesis and Polymorphism Data. PLOS Computational Biology. 2014;10(12):e1003951.
43. Munro JB, Nath A, Färber M, et al. A conformational transition observed in single HIV-1 Gag molecules during in vitro assembly of virus-like particles. Journal of Virology. 2014;88(6):3577-3585.
44. Jiang J, Ablan SD, Derebail S, et al. The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability. Virology. 2011;421(2):253-265.
45. Lyons DM, Lauring AS. Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses. Molecular Biology and Evolution. 2017;34(12):3205-3215.
46. Stoltzfus A, Norris RW. On the Causes of Evolutionary Transition:Transversion Bias. Molecular Biology and Evolution. 2016;33(3):595-602.
47. Rosenberg MS. Patterns of Transitional Mutation Biases Within and Among Mammalian Genomes. Molecular Biology and Evolution. 2003;20(6):988-993.
48. Drake JW, Charlesworth B, Charlesworth D, et al. Rates of spontaneous mutation. Genetics. 1998;148(4):1667-1686.
49. Potapov V, Ong JL. Examining sources of error in PCR by single-molecule sequencing. PloS One. 2017;12(1):e0169774.
50. Rihn SJ, Wilson SJ, Loman NJ, et al. Extreme genetic fragility of the HIV-1 capsid. PLoS pathogens. 2013;9(6):e1003461.
51. Wacharapornin P, Lauhakirti D, Auewarakul P. The effect of capsid mutations on HIV-1 uncoating. Virology. 2007;358(1):48-54.
52. Brun S, Chaloin L, Gay B, et al. Electrostatic repulsion between HIV-1 capsid proteins modulates hexamer plasticity and in vitro assembly. Proteins: Structure, Function, and Bioinformatics. 2010;78(9):2144-2156.
53. B. Nachega J, C. Marconi V, U. van Zyl G, et al. HIV Treatment Adherence, Drug Resistance, Virologic Failure: Evolving Concepts. Infectious Disorders - Drug Targets. 2011;11(2):167-174.
54. Sutherland KA, Mbisa JL, Cane PA, et al. Contribution of Gag and protease to variation in susceptibility to protease inhibitors between different strains of subtype B human immunodeficiency virus type 1. Journal of General Virology. 2014;95:190-200.
55. Mammano F, Ohagen A, Hoglund S, et al. Role of the Major Homology Region of Human Immunodeficiency Virus Type 1 in Virion Morphogenesis. Journal of Virology. 1994;68:10.

56. von Schwedler UK, Stray KM, Garrus JE, et al. Functional Surfaces of the Human Immunodeficiency Virus Type 1 Capsid Protein. Journal of Virology. 2003;77(9):5439-5450.
57. McKinstry WJ, Hijnen M, Tanwar HS, et al. Expression and purification of soluble recombinant full length HIV-1 Pr55Gag protein in Escherichia coli. Protein Expression and Purification. 2014;100:10-18.
58. Chung S, Miller JT, Lapkouski M, et al. Examining the role of the HIV-1 reverse transcriptase p51 subunit in positioning and hydrolysis of RNA/DNA hybrids. Journal of Biological Chemistry. 2013;288(22):16177-16184.
59. Betancor G, Alvarez M, Marcelli B, et al. Effects of HIV-1 reverse transcriptase connection subdomain mutations on polypurine tract removal and initiation of (+)-strand DNA synthesis. Nucleic Acids Research. 2015;43(4):2259-2270.
60. Hu W-S, Temin HM. 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. 1990;87(4):1556-1560.
61. Yuste E, Sánchez-Palomino S, Casado C, et al. Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. Journal of Virology. 1999;73(4):2745-2751.
62. Ormond L, Liu P, Matuszewski S, et al. The combined effect of oseltamivir and favipiravir on influenza A virus evolution. Genome biology and evolution. 2017;9(7):1913-1924.
63. Granier C, Battivelli E, Lecuroux C, et al. Pressure from TRIM5 Contributes to Control of HIV-1 Replication by Individuals Expressing Protective HLA-B Alleles. Journal of Virology. 2013;87(18):10368-10380.
64. Schneidewind A, Tang Y, Brockman MA, et al. Maternal Transmission of Human Immunodeficiency Virus Escape Mutations Subverts HLA-B57 Immunodominance but Facilitates Viral Control in the Haploidentical Infant. Journal of Virology. 2009;83(17):8616-8627.
65. van Bockel DJ, Price DA, Munier ML, et al. Persistent Survival of Prevalent Clonotypes within an Immunodominant HIV Gag-Specific CD8+ T Cell Response. The Journal of Immunology. 2011;186(1):359-371.
66. Sanchez-Merino V, Nie S, Luzuriaga K. HIV-1-Specific CD8+ T Cell Responses and Viral Evolution in Women and Infants. The Journal of Immunology. 2005;175(10):6976-6986.
67. Sanchez-Merino V, Farrow Melissa A, Brewster F, et al. Identification and Characterization of HIV-1 CD8+ T Cell Escape Variants with Impaired Fitness. The Journal of Infectious Diseases. 2008;197(2):300-308.
68. Amano M, Tojo Y, Salcedo-Gómez PM, et al. GRL-0519, a novel oxatricyclic ligand-containing nonpeptidic HIV-1 protease inhibitor (PI), potently suppresses replication of a wide spectrum of multi-PI-resistant HIV-1 variants in vitro. Antimicrobial Agents and Chemotherapy. 2013;57(5):2036-2046.
69. Kletenkov K, Hoffmann D, Böni J, et al. Role of Gag mutations in PI resistance in the Swiss HIV cohort study: bystanders or contributors? Journal of Antimicrobial Chemotherapy. 2016:866-875.
70. Dykes C, Demeter LM. Clinical Significance of Human Immunodeficiency Virus Type 1 Replication Fitness. Clinical Microbiology Reviews. 2007;20(4):550-578.
71. Shi J, Zhou J, Shah VB, et al. Small-Molecule Inhibition of Human Immunodeficiency Virus Type 1 Infection by Virus Capsid Destabilization. Journal of Virology. 2011;85(1):542-549.
72. Price AJ, Jacques DA, McEwan WA, et al. Host Cofactors and Pharmacologic Ligands Share an Essential Interface in HIV-1 Capsid That Is Lost upon Disassembly. PLoS Pathogens. 2014;10(10):e1004459.
73. Svarovskaia ES, Cheslock SR, Zhang WH, et al. Retroviral mutation rates and reverse transcriptase fidelity. Front Biosci. 2003 2003;8(4):d117-134.

74. Abram ME, Ferris AL, Shao W, et al. Nature, Position, and Frequency of Mutations Made in a Single Cycle of HIV-1 Replication. Journal of Virology. 2010;84(19):9864-9878.
75. Rezende L. Nucleoside-analog resistance mutations in HIV-1 reverse transcriptase and their influence on polymerase fidelity and viral mutation rates. The International Journal of Biochemistry & Cell Biology. 2004;36(9):1716-1734.
76. Poh JJ, Phua SX, Chan KF, et al. Commentary: Augmented Reality Scientific Phone Apps –making the APD AR Holistic Review app and using existing AR apps for scientific publications. Scientific Phone Apps and Mobile Devices. 2018;4:4.

Table 1: The eight mutations in the 269 variants that was previously reported.

Mutation	Reported Function	References
I223V	CTL immune evasion resistance. Known compensatory mutation for T242N escape mutation.	[16,63,64]
N271S	CTL immune evasion resistance. Known rare and transient mutation.	[65]
E17K	CTL immune evasion resistance. Known rare and transient mutation. HIV-1 variants selected by a novel PI (GRL-0519)	[66-68]
K418R	Associated with PI resistant variants	[69]
K202R	Capsid inhibitor (PF74) resistance	[11,12,70-72]
F484S	No previously reported function	N.A
H124R	No previously reported function	N.A.
T239A	No previously reported function	N.A

Table 2. Estimated error rate of HIV-1 RT

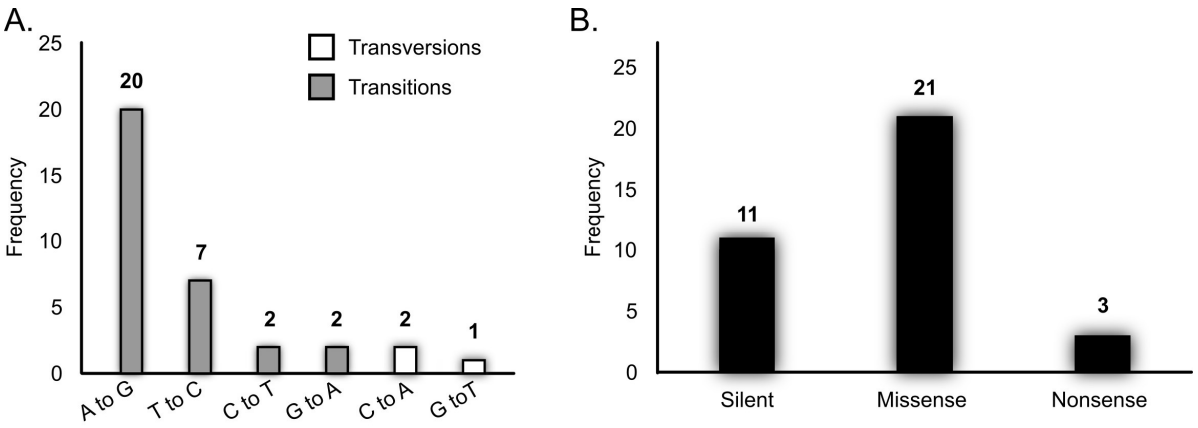
Total no. of Gag Sequences	Total no. of bases	Total no. of mutations	Error rate
269	403,500	35	~8.7 x 10^{-5}

The error rate is calculated as the ratio of total number of mutations and the total number of bases (i.e. 35 / 403,500), which is comparable to the reported range of 5.9 x 10^{-4} to 5.3 x 10^{-5} (reviewed extensively in [73-75]).

Table 3: Estimated changes of allosteric free energy of the capsid linker and the first Gag cleavage site caused by the mutations, using both compact and extended Gag structures.

Variant	Mutations	**Change in allosteric free energies (ΔG_{CA_linker} *or* ΔG_{site})**			
		Compact Gag		Extended Gag	
		^{277}YSPTSIL^{283} (Capsid Linker)	^{374}ATIMIQK^{380} (Cleavage Site)	^{277}YSPTSIL^{283} (Capsid Linker)	^{374}ATIMIQK^{380} (Cleavage Site)
1	E17K	0.21	-0.02	0.90	0.26
2	G192R	-1.33	0.6	0	-0.07
3	S241I	0.04	-0.02	-0.02	-0.02
4	F433L	-0.01	-0.01	0.07	-0.06
5	N271S	0.02	-0.1	-0.01	0.01
6	F484S	0.01	0	0.04	0
7	H124R / K290R	0.08	-0.34	-0.13	0.1
	H124R / ~~K290R~~	0.43	-0.15	-0.07	0.13
	~~H124R~~ / K290R	-0.23	-0.17	-0.06	-0.03
8	T239A / K418R	0.39	-1.33	0.79	0.04
	T239A / ~~K418R~~	-0.05	-0.03	0.01	-0.01
	~~T239A~~ / K418R	0.4	-1.28	0.78	0.04
9	K202R / T204A / R214G / I223V / S462R	0.53	0.11	0.15	0.61
	~~K202R~~ / T204A / R214G / I223V / S462R	0.35	0.22	1.09	0.53
	K202R / ~~T204A~~ / R214G / I223V / S462R	0.56	0.12	-0.09	0.61
	K202R / T204A / ~~R214G~~ / I223V / S462R	0.57	0.11	0.18	0.59
	K202R / T204A / R214G / ~~I223V~~ / S462R	0.90	0.03	0.18	0.55
	K202R / T204A / R214G / I223V / ~~S462R~~	0.35	-0.07	-0.21	0.09
	K202R / ~~T204A~~ / ~~R214G~~ / ~~I223V~~ / ~~S462R~~	0.42	-0.08	-0.46	0.07
	~~K202R~~ / T204A / ~~R214G~~ / ~~I223V~~ / ~~S462R~~	-0.02	-0.01	0.08	0.04
	~~K202R~~ / ~~T204A~~ / S241I / ~~I223V~~ / ~~S462R~~	0.04	-0.02	-0.02	-0.02
	~~K202R~~ / ~~T204A~~ / ~~R214G~~ / I223V / ~~S462R~~	-0.02	0.04	0	0
	~~K202R~~ / ~~T204A~~ / ~~R214G~~ / ~~I223V~~ / S462R	0.31	0.2	0.89	0.59

Figure legends:



C.

Gag variants	Mutations	Domain
1	E17K	MA
2	G192R	CA
3	S241I	CA
4	F433L	sp1
5	N271S	CA
6	F484S	p6
7	H124R, K290R	MA,CA
8	T239R, K418R	CA,NC
9*	K202R, T204A, R214G, I223V, S462R	CA,CA,CA,CA,p6
10	Nonsense (insertion)	
11*	Nonsense (Premature stop codon)	

Figure 1: List of generated HIV-1 Gag mutations. (A) Histogram showing the mutations captured on the 18 mutated Gag sequences. Transversion and Transition mutations are shown in white and grey, respectively. (B) The 35 point mutations detected in this study included 11 silent, 21 missense and 3 nonsense mutations. (C) List of Gag variants possessing non-synonymous mutations. Single mutations (variants 1 to 6), multiple mutations (variants 7 to 9) and truncating mutations (variants 10 & 11) were identified. Eight known mutations listed in the Los Alamos HIV database are underlined. The Gag domains where the mutations are located are indicated in the third column.

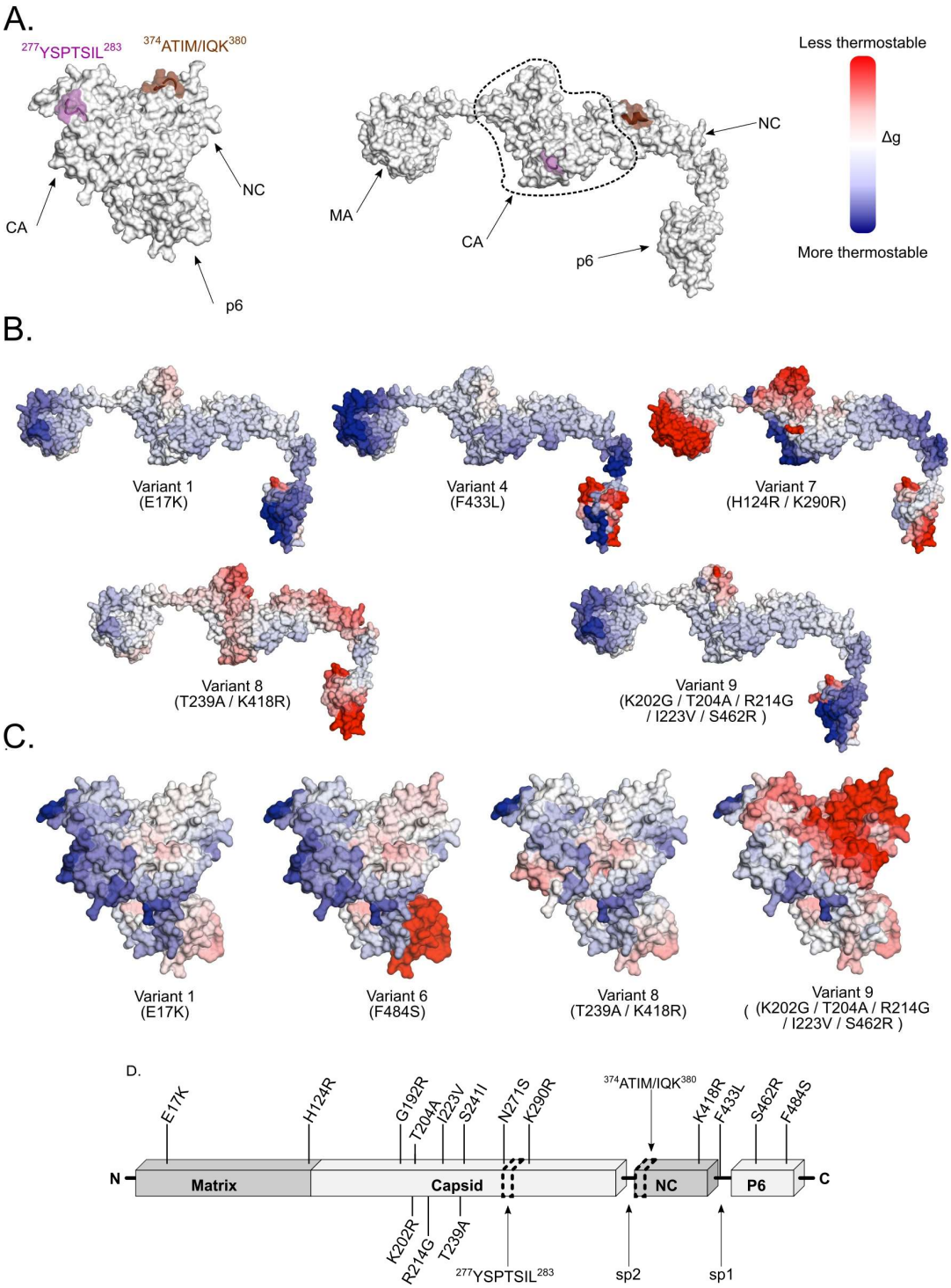


Figure 2: Changes in the Gag thermostability (A). Wild type Gag with the capsid linker (^{277}YSPTSIL^{283}) and the first Gag cleavage site (^{374}ATIM/IQK^{380}). Representative thermostability profiles of the extended (B) and compact (C) Gag structures. (D) Schematic of the Gag structure including domains with the identified mutations.

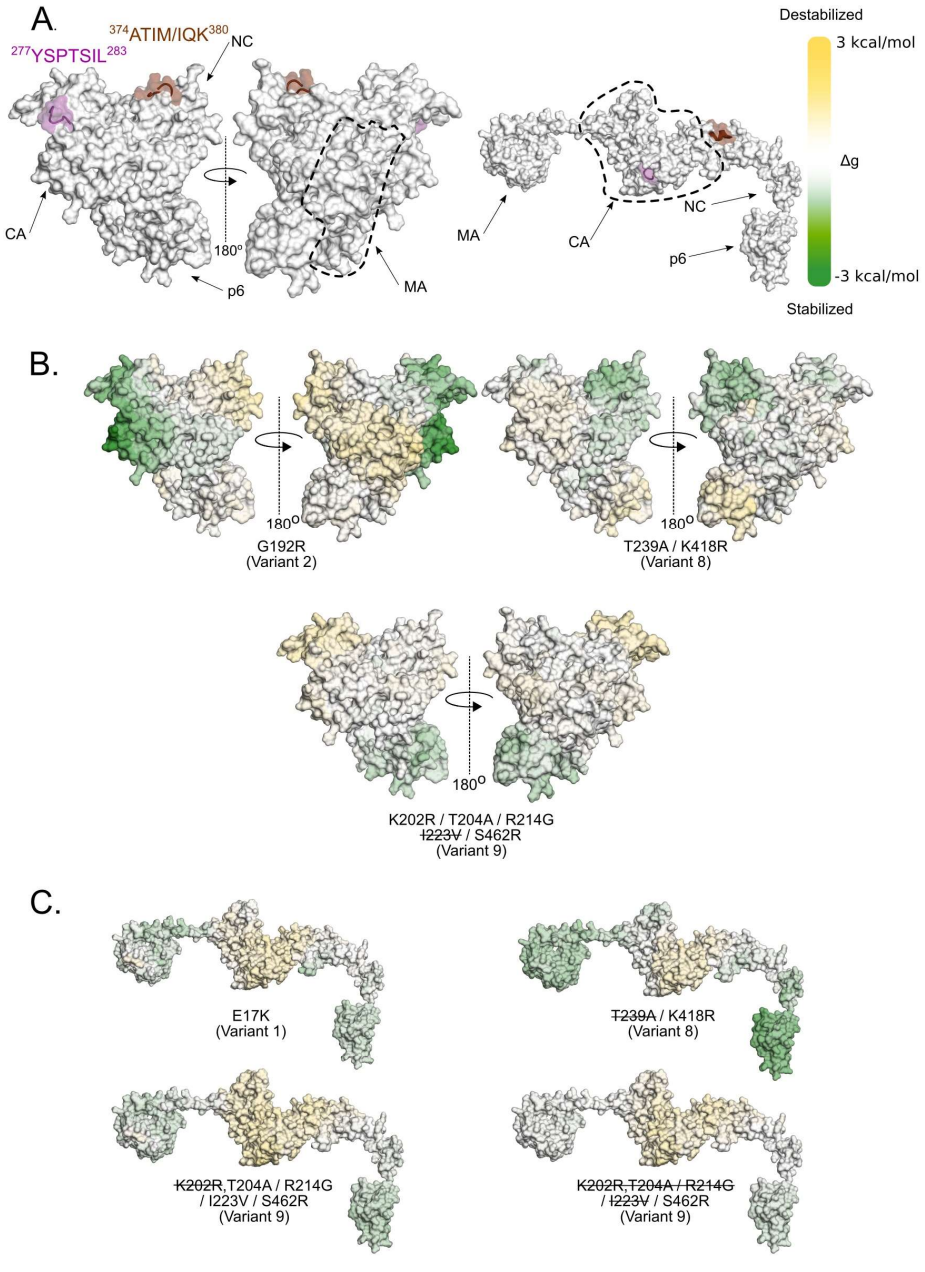


Figure 3: Allosteric communications caused by the identified mutations. (A). Wild type Gag with the capsid linker (^{277}YSPTSIL^{283}) and the first Gag cleavage site (^{374}ATIM/IQK^{380}). Representative allostery profiles of the compact (B) and extended (C) Gag structures. An augmented reality animation of the compact Gag structures can be viewed using the "APD AR Holistic Review App" available on both Google and Apple app stores (for more details, see Poh *et al.* [76]).