






## RESEARCH ARTICLE

# Predominance of positive epistasis among drug resistance-associated mutations in HIV-1 protease

Tian-hao Zhang<sup>1</sup> , Lei Dai<sup>2</sup> \*, John P. Barton<sup>3</sup>, Yushen Du<sup>4,5</sup> , Yuxiang Tan<sup>2</sup>, Wenwen Pang<sup>6</sup>, Arup K. Chakraborty<sup>7,8</sup>, James O. Lloyd-Smith<sup>9</sup> , Ren Sun<sup>5</sup> \*

**1** Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA, **2** CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, **3** Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA, **4** School of Medicine, ZheJiang University, Hangzhou, 210000, China, **5** Molecular and Medical Pharmacology, University of California, Los Angeles, CA 90095, USA, **6** Department of Public Health Laboratory Science, West China School of Public Health, Sichuan University, Chengdu 610041, China, **7** Institute for Medical Engineering and Science, Departments of Chemical Engineering, Physics, & Chemistry, Massachusetts Institute of Technology, MA 21309, USA, **8** Ragon Institute of MGH, MIT, & Harvard, Cambridge, MA 21309, USA, **9** Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA

 These authors contributed equally to this work.

\* [rensun@hku.hk](mailto:rensun@hku.hk) (RS); [lei.dai@siat.ac.cn](mailto:lei.dai@siat.ac.cn) (LD)



## OPEN ACCESS

**Citation:** Zhang T-h, Dai L, Barton JP, Du Y, Tan Y, Pang W, et al. (2020) Predominance of positive epistasis among drug resistance-associated mutations in HIV-1 protease. *PLoS Genet* 16(10): e1009009. <https://doi.org/10.1371/journal.pgen.1009009>

**Editor:** Jianzhi Zhang, University of Michigan, UNITED STATES

**Received:** November 13, 2019

**Accepted:** July 24, 2020

**Published:** October 21, 2020

**Copyright:** © 2020 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All codes are available on <https://github.com/Tian-hao/protease-inhibitor>. All data were deposited in SRA (short read archive) database under accession PRJNA546460 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA546460>).

**Funding:** The work is supported by National Institutes of Health grant AI131294 and AI145038. All works in UCLA virology core lab is supported by University of California, Los Angeles, Center For AIDS Research grant 5P30 AI028697. Lei Dai is

## Abstract

Drug-resistant mutations often have deleterious impacts on replication fitness, posing a fitness cost that can only be overcome by compensatory mutations. However, the role of fitness cost in the evolution of drug resistance has often been overlooked in clinical studies or *in vitro* selection experiments, as these observations only capture the outcome of drug selection. In this study, we systematically profile the fitness landscape of resistance-associated sites in HIV-1 protease using deep mutational scanning. We construct a mutant library covering combinations of mutations at 11 sites in HIV-1 protease, all of which are associated with resistance to protease inhibitors in clinic. Using deep sequencing, we quantify the fitness of thousands of HIV-1 protease mutants after multiple cycles of replication in human T cells. Although the majority of resistance-associated mutations have deleterious effects on viral replication, we find that epistasis among resistance-associated mutations is predominantly positive. Furthermore, our fitness data are consistent with genetic interactions inferred directly from HIV sequence data of patients. Fitness valleys formed by strong positive epistasis reduce the likelihood of reversal of drug resistance mutations. Overall, our results support the view that strong compensatory effects are involved in the emergence of clinically observed resistance mutations and provide insights to understanding fitness barriers in the evolution and reversion of drug resistance.

## Author summary

Antiretroviral drugs have achieved great success in controlling the HIV pandemic. However, the therapy fails sometimes owing to the low drug adherence and/or the emergence

supported by National Key R&D Program of China (2019YFA09006700), National Natural Science Foundation of China (31971513) and Shenzhen Institute of Synthetic Biology(DWKF20190001). James O. Lloyd-Smith is supported by National Science Foundation (DEB 1557022). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

of resistance associated mutations on viral genome. The persistence of drug resistance poses challenges in using antiretroviral drugs for long term control or pre-exposure prophylaxis. To understand the mechanisms of resistance evolution and persistence, we profiled the replication fitness of over 1000 HIV-1 mutants with combinations of resistance associated mutations on its protease gene. We found that although resistance associated mutations greatly reduce replication fitness, they interact positively to alleviate the mutational load. These genetic interactions, termed epistasis, increase the ruggedness along the evolution paths, restricting resistance associated mutations from reversal. Our data support the clinical observations that drug resistance mutations tend to persist even when antiretroviral drug is discontinued.

## Introduction

Antibiotics and antiviral drugs have achieved great success in recent history [1]. However, therapeutic failure may occur due to low adherence and the emergence of drug resistance [2, 3]. The increasing amount of drug resistant pathogens is a global threat to public health [4–11]. The genetic barrier to drug resistance, defined as the number of mutations needed to acquire resistance, is a major determining factor of treatment outcomes [12–14]. Another important but often overlooked aspect of drug resistance is the fitness barrier [15–17]. Resistance associated mutations (RAMs) in pathogen proteins may decrease enzymatic activities, interfere with molecular interactions, or destabilize the protein structure [18–22]. Because of the impaired replication capacity without drug selection, drug-resistant mutants cannot normally outcompete wild-type or establish in the population [23–25]. However, drug-resistant mutants can sometimes reach substantial frequency in the population. Fluctuating drug concentrations may create time windows when drug-resistant mutants replicate better than wild-type virus [26]. Moreover, compensatory mutations can rescue the impaired replication capacity of mutants and stabilize drug resistance [22, 27, 27–29]. Thus, comprehensive quantification of the fitness landscape is needed to predict the evolution of drug resistance [30, 31].

Epistasis, i.e. genetic interactions between mutations, is prevalent in molecular evolution [30–34]. Negative epistasis decreases fitness of the double mutant, posing constraints on gaining multiple mutations [35, 36]. It plays an important role in shaping the local fitness landscape [37]. Positive epistasis increases replication capacity of the double mutant, facilitating pathogens to acquire and maintain drug resistance [38–40]. Positive epistasis may create a fitness valley that prevents drug resistant mutations from reversal [41]. Collectively, positive and negative epistasis determine the topography of the fitness landscape [42] and the course of drug resistance evolution [32]. Empirical studies on the genetic interactions between RAMs, especially in high-order mutants, are still rare [43, 44].

HIV-1 protease inhibitors are important components of combination antiretroviral therapy [45] that target HIV-1 protease enzymatic activity [46, 47]. Second-generation protease inhibitors have extremely high binding affinity to viral protein [48]. Resistance to them typically requires more mutations than resistance to first-generation protease inhibitors and other antiretroviral drugs [49, 50]. For example, mutation K103N on reverse transcriptase is sufficient to confer HIV-1 nevirapine (NVP) resistance [51], while more than 4 *de novo* mutations are needed for protease inhibitor Darunavir (DRV) resistance [52]. Protease inhibitor-resistant viruses with multiple RAMs also have significantly reduced fitness [53, 54]. HIV-1 gained RAMs on protease during sub-optimal protease inhibitor therapy [55]. Most resistance mutations directly affect the binding affinity between HIV-1 protease and the inhibitor, but they are

likely to be deleterious because they also reduce binding to the native substrate of HIV-1 protease. To compensate the deleterious effect, some other RAMs stabilize HIV-1 protease, allowing drug-resistant virus to replicate as efficiently as its parental wild-type virus [27, 56]. The compensatory effects between pairs of RAMs have been studied in several studies and are available on the Stanford HIV drug resistance database [22, 57–61]. Meanwhile, reversals of protease inhibitor resistance-associated mutations were rarely seen clinically, even when therapy was interrupted [62] or when mutant virus infected drug-naïve patients [63, 64]. These observations indicate that epistasis may be important for the evolution of protease inhibitor resistance. Recent analyses of sequence co-variation in drug-targeted HIV Pol proteins (protease, reverse transcriptase and integrase) and co-evolutionary Potts model provide evidence that epistasis plays an important role in drug resistance. Despite being disfavored in the wild-type background, primary resistance mutations can become entrenched by the complex mutation patterns which arise in response to drug therapy [65, 66].

Here, we present a quantitative high-throughput genetics approach [67, 68] to study the fitness distribution and epistasis of HIV-1 protease inhibitor RAMs. Combining these data with clinical data and fitness models, we found that positive epistasis was predominant and especially enriched among RAMs, and prevalent along drug resistance evolutionary paths. Our results suggest that fitness hills created by epistasis result in barriers that entrench RAMs, and thus drug-resistant viruses are unlikely to revert after transmission to drug-naïve patients or discontinuation of anti-retroviral drug treatment.

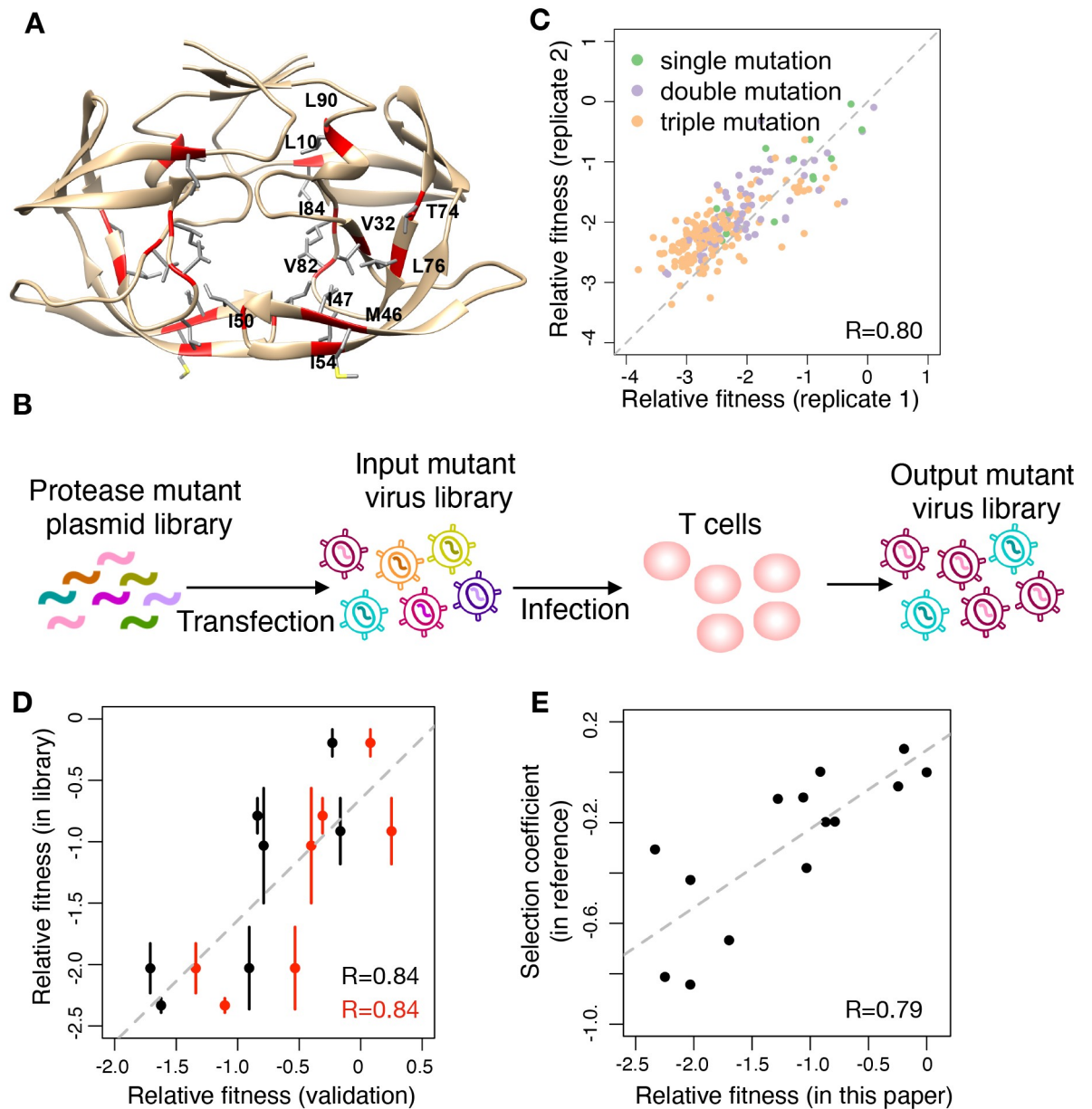
## Results

### Fitness profiling of RAMs in HIV protease

To study the interactions among RAMs in HIV protease, we constructed a library of virus mutants that covers combinations of amino acid substitutions at 11 resistance-associated sites in HIV protease (Fig 1A, Table 1,  $2^9 \times 3^2 = 4608$  genotypes). To ensure sufficient coverage, we harvested more than 30000 colonies after transforming *E. coli*. These sites have been annotated as major drug resistance sites in Stanford Drug Resistance Database [61, 69], and all have been shown to be strongly associated with drug resistance [3]. In our mutant library, 9 sites have one amino acid substitution and the other 2 sites have 2 amino acid substitutions (Fig 1A, Table 1). 2736 out of 4608 possible genotypes (59.38%) were covered in the plasmid library.

We quantified the relative fitness of mutants using high-throughput fitness profiling (Fig 1B, See Material and methods for details). We performed 3 independent transfection experiments to validate the reproducibility of fitness profiling. 20 million 293T cells were transfected and 50 million T cells were infected in each experiment. For each biological replicate, relative fitness was calculated independently. The Pearson's correlation coefficients of single, double and triple mutations between replicates range from 0.80 to 0.82 (Fig 1C and S1 Fig). After filtering out mutants with low frequency or low reproducibility among replicates of input virus libraries (see Material and methods for details), we were able to estimate the relative fitness of 1219 genotypes. The fitnesses of all single mutants, and more than 70% of double and triple mutants, were quantified (S2 Fig).

To validate the quantification of relative fitness, we conducted competition experiments with individually constructed protease mutants. We performed two sets of validation experiments. For the first set, we packaged the mutant virus and wild-type virus independently and mixed them in pairs for head-to-head competition. The frequency of the mutant virus and wild-type virus were quantified by deep sequencing and the relative fitness was calculated in the same way as we did in library screening. A total of 7 mutants were constructed and validated. For the second set of experiments, we mixed all 7 single mutants with wild-type virus in



**Fig 1. High-throughput fitness profiling of combinatorial HIV-1 protease mutant library.** (A) The structure of protease dimer (PDB: 4LL3). The side chains of selected resistance associated residues are shown. (B) Workflow of the fitness profiling. Protease mutations were introduced into NL4-3 background. T cells were infected by the mutant virus library. The frequency of mutants before (input library) and after (output library) selection were deep sequenced. (C) The correlation of relative fitness between two biological replicates. Pearson correlation coefficient ( $R$ ) is 0.80. (D) Two independent validation experiments were performed. We constructed 7 protease single mutant plasmids and recovered viruses independently. We mixed each mutant virus with wild-type virus (validation 1, black dots) and passaged in T cells for 6 days. We also mixed all 7 mutant viruses together with wild-type (validation 2, red dots) and infected T cells for 6 days. The relative fitness of each mutant was quantified by the same means as that in the library. Pearson correlation coefficients ( $R$ ) for validation 1 and validation 2 are both 0.84. Error bar is standard deviation ( $n = 3$ ). (E) The correlation of relative fitness in this study with the experimental selection coefficients in [71]. Pearson correlation coefficients ( $R$ ) is 0.79.

<https://doi.org/10.1371/journal.pgen.1009009.g001>

competition experiments. The relative fitness was defined in the same way. The fitness measured in validation experiments was highly correlated with the fitness in library screening (Fig 1D,  $R = 0.84$  for each independent validation, Pearson's correlation test). In addition, we compared the selection coefficients of HIV-1 protease mutants measured in an independent study

**Table 1. List of protease inhibitor resistance associated mutations covered in the library.** <sup>a</sup> From 148840 subtype B protease sequences in Los Alamos Database [70]. <sup>b</sup> From 1951 isolates tested in PhenoSense assay [61].

Residue number	Consensus	Mutation	Prevalence in clinical dataset <sup>a</sup>	Occurrence in <i>in vitro</i> dataset <sup>b</sup>
10	L	F	1.54%	10.20%
32	V	I	1.37%	7.53%
46	M	I	4.32%	22.19%
47	I	V	0.88%	4.36%
50	I	V	0.30%	1.85%
54	I	L	0.68%	4.92%
54	I	M	0.48%	3.02%
74	T	P	0.37%	2.15%
76	L	V	0.46%	2.92%
82	V	T	0.64%	4.05%
82	V	F	0.33%	1.54%
84	I	V	3.00%	17.12%
90	L	M	7.71%	31.78%

<https://doi.org/10.1371/journal.pgen.1009009.t001>

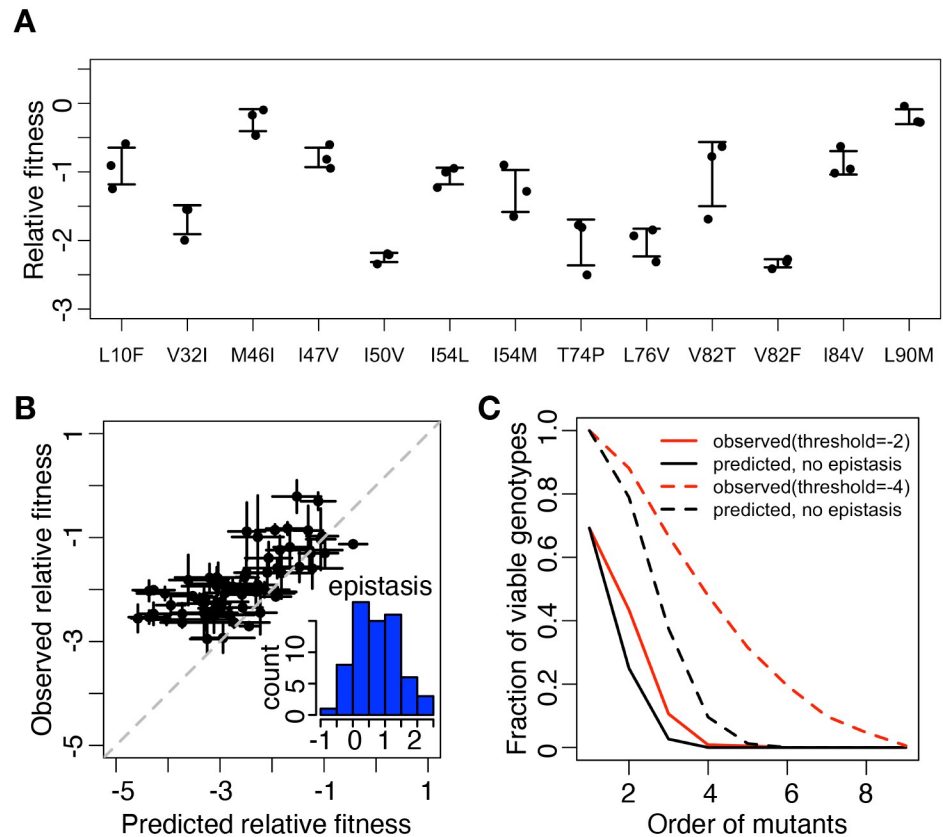
by Boucher et al. [71] and the relative fitness values in our experiment (Fig 1E, S2 Table). The experimental results from two studies show a good correlation (Pearson's correlation coefficient is 0.79), supporting the reliability of our experimental methods.

### Positive epistasis rescues the mutational load of RAMs

We first looked at fitness effect of RAMs. In our definition, a mutant virus of relative fitness  $-1$  means that the relative frequency of this mutant drops 10 fold after infection in cell culture. All single mutations were deleterious to virus replication (Fig 2A). The relative fitness of single mutants ranged from  $-2.33$  (V82F) to  $-0.19$  (L90M). This is consistent with previous reports that randomly introduced mutations were mostly deleterious to protease enzymatic activity or HIV-1 replication capacity [34, 72–74]. Random mutagenesis in other viruses also revealed a lack of beneficial mutations in well-adapted systems [73, 75–77]. RAMs in particular were also reported to be deleterious to virus replication [31, 44]. They may destabilize viral protein, affect enzymatic activities or impact other protein-protein interactions [21, 78].

We then analyzed epistasis between all pairs of RAMs. Previous studies have shown the prevalence of epistasis among pairs of random mutations [34, 37, 75] or spontaneously accumulated mutations [79]. However, studies focused on the epistasis among drug resistance mutations are still limited [30, 39, 72, 75, 80]. Based on the fitness effect of single RAMs, we predicted the relative fitness of double mutants with the assumption that no epistasis existed among any two single mutations (i.e., the predicted relative fitness of a double mutant was the sum of those of two single mutants)(Fig 2B). Surprisingly, the observed relative fitness of most double mutants were significantly higher than the predicted values ( $p = 2.2 \times 10^{-6}$ , two-sided Wilcoxon rank sum test), suggesting that positive epistasis is prevalent among RAMs (Fig 2B inset). Pairwise epistasis between two RAMs is quantified as  $\epsilon_{i,j} = f_{i,j} - f_i - f_j$ ,  $f_i$  represents the relative fitness of mutants  $i$ . The distribution of epistasis ranged from  $-0.69$  (M46I and L90M) to  $2.34$  (L76V and V82F) and 86.6% of pairwise interactions between RAMs are positive.

We also analyzed the extent of epistasis among high-order mutants. We observed a trend that relative fitness decreased as the order of mutants increased (S3 Fig). This is consistent with previous reports that mutational load restricted virus replication capacity [30, 81, 82]. To better quantify the fitness cost of multiple mutations, we calculated the frequency of viable mutants by different thresholds,  $f > -2$  or  $f > -4$ . The frequency of viable mutant virus



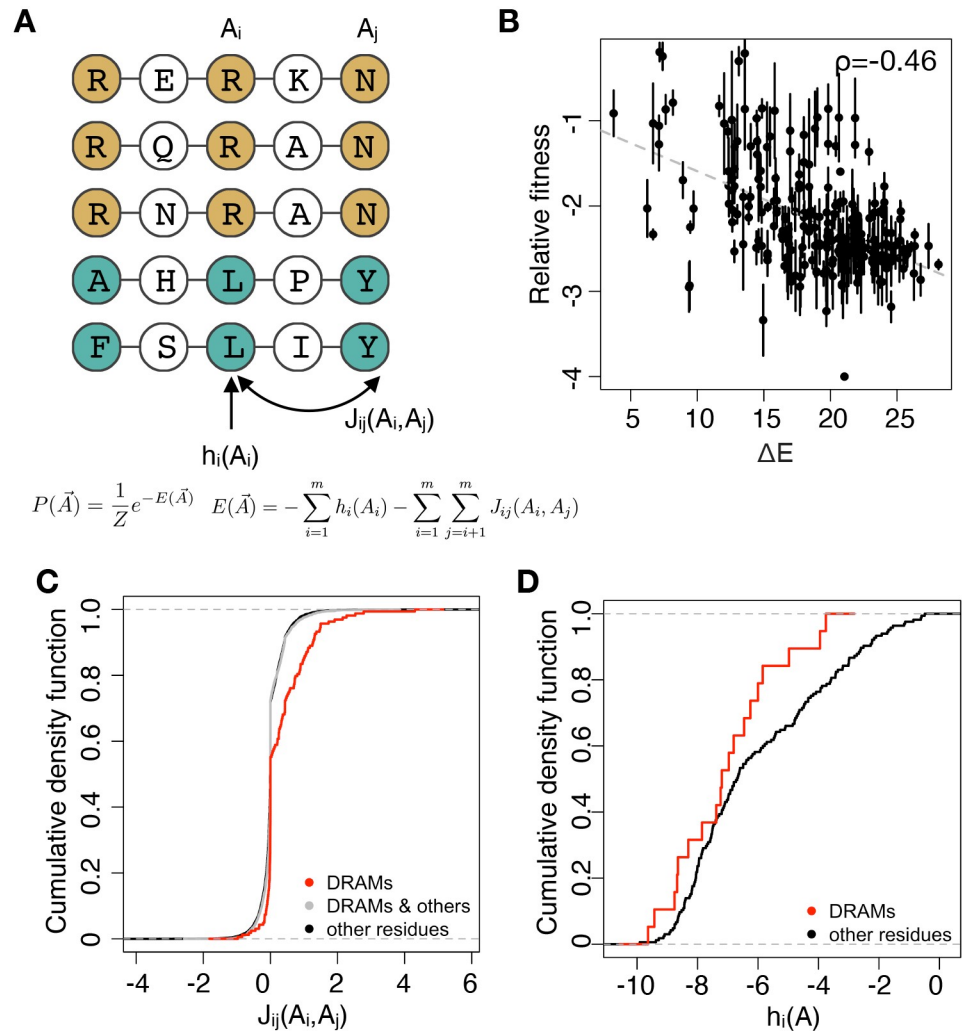
**Fig 2. Positive epistasis is enriched among RAMs.** (A) Relative fitness of single mutants. Error bar is standard deviation ( $n = 3$ ). (B) The predicted relative fitness and observed relative fitness of double mutants. The predicted relative fitness was the sum of that of the two single mutants. Inset, the distribution of epistasis between double mutants. Error bar is standard deviation ( $n = 3$ ). (C) The predicted and observed fraction of viable mutants. A mutant was defined as viable if its relative fitness is higher than  $-4$  (dashed line) or  $-2$  (solid line).

<https://doi.org/10.1371/journal.pgen.1009009.g002>

decreased as the number of mutations increased (Fig 2C), consistent with previous observations in HIV-1 and other RNA viruses [83–86]. We then predicted the relative fitness of high-order mutants by summing the relative fitness of corresponding single mutants. We observed more viable mutants than would be predicted without epistasis (Fig 2C). This indicated pervasive positive epistasis rescued high-order mutants from lethal relative fitness, which is consistent with other clinical observations in protease inhibitor resistant virus [30, 44]. As a result, positive epistasis partially relieved HIV-1 mutational load and allowed viruses to explore more sequence space.

### Enrichment of positive epistasis among RAMs

There are two possible explanations for the observed positive epistasis among RAMs of HIV protease. The first hypothesis is that all mutations in HIV protease tend to interact positively. The second hypothesis is that epistasis among random mutations in HIV protease is on average zero, but positive epistasis is enriched among RAMs. We introduced the Potts model to test our hypotheses, while simultaneously testing whether our finding of prevalent positive epistasis among RAMs carries over to the clinical setting. Potts models, originally developed in statistical physics, have been employed previously to use the population-level frequencies and correlations between different mutations to estimate their fitness effects [87–90]. In the Potts



**Fig 3. Positive epistasis rescues the mutational load of RAMs.** (A) The conceptual graph of Potts model. Potts model uses the probability of mutations occurring with other mutations to estimate the statistical energy.  $h_i$  is the field parameter while  $J_{ij}$  is the coupling parameter. (B) The correlation of Potts energy ( $\Delta E = E_{mut} - E_{WT}$ ) and relative fitness of mutants with lower than 4 RAMs. Spearman correlation coefficient ( $\rho$ ) is  $-0.46$ . (C) The cumulative density function of coupling parameters of RAMs and all other mutations. Coupling parameters between RAMs are more positive than those between RAMs and others ( $D = 0.22, p = 2.1 \times 10^{-7}$ , two-sided K-S test) and those between other residues ( $D = 0.22, p = 5.1 \times 10^{-7}$ , two-sided K-S test). (D) The cumulative density function of field parameters of RAMs and all other mutations. Field parameters of RAMs and other residues are not significantly different ( $D = 0.25, p = 0.20$ , two-sided K-S test).

<https://doi.org/10.1371/journal.pgen.1009009.g003>

model, the probability of observing a genotype  $\vec{A} = \{A_1, A_2, \dots, A_{99}\}$  is given by equations in Fig 3A. Here the  $A_i, i \in \{1, 2, \dots, 99\}$  are variables that represent the amino acid at site  $i$  on each of the 99 sites of protease. Two sets of Potts parameters, fields  $h_i(A_i)$  and couplings  $J_{ij}(A_i, A_j)$ , give the statistical energy  $E(\vec{A})$ , which is negatively correlated with fitness. These parameters are estimated in order to reproduce the frequencies and correlations between mutations that are observed in the data. The fields  $h_i(A_i)$  represent the fitness effect of amino acids  $A_i$  at sites  $i$  alone, while the couplings  $J_{ij}(A_i, A_j)$  describe epistatic interactions between amino acids  $A_i$  at site  $i$  and  $A_j$  at site  $j$ . For both the couplings and the fields, positive parameter values correspond to beneficial effects on fitness, while negative values correspond to deleterious fitness

effects. We applied a maximum entropy method [91] to an alignment of 20911 HIV-1 clade B protease sequences from drug naïve patients, obtained from the Los Alamos National Laboratory HIV sequence database (hiv.lanl.gov, accessed 24 March 2017) to calculate these two sets of Potts parameters.

Then we calculated  $E(\vec{A})$  for all mutants in our protease library. We found that the Potts energy for single, double or triple mutants ( $\Delta E = E_{mut} - E_{WT}$ ) is significantly correlated with the relative fitness we measured in our screening ( $\rho = -0.46$ ,  $p = 1.2 \times 10^{-14}$ , Spearman's correlation test, Fig 3B). The correlation was lower than previous analysis in HIV-1 Gag and Env region [88, 90]. This may be due in part to strong phylogenetic bias on the inferred Potts parameters, because protease is highly conserved. It is also possible that epistatic interactions with cleavage sites on other parts of the HIV-1 genome and complicated anti-innate immunity functions of protease [57, 59, 92] obscure the effects of individual mutations on replicative fitness *in vitro*.

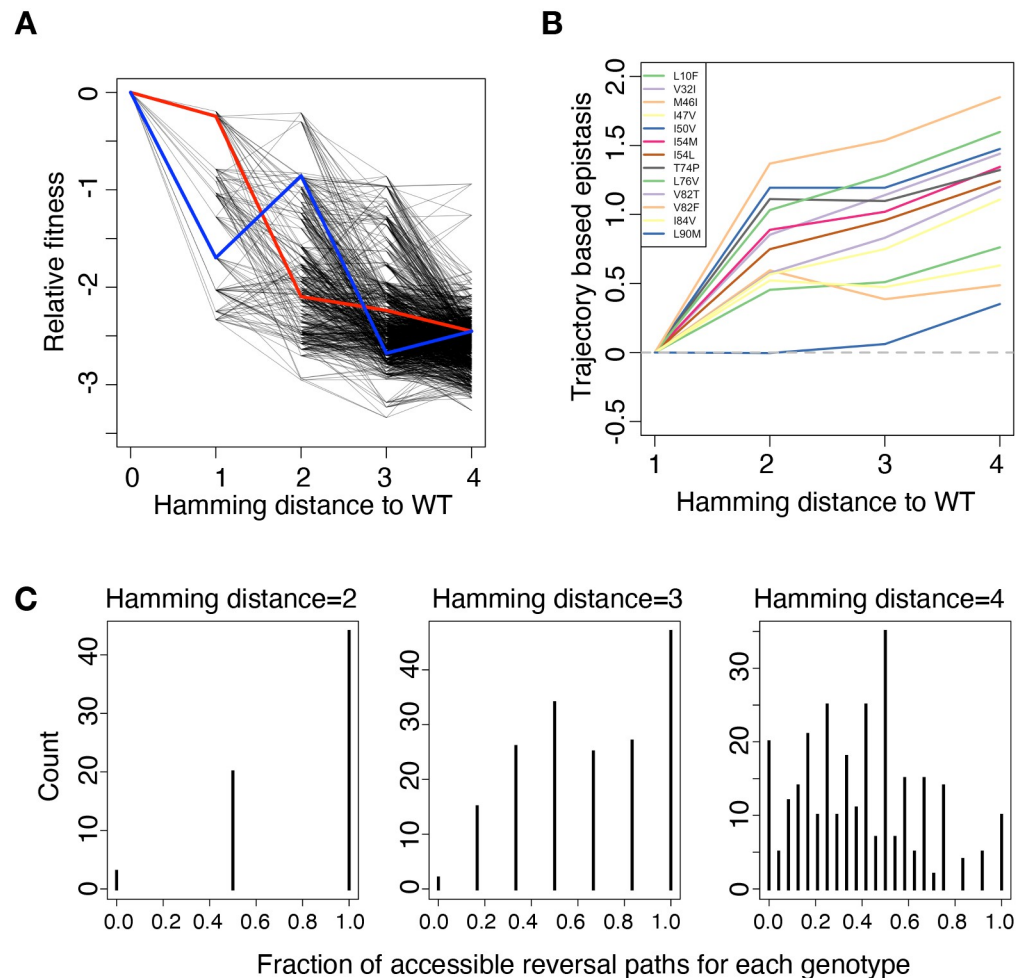
The Potts couplings  $J_{ij}(A_i, A_j)$  give the contribution of pairwise epistatic interactions between amino acids  $A_i$  and  $A_j$  at sites  $i$  and  $j$ , respectively. We compared the couplings among RAMs and among all other possible mutations on protease (Fig 3C). Couplings of other protease mutations clustered near 0, while those of RAMs are significantly more positive than that of other mutations ( $D = 0.22$ ,  $p = 2.1 \times 10^{-07}$ , two-sided K-S test). Moreover,  $J_{ij}(A_i, A_j)$  among RAMs were also more positive than those between RAMs and other residues ( $D = 0.22$ ,  $p = 5.1 \times 10^{-07}$ , two-sided K-S test). Although the fields  $h_i(A_i)$  of RAMs are more negative than other mutations, the difference is not significant (Fig 3D,  $D = 0.25$ ,  $p = 0.20$ , two-sided K-S test). We note that the magnitude and the variation of field parameters is much larger than that of coupling parameters (Fig 3C and 3D). The Interquartile Range (IQR, i.e. the middle 50%) of field parameters is 3.55, while the IQR of coupling parameters is 0.15. The standard deviation of field parameters is 2.29, while the standard deviation of coupling parameters is 0.37. Overall, analysis based on the Potts model is consistent with our experimental results that positive epistasis is enriched among RAMs, and lends support to our second hypothesis that epistasis among random mutations in HIV protease is on average zero.

## Implications of positive epistasis in evolution

To study the role of epistasis in evolution, we analyzed the evolutionary pathways covering all genotypes with up to 4 amino acid substitutions from the wild-type virus (13 single mutants, 67 double mutants, 176 triple mutants and 290 quadruple mutants) (Fig 4A). Mutants are linked if they differ by one amino acid substitution.

We have found that all 13 RAMs are deleterious on the wild-type background (Fig 2A). However, the fitness effect of a single RAM becomes less deleterious on genetic backgrounds where other RAMs have been fixed (S4 Fig). Following the generalized definition of epistasis proposed by Shah et al. [93], we define trajectory-based epistasis  $\epsilon_{M,j}$  that measures the deviation of the fitness effect if the order of mutations were reversed.  $\epsilon_{M,j} = f_{M,j} - f_M - f_j$ , where  $f_M$  and  $f_j$  represent the relative fitness of background  $M$  and single mutant  $j$  [94]. For example, mutation  $j$  can be deleterious on the wild-type background but beneficial on another genetic background that mutation  $i$  has been fixed. Trajectory-based epistasis is calculated for each amino acid substitution and averaged over genetic backgrounds with a certain Hamming distance to the wild-type (Fig 4B). For all RAMs profiled in this study, we find that trajectory-based epistasis is overall positive and increases steadily with the number of substitutions, i.e. the fitness contribution of a specific amino acid substitution becomes more positive if more RAMs have been fixed. Our results are consistent with previous analyses of sequence co-variation in HIV-1 protease [65, 66], where inferred epistatic interactions among mutations at PI





**Fig 4. Ruggedness in fitness landscapes prevents RAMs from reversion to wild-type.** (A) Fitness with possible evolutionary trajectories. Mutants are linked if they only have one residue difference. Red line represents an accessible path that a quadruple mutant can take and reverse to wild-type. Blue line represents an inaccessible reversal path to wild-type for that mutant. (B) Trajectory-based epistasis is calculated for each amino acid substitution and averaged over genetic backgrounds with a certain Hamming distance to the wild-type. The fitness effect of a single mutation becomes less deleterious on genetic backgrounds where other RAMs have been fixed. (C) The distribution of accessible paths for all genotypes with a certain hamming distance to wild type.

<https://doi.org/10.1371/journal.pgen.1009009.g004>

resistance associated sites lead to entrenchment of primary drug resistance mutations. In this study, we combine the analyses of co-variation (Potts model) with comprehensive experimental fitness data of HIV-1 protease mutants (including a large number of higher-order mutants) to provide direct evidence of positive epistasis among RAMs of second-generation PIs.

We tested the hypothesis that positive epistasis prevented resistance associated genotypes from reverting to wild-type [41, 95, 96]. Although RAMs incurred significant fitness cost, some drug resistant mutants would not revert to wild-type after transmitting to a drug naïve patient. We quantified the frequency of accessible evolutionary pathways between mutants and wild-type in our experimentally measured fitness landscape of HIV protease RAMs. A reversal path is defined to be accessible if and only if the virus fitness increases monotonically along the path. For example, quadruple mutant V32I\_M46I\_I54L\_V82F has many paths to revert to wild-type (Fig 4A). Among them, reversing V32I, I54L, V82F and M46I in order is an accessible path (Fig 4A, red line). On the contrary, reversing I54L, V82F, M46I and V32I is not

an accessible path because there are 2 steps with decreasing fitness (Fig 4A, blue line). We found that among double mutants, 44 have two accessible reversal paths to the wild type, 20 have only one accessible reversal path, and interestingly 3 of them have none. These 3 mutants (I50V\_T74P, M46I\_I54M and L76V\_V82F) represent local fitness peaks and the reversal to wild-type is blocked by a fitness valley. We found that the number of accessible reversal paths decreased with the accumulation of RAMs (Fig 4C). This indicates that protease mutants become less likely to revert to wild-type as the number of RAMs increases. Our results are consistent with clinical observations that protease inhibitor resistance associated mutations seldom reverted even when therapies were interrupted [25, 62] or drug-naïve patients were infected [63, 64]. The difficulty of reversal also explains the rising frequency of drug resistant HIV-1 viruses in acute phase patients [41, 96].

## Discussion

In this study, we systematically quantified the fitness effect of RAMs of HIV-1 protease. While all RAMs reduced the virus replication fitness, pervasive positive epistasis among RAMs alleviated the fitness cost substantially. Moreover, we analyzed the HIV sequence data from patients by the Potts model. We found the statistical energy inferred from HIV sequences *in vivo* correlated well with the replication fitness measured *in vitro*. Based on our fitness data and the mutational couplings inferred by the Potts model, we showed that positive epistasis is enriched among RAMs of HIV-1 protease, in both local fitness landscape and evolutionary paths. Finally, we studied the role of epistasis in evolutionary pathways. We found that positive epistasis among RAMs entrenches drug resistance and blocks the reversal paths to wild-type virus, which has important implications for the design of anti-retroviral therapies. Through this project, we also established a high-throughput platform to quantify the genetic interactions among a group of mutations. Another independent study profiled the fitness effect of all single amino acid change on HIV protease [71]. The data showed significant correlation with our study (Fig 1E, Pearson's correlation coefficient ( $R$ ) is 0.79).

There are a few limitations of this study. Firstly, we only measured the fitness effect of RAMs in the absence of protease inhibitors. We are not able to quantify drug resistance of RAMs because protease inhibitors block multiple rounds of virus infection and prevent us from accurate examination of mutant frequency under drug selection. Also, we did not sequence other genes of HIV-1. HIV-1 mutates rapidly due to low fidelity of reverse transcriptase [97, 98]. There might be compensatory mutations occurring on other proteins that rescued the protease RAMs. Secondly, the correlation between our validation experiments and high-throughput screening experiments was less than the correlation observed in similar experiments in bacteria and yeast [99, 100]. The correlation between Potts energy and experimental fitness is also lower than previous reports on Gag and Env regions [88, 90]. Mechanistic difference between logistic growth and viral growth may complicate the quantification of viral fitness [101]. Direct measurement of viral frequency may not linearly correlate to the probability of replication [102]. Moreover, we tested a large number of higher-order mutants (i.e. multiple mutations from the wild-type virus). Our experimental dataset not only contains clinically observed genotypes but also combinations of mutations that was not observed in patients, which are highly deleterious and may suffer from higher experimental errors. If we exclude higher-order mutants and very deleterious genotypes (S5 Fig), the Spearman's correlation between fitness and Potts energy is higher ( $\rho = -0.54$ , compared to  $\rho = -0.46$  in Fig 3B). Thirdly, we did not cover all clinically observed polymorphism, given the bottlenecks in virus library screening. We chose to prioritize for RAMs of second-generation protease inhibitors Darunavir (DRV) and Tipranavir (TPV), which are considered to have high genetic barriers

(i.e. multiple RAMs are involved in the emergence and reversal of drug resistance) [52]. According to Stanford Drug Resistance Database [61, 69], the RAMs that we chose contribute to the resistance to DRV and TPV (S2 Table). The only exception is L90M, which is frequently found in drug resistant viruses. The RAMs and the combinatorial genotypes in our library are prevalent in patients and documented in Stanford Drug Resistance Databases (Table 1). Future work could be extended to cover more clinically observed polymorphism in HIV-1 protease and other drug-targeted proteins. Finally, the correlation between Potts energy and experimental fitness is confounded by many factors, like different selection pressures *in vivo* and *in vitro*, or phylogenetic bias. Nonetheless, we observe moderate but statistically significant correlation between the coupling parameters in the Potts model and the experimental epistasis (S6 Fig, Spearman's correlation test,  $p = 6.8 \times 10^{-3}$ ). We note that the coupling parameters in the Potts model and the experimental measure of epistasis (calculated for WT genetic background) are conceptually different, representing Fourier coefficients and Taylor coefficients of the fitness landscape [103]. Our findings are consistent with the literature that Potts model couplings are strongly associated with contact residues in the three-dimensional structure of protein families [104, 105]. We tested a series of different statistical models, including the binary (Ising) model inferred via ACE, the Potts model inferred via pseudo-likelihood maximization (a popular approach to analyzing sequence data from protein families), and the Potts model inferred via ACE, to examine the epistatic effects among drug resistance mutations (S7 Fig). We found that the Potts model inferred via ACE is the best choice to analyze epistasis in our study.

Statistical models suggest a pervasive negative distribution of fitness effect for single mutations on HIV-1 [31, 88, 106]. Previous models also predicted the entrenchment of deleterious RAMs by positive epistasis [65, 66]. This dataset provides a unique chance to experimentally test these statistical hypotheses. The predominance of positive epistasis is also observed in HIV-1 [30] and in other organisms [39, 107]. However, they either relied on naturally-occurring resistant clones or indirectly activating gene functions. This report is the first dataset to systematically quantify the epistasis among functional residues in HIV-1 drug resistance evolution, without the bias of drug selection and *in vivo* evolution. Overall, our results are important for understanding drug resistance evolution. We found positive epistasis plays a critical role in HIV-1 gaining and maintaining drug resistance. Epistasis makes the fitness landscape rugged, preventing RAMs from reversion to wild-type, even when antiviral therapy is interrupted or virus transmits to a healthy individual [95, 108].

Positive epistasis involves many kinds of molecular mechanisms. We find that the relative fitness of single mutants is not a significant factor of positive epistasis. We compared  $h_i$  in the Potts model for all RAMs and other single mutants. They were not significantly different ( $p = 0.20$ , K-S test). Physical distance between residues is a significant factor contributing to positive epistasis. The physical distances between these residues were significantly less than those between any two random residues on HIV-1 protease ( $D = 0.32$ ,  $p = 3.9 \times 10^{-10}$ , two-sided K-S test, S8 Fig), suggesting that physical contact among RAMs might contribute to the observed positive epistasis. Notably, their average distance was more than 10 Å, indicating most of them did not have direct contact. Some mutations may have structurally stabilizing effect to other residues. We used FoldX and Rosetta to predict the folding free energy ( $\Delta\Delta G$ ) as a quantification of protein stability [109, 110] for all mutants in our library (S8 Fig). We notice that mutation V82F contributed to the positive epistasis on many genetic backgrounds (Fig 4B), but it did not contribute much to the stabilizing effect. Thus, structurally stabilizing effects cannot fully explain the predominance of positive epistasis observed in this study. Future studies on the structure and function of HIV-1 protease mutants will help elucidate the molecular mechanisms underlying the interactions among RAMs.

## Material and methods

### Plasmid library construction

HIV-1 RAMs were picked according to their prevalence in protease inhibitor treated patients [3]. We chose 11 residues with 13 mutations to construct a combination of HIV-1 protease mutant library (Table 1).

We used a ligation-PCR method to construct the library on NL4-3 backbone, which is an infectious subtype B strain. All possible combinations of these 13 mutations are  $2^9 \times 3^2 = 4608$  genotypes. The mutagenesis region spanned 243 nucleotides on HIV-1 genome. We split the region into 5 oligonucleotides and ligate them in order by T4 ligase (from New England BioLabs). The sequence of oligonucleotides are shown in S3 Table. After each ligation, we recovered the product by PCR and used restriction enzyme BsaI-HF (from New England BioLabs) to generate a sticky end for the next step ligation.

After making the 243-nucleotide mutagenesis fragment, we PCR amplified the upstream and downstream regions near this fragment and used overlap extension PCR to ligate them together. We then cloned it into full length HIV-1 NL4-3 background. We harvested more than 30,000 *E. coli* colonies to ensure sufficient coverage of the library complexity.

### Virus production

The plasmid DNA was purified by HiPure Plasmid Midi Prep Kit (from Thermo Fisher Scientific). To produce virus, we used 16  $\mu$ g plasmid DNA and 40  $\mu$ L lipofectamine 2000 (from Thermo Fisher Scientific) to transfect  $2 \times 10^7$  293T cells, in 3 independent biological replicates. We changed media 12 hours post transfection. The supernatant was harvested 48 hours post transfection, labeled as input virus and frozen at  $-80^\circ\text{C}$ . We harvested 40mL viruses from each transfection. Virus was quantified by p24 antigen ELISA kit (from PerkinElmer).

### Library screening

CEM cells were cultured in RPMI 1640 (from Corning) with 10% FBS (from Corning). To passage library in T cells, we added 25 mL viruses and 120  $\mu$ g polybrene to 50 million CEM cells. We achieved 10 ng p24 ( $10^8$  physical viral particles) for every million CEM cells during infection. We washed cells and completely changed media 6 hours post infection. We supplemented the cells with fresh media 3 days post infection and harvested supernatant 6 days post infection. We centrifuged supernatant at  $500 \times g$  for 3 minutes to remove the cells and cell debris. The rest of supernatant was frozen at  $-80^\circ\text{C}$ .

In summary, we carefully controlled the experiment scales to ensure the library complexity was maintained in every step. Briefly, we harvested  $>3 \times 10^4$  *E. coli* colonies during bacteria transformation, which ensured  $\sim 6$ -fold coverage of the expected complexity (4608 genotypes). We then transfected  $2 \times 10^7$  HEK 293T cells with 16  $\mu$ g plasmid library to package infectious viruses. We used 25 mL viruses (500 ng p24,  $\sim 5 \times 10^9$  viral particles) to infect  $2 \times 10^7$  million CEM cells for each biological replicate.

### Sequencing library preparation

We used QIAamp viral RNA mini kit (from QIAGEN) to extract virus RNA from supernatant. We then used DNase I (from Thermo Fisher Scientific) to remove the residual DNA. We used random hexamer and SuperScript III (from Thermo Fisher Scientific) to synthesize cDNA. The virus genome copy number was quantified by qPCR. The qPCR primers are 5' -CCTTGT TGGTCCAAAATGCGAAC-3' and 5' -ATGGCCGGTCCCCCACTCCCT-3'.

At least  $2 \times 10^5$  copies of viral genome were used to make sequencing libraries. We PCR amplified the mutagenesis regions using the following primers: 5' -CTAATCCTGGAGTCTTTGGCAGCGACCC-3' and 5' -GAAGACCTGGAGTGCAGCCAATCTGAGT-3'. We then used BpmI (from New England BioLabs) to cleave the primers and ligate the sequencing adapter to the amplicon. We used PE250 program on Illumina MiSeq platform to sequence the amplicon.

### Calculation of fitness and epistasis

We used custom python codes to map the sequencing reads to reference NL4-3 genome. Mutations were called if both forward and reverse reads have the same mutation and phred quality scores are both above 30. All codes are available on <https://github.com/Tian-hao/teprotease-inhibitor>. All data were deposited in SRA (short read archive) database under accession [PRJNA546460](https://www.ncbi.nlm.nih.gov/sra/PRJNA546460). For each replicate of the virus library from the transfected 293T cells, we reached  $4.45 \times 10^5$  to  $6.05 \times 10^5$  sequencing depth. We filtered out the genotypes with frequency fewer than  $5 \times 10^{-5}$  in any biological replicate and the genotypes whose frequency differ more than 10 folds between any two biological replicates.

Relative fitness  $f_{m,r}$  of mutant  $m$  in experiment  $r$  (biological replicates) was defined as Eq 1.

$$f_{m,r} = \log_{10} \left( \frac{F_{m,r,output}}{F_{m,r,input}} \frac{F_{WT,r,output}}{F_{WT,r,input}} \right) \quad (1)$$

$F_{m,r,input}$  is the frequency of mutant  $m$  before screening.  $F_{m,r,output}$  is the frequency of mutant  $m$  after passaging.  $F_{WT,r,input}$  is the frequency of wild-type virus before screening.  $F_{WT,r,output}$  is the frequency of wild-type virus after passaging.

The relative fitness  $f_m$  was defined as the average of 3 biological replicates (Eq 2). However, if relative fitness was missing in one replicate, we only average the other two replicates. The relative fitness value of all mutants was shown in S1 Table.

$$f_m = \sum_{t=1}^R f_{m,r} / R, \quad (2)$$

where  $R$  is the number of biological replicates.

Pairwise epistasis  $\varepsilon_{i,j}$  between mutant  $i$  and mutant  $j$  was defined as:

$$\varepsilon_{i,j} = f_{i,j} - f_i - f_j, \quad (3)$$

where  $f_{i,j}$  refers to the relative fitness of double mutant  $i$  and  $j$ .

Trajectory-based epistasis  $\varepsilon_{M,j}$  between a multi-mutation genotype  $M$  and another genotype differ by one mutation  $j$  was defined as:

$$\varepsilon_{M,j} = f_{M,j} - f_M - f_j \quad (4)$$

### Potts model

Data used to infer parameters for the Potts model were downloaded from the Los Alamos National Laboratory HIV sequence database, as described in the main text. Sequences were processed as previously described [111]. Briefly, we first removed insertions relative to the HXB2 reference sequence. We also excluded sequences labeled as “problematic” in the database, and sequences with gaps or ambiguous amino acids present at >5% of residues were removed. Remaining ambiguous amino acids were imputed using simple mean imputation.

Each sequence in the multiple sequence alignment (MSA) is represented as a vector of variables  $\vec{A} = \{A_1, A_2, \dots, A_N\}$ , where  $N = 99$  is the length of the sequence. Each of the  $A_i$  represents a (set of) amino acid(s) present at residue  $i$  in the protein sequence. To choose the amino acids at each site that would be explicitly represented in the model, we first computed the frequency  $p_i^*(A)$  of each amino acid  $A$  at each site  $i$  in the MSA. To compute these frequencies, we weighted the sequences such that the weight of all sequences from each unique patient was equal to one, thereby avoiding overcounting in cases where many sequences were isolated from a single individual. We then explicitly modeled the  $q_i$  most frequently observed amino acids at each site that collectively capture at least 90% of the Shannon entropy of the distribution of amino acids at that site [111]. All remaining, rarely observed amino acids were grouped together into a single aggregate state. For these data, this choice resulted in an average of three explicitly modeled states at each site (minimum of 2, maximum of 6).

The Potts model is a probabilistic model for the ‘compressed’ sequences  $\vec{A}$ , where the probability of observing a sequence  $\vec{A}$  is

$$P(\vec{A}) = \frac{1}{Z} e^{-E(\vec{A})}, \quad (5)$$

$$E(\vec{A}) = - \sum_{i=1}^m h_i(A_i) - \sum_{i=1}^m \sum_{j=i+1}^m J_{ij}(A_i, A_j). \quad (6)$$

Here the normalizing factor

$$Z = \sum_{\vec{A}} e^{-E(\vec{A})} \quad (7)$$

ensures that the probability distribution is normalized. We used ACE [91] to infer the set of Potts fields  $h_i(A_i)$  and couplings  $J_{ij}(A_i, A_j)$  that result in average frequencies and correlations between amino acids in the model (5) that match the frequencies  $p_i^*(A_i)$  and correlations  $p_{ij}^*(A_i, A_j)$  observed in the data. We used a regularization strength of  $\gamma = 7 \times 10^{-5}$  in the inference, which is roughly equal to one divided by the number of unique patients from which the sequence data were obtained. We used “consensus gauge,” where the fields and couplings for the most frequent residue at each site in the protein are set to zero. We confirmed that the parameters inferred by ACE resulted in a Potts model that accurately recovered the correlations present in the data.

## Validation experiments

We constructed 7 single mutants by site-directed mutagenesis. The primers used in this experiment are listed in S3 Table. We used overlap-extension PCR to amplify the fragment with mutated nucleotides. We ligated the fragment with NL4-3 backbone using ApaI and SbfI. We transformed competent *E.coli* and picked single colonies. We sequenced the protease region of plasmids to make sure there is only desired mutant in this region. 7 mutants were L10F, I47V, T74P, L76V, V82F, V82T, L90M.

We produced mutant viruses in 293T cells, mixed them with wild-type and infected CEM cells. The frequencies of mutant virus before and after infection were quantified by deep sequencing. We did 2 biological replicates with each validation method. For validation 1, we pairwise mixed the mutant and wild-type virus for competition. For validation 2, we mixed all 7 mutants and wild-type virus.

## Protein stability prediction

Mutants stability was predicted using either FoldX or Rosetta. For FoldX, we used the protease structure (PDB: 3S85) as reference and repaired the structure using the RepairPDB function. The free energy of the mutants was computed by using the BuildModel function under default parameters. For Rosetta analysis, we used the protease crystal structure (PDB: 6DGX) as reference and score function `ddg_monomer` to evaluate the effect of mutations. Each mutants were evaluated 10 times and the average score was used as  $\Delta\Delta G$ .

## Ethics statement

Reagents were acquired from the NIH AIDS Reagent program. The work is approved by UCLA IRB.

## Supporting information

**S1 Fig. The correlation of relative fitness among biological replicates.** All single mutants, double mutants and triple mutants are shown.  $R$  stands for Pearson correlation coefficient. (TIFF)

**S2 Fig. Coverage of protease mutant library.** (A) Fraction of expected protease mutants in each transfection virus library. (B) Number of mutant in each transfection virus library. Dashed line represents the number of all possible combinations of mutations. (TIFF)

**S3 Fig. Relative fitness of different order of mutations.** (TIFF)

**S4 Fig. Relative fitness of single RAMs on different genetic backgrounds.** (TIFF)

**S5 Fig. Correlation between Potts energy and relative fitness for low order mutants.** Mutants with relative fitness higher than  $-2.5$  and numbers of mutations lower than 4 is shown. The Pearson's correlation coefficient is  $-0.57$ . The Spearman's correlation coefficient is  $-0.54$ . (TIFF)

**S6 Fig. The correlation between Potts' coupling parameters with experimental epistasis.** The pairwise epistasis between all RAMs in our library was compared with Potts' coupling parameters. The Spearman's correlation coefficient is  $-0.33$ . The  $p$  value for the Spearman's correlation coefficient is  $6.8 \times 10^{-3}$ . (TIFF)

**S7 Fig. Correlation between relative fitness and different statistical models.** (A, B & C) The correlation between relative fitness with (A, bin) binary (Ising) model inferred via ACE, (B, plm) the Potts model inferred via pseudo-likelihood maximization, or (C, potts) the Potts model inferred via ACE. (D) Spearman's correlation coefficients for different models. Mutants were classified according to their HD to wild-type. HD, hamming distance. (TIFF)

**S8 Fig. Structure insights on resistance associated mutations.** (A) Distribution of pairwise distance among resistance associated residues and other residues. The distance between the  $C-\alpha$  of two residues was shown. (B & C) Correlation between mutants' relative fitness and protein stability ( $\Delta\Delta G$ ).  $\Delta\Delta G$  is predicted by FoldX (B) or Rosetta (C). The correlation coefficients

were calculated for mutants with lower than 5 mutations.  $\rho$  stands for Spearman's correlation coefficient.

(TIFF)

**S1 Table. Relative fitness of all mutants in this research.**

(PDF)

**S2 Table. Information of protease inhibitor resistance associated mutations covered in the library.**

(PDF)

**S3 Table. Sequence of oligonucleotides used in this research.**

(PDF)

**S4 Table. Protein stability simulated by Rosetta or FoldX.**

(TSV)

## Acknowledgments

We thank UCLA/CFAR Virology Core Lab for doing p24 ELISA.

## Author Contributions

**Conceptualization:** Tian-hao Zhang, Lei Dai, James O. Lloyd-Smith.

**Data curation:** Tian-hao Zhang, Lei Dai, John P. Barton.

**Formal analysis:** Tian-hao Zhang, Lei Dai, John P. Barton.

**Funding acquisition:** Lei Dai, Ren Sun.

**Investigation:** Tian-hao Zhang, Lei Dai, John P. Barton, James O. Lloyd-Smith.

**Methodology:** Tian-hao Zhang, Lei Dai, John P. Barton, James O. Lloyd-Smith.

**Project administration:** Lei Dai, Ren Sun.

**Resources:** Ren Sun.

**Software:** Tian-hao Zhang, Lei Dai, John P. Barton.

**Supervision:** Lei Dai, James O. Lloyd-Smith, Ren Sun.

**Validation:** Tian-hao Zhang, Lei Dai.

**Visualization:** Tian-hao Zhang.

**Writing – original draft:** Tian-hao Zhang, Lei Dai, John P. Barton.

**Writing – review & editing:** Tian-hao Zhang, Lei Dai, John P. Barton, Yushen Du, Yuxiang Tan, Wenwen Pang, Arup K. Chakraborty, James O. Lloyd-Smith, Ren Sun.

## References

1. Palella FJ Jr, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *New England Journal of Medicine*. 1998; 338(13):853–860. <https://doi.org/10.1056/NEJM199803263381301>
2. Maggiolo F, Airoldi M, Kleinloog HD, Callegaro A, Ravasio V, Arici C, et al. Effect of adherence to HAART on virologic outcome and on the selection of resistance-conferring mutations in NNRTI-or PI-treated patients. *HIV Clinical Trials*. 2007; 8(5):282–292. <https://doi.org/10.1310/hct0805-282> PMID: 17956829



3. Shafer RW. Rationale and Uses of a Public HIV Drug-Resistance Database. *The Journal of Infectious Diseases*. 2006; 194(Supplement\_1):S51–S58. <https://doi.org/10.1086/505356> PMID: 16921473
4. Lin J, Nishino K, Roberts MC, Tolmasky M, Aminov RI, Zhang L. Mechanisms of antibiotic resistance. *Frontiers in Microbiology*. 2015; 6:34. <https://doi.org/10.3389/fmicb.2015.00034> PMID: 25699027
5. Lontok E, Harrington P, Howe A, Kieffer T, Lennerstrand J, Lenz O, et al. Hepatitis C virus drug resistance-associated substitutions: state of the art summary. *Hepatology*. 2015; 62(5):1623–1632. <https://doi.org/10.1002/hep.27934> PMID: 26095927
6. McKimm-Breschkin JL. Resistance of influenza viruses to neuraminidase inhibitors—a review. *Antiviral Research*. 2000; 47(1):1–17. [https://doi.org/10.1016/S0166-3542\(00\)00103-0](https://doi.org/10.1016/S0166-3542(00)00103-0) PMID: 10930642
7. Alexander BD, Perfect JR. Antifungal resistance trends towards the year 2000. *Drugs*. 1997; 54(5):657–678. <https://doi.org/10.2165/00003495-199754050-00002> PMID: 9360056
8. Kontoyiannis DP, Lewis RE. Antifungal drug resistance of pathogenic fungi. *The Lancet*. 2002; 359(9312):1135–1144. [https://doi.org/10.1016/S0140-6736\(02\)08162-X](https://doi.org/10.1016/S0140-6736(02)08162-X)
9. on Antimicrobial Resistance R. Tackling drug-resistant infections globally: final report and recommendations. *Review on Antimicrobial Resistance*; 2016.
10. Forum WE. *The Global Risks Report 2018, 13th Edition*. World Economic Forum; 2018.
11. Blair JM, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJ. Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*. 2015; 13(1):42. <https://doi.org/10.1038/nrmicro3380> PMID: 25435309
12. Altmann A, Beerenwinkel N, Sing T, Savenkov I, Däumer M, Kaiser R, et al. Improved prediction of response to antiretroviral combination therapy using the genetic barrier to drug resistance. *Antiviral Therapy*. 2007; 12(2):169. PMID: 17503659
13. Brenner BG, Wainberg MA. Clinical benefit of dolutegravir in HIV-1 management related to the high genetic barrier to drug resistance. *Virus Research*. 2017; 239:1–9. <https://doi.org/10.1016/j.virusres.2016.07.006> PMID: 27422477
14. Deforche K, Cozzi-Lepri A, Theys K, Clotet B, Camacho RJ, Kjaer J, et al. Modelled in vivo HIV fitness under drug selective pressure and estimated genetic barrier towards resistance are predictive for virological response. *Antiviral Therapy*. 2008; 13(3):399. PMID: 18572753
15. Devereux HL, Emery VC, Johnson MA, Loveday C. Replicative fitness in vivo of HIV-1 variants with multiple drug resistance-associated mutations. *Journal of Medical Virology*. 2001; 65(2):218–224. <https://doi.org/10.1002/jmv.2023> PMID: 11536226
16. Andersson DI, Levin BR. The biological cost of antibiotic resistance. *Current Opinion in Microbiology*. 1999; 2(5):489–493. [https://doi.org/10.1016/S1369-5274\(99\)00005-3](https://doi.org/10.1016/S1369-5274(99)00005-3) PMID: 10508723
17. Andersson DI, Hughes D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature Reviews Microbiology*. 2010; 8(4):260. <https://doi.org/10.1038/nrmicro2319> PMID: 20208551
18. Götte M. The distinct contributions of fitness and genetic barrier to the development of antiviral drug resistance. *Current Opinion in Virology*. 2012; 2(5):644–650. <https://doi.org/10.1016/j.coviro.2012.08.004> PMID: 22964133
19. Mesplède T, Quashie PK, Osman N, Han Y, Singhroy DN, Lie Y, et al. Viral fitness cost prevents HIV-1 from evading dolutegravir drug pressure. *Retrovirology*. 2013; 10(1):22. <https://doi.org/10.1186/1742-4690-10-22> PMID: 23432922
20. Sibley CH, Hyde JE, Sims PF, Plowe CV, Kublin JG, Mberu EK, et al. Pyrimethamine-sulfadoxine resistance in *Plasmodium falciparum*: what next? *Trends in Parasitology*. 2001; 17(12):570–571. [https://doi.org/10.1016/S1471-4922\(01\)02085-2](https://doi.org/10.1016/S1471-4922(01)02085-2)
21. Zhou J, Price AJ, Halambage UD, James LC, Aiken C. HIV-1 resistance to the capsid-targeting inhibitor PF74 results in altered dependence on host factors required for virus nuclear entry. *Journal of Virology*. 2015; 89(17):9068–9079. <https://doi.org/10.1128/JVI.00340-15> PMID: 26109731
22. Piana S, Carloni P, Rothlisberger U. Drug resistance in HIV-1 protease: flexibility-assisted mechanism of compensatory mutations. *Protein Science*. 2002; 11(10):2393–2402. <https://doi.org/10.1110/ps.0206702> PMID: 12237461
23. Deeks SG, Wrin T, Liegler T, Hoh R, Hayden M, Barbour JD, et al. Virologic and immunologic consequences of discontinuing combination antiretroviral-drug therapy in HIV-infected patients with detectable viremia. *New England Journal of Medicine*. 2001; 344(7):472–480. <https://doi.org/10.1056/NEJM200102153440702> PMID: 11172188
24. Frost SD, Nijhuis M, Schuurman R, Boucher CA, Brown AJL. Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection. *Journal of Virology*. 2000; 74(14):6262–6268. <https://doi.org/10.1128/jvi.74.14.6262-6268.2000> PMID: 10864635

25. Deeks SG, Hoh R, Neilands TB, Liegler T, Aweeka F, Petropoulos CJ, et al. Interruption of treatment with individual therapeutic drug classes in adults with multidrug-resistant HIV-1 infection. *Journal of Infectious Diseases*. 2005; 192(9):1537–1544. <https://doi.org/10.1086/496892> PMID: 16206068
26. Rosenbloom DI, Hill AL, Rabi SA, Siliciano RF, Nowak MA. Antiretroviral dynamics determines HIV evolution and predicts therapy outcome. *Nature Medicine*. 2012; 18(9):1378. <https://doi.org/10.1038/nm.2892> PMID: 22941277
27. Nijhuis M, Schuurman R, De Jong D, Erickson J, Gustchina E, Albert J, et al. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *Aids*. 1999; 13(17):2349–2359. <https://doi.org/10.1097/00002030-199912030-00006> PMID: 10597776
28. zur Wiesch PS, Engelstädter J, Bonhoeffer S. Compensation of fitness costs and reversibility of antibiotic resistance mutations. *Antimicrobial Agents and Chemotherapy*. 2010; 54(5):2085–2095. <https://doi.org/10.1128/AAC.01460-09>
29. Maisnier-Patin S, Andersson DI. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Research in Microbiology*. 2004; 155(5):360–369. <https://doi.org/10.1016/j.resmic.2004.01.019> PMID: 15207868
30. Bonhoeffer S, Chappey C, Parkin NT, Whitcomb JM, Petropoulos CJ. Evidence for positive epistasis in HIV-1. *Science*. 2004; 306(5701):1547–1550. <https://doi.org/10.1126/science.1101786> PMID: 15567861
31. Hinkley T, Martins J, Chappey C, Haddad M, Stawiski E, Whitcomb JM, et al. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nature Genetics*. 2011; 43(5):487. <https://doi.org/10.1038/ng.795> PMID: 21441930
32. Starr TN, Thornton JW. Epistasis in protein evolution. *Protein Science*. 2016; 25(7):1204–1218. <https://doi.org/10.1002/pro.2897> PMID: 26833806
33. Michalakis Y, Roze D. Epistasis in RNA viruses. *Science*. 2004; 306(5701):1492–1493. <https://doi.org/10.1126/science.1106677> PMID: 15567846
34. Parera M, Perez-Alvarez N, Clotet B, Martínez MA. Epistasis among deleterious mutations in the HIV-1 protease. *Journal of Molecular Biology*. 2009; 392(2):243–250. <https://doi.org/10.1016/j.jmb.2009.07.015> PMID: 19607838
35. Olson CA, Wu NC, Sun R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Current Biology*. 2014; 24(22):2643–2651. <https://doi.org/10.1016/j.cub.2014.09.072> PMID: 25455030
36. Bank C, Hietpas RT, Jensen JD, Bolon DN. A systematic survey of an intragenic epistatic landscape. *Molecular Biology and Evolution*. 2014; 32(1):229–238. <https://doi.org/10.1093/molbev/msu301> PMID: 25371431
37. Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, et al. Local fitness landscape of the green fluorescent protein. *Nature*. 2016; 533(7603):397. <https://doi.org/10.1038/nature17995> PMID: 27193686
38. Borrell S, Gagneux S. Strain diversity, epistasis and the evolution of drug resistance in *Mycobacterium tuberculosis*. *Clinical Microbiology and Infection*. 2011; 17(6):815–820. <https://doi.org/10.1111/j.1469-0691.2011.03556.x> PMID: 21682802
39. Trindade S, Sousa A, Xavier KB, Dionisio F, Ferreira MG, Gordo I. Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genetics*. 2009; 5(7):e1000578. <https://doi.org/10.1371/journal.pgen.1000578> PMID: 19629166
40. Silva RF, Mendonça SC, Carvalho LM, Reis AM, Gordo I, Trindade S, et al. Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. *PLoS Genetics*. 2011; 7(7):e1002181. <https://doi.org/10.1371/journal.pgen.1002181> PMID: 21829372
41. Yang WL, Kouyos RD, Böni J, Yerly S, Klimkait T, Aubert V, et al. Persistence of transmitted HIV-1 drug resistance mutations associated with fitness costs and viral genetic backgrounds. *PLoS Pathogens*. 2015; 11(3):e1004722. <https://doi.org/10.1371/journal.ppat.1004722> PMID: 25798934
42. Fragata I, Blanckaert A, Louro MAD, Liberles DA, Bank C. Evolution in the light of fitness landscape theory. *Trends in Ecology & Evolution*. 2018.
43. Cong Me, Heneine W, García-Lerma JG. The fitness cost of mutations associated with human immunodeficiency virus type 1 drug resistance is modulated by mutational interactions. *Journal of Virology*. 2007; 81(6):3037–3041. <https://doi.org/10.1128/JVI.02712-06> PMID: 17192300
44. Martínez-Picado J, Savara AV, Sutton L, Richard T. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *Journal of Virology*. 1999; 73(5):3744–3752. <https://doi.org/10.1128/JVI.73.5.3744-3752.1999> PMID: 10196268

45. Lv Z, Chu Y, Wang Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV/AIDS (Auckland, NZ)*. 2015; 7:95.
46. Strack PR, Frey MW, Rizzo CJ, Cordova B, George HJ, Meade R, et al. Apoptosis mediated by HIV protease is preceded by cleavage of Bcl-2. *Proceedings of the National Academy of Sciences*. 1996; 93(18):9571–9576. <https://doi.org/10.1073/pnas.93.18.9571>
47. Gougeon ML. Cell death and immunity: apoptosis as an HIV strategy to escape immune attack. *Nature Reviews Immunology*. 2003; 3(5):392. <https://doi.org/10.1038/nri1087> PMID: 12766761
48. Velazquez-Campoy A, Kiso Y, Freire E. The binding energetics of first-and second-generation HIV-1 protease inhibitors: implications for drug design. *Archives of Biochemistry and Biophysics*. 2001; 390(2):169–175. <https://doi.org/10.1006/abbi.2001.2333> PMID: 11396919
49. Harrigan PR, Hogg RS, Dong WW, Yip B, Wynhoven B, Woodward J, et al. Predictors of HIV drug-resistance mutations in a large antiretroviral-naive cohort initiating triple antiretroviral therapy. *The Journal of Infectious Diseases*. 2005; 191(3):339–347. <https://doi.org/10.1086/427192> PMID: 15633092
50. Lu Z. Second generation HIV protease inhibitors against resistant virus. *Expert opinion on drug discovery*. 2008; 3(7):775–786. <https://doi.org/10.1517/17460441.3.7.775> PMID: 23496220
51. Eshleman SH, Jones D, Galovich J, Paxinos EE, Petropoulos CJ, Jackson JB, et al. Phenotypic drug resistance patterns in subtype A HIV-1 clones with nonnucleoside reverse transcriptase resistance mutations. *AIDS Research & Human Retroviruses*. 2006; 22(3):289–293. <https://doi.org/10.1089/aid.2006.22.289>
52. De Meyer S, Vangeneugden T, Van Baelen B, De Paepe E, Van Marck H, Picchio G, et al. Resistance profile of darunavir: combined 24-week results from the POWER trials. *AIDS Research and Human Retroviruses*. 2008; 24(3):379–388. <https://doi.org/10.1089/aid.2007.0173> PMID: 18327986
53. Barbour JD, Wrin T, Grant RM, Martin JN, Segal MR, Petropoulos CJ, et al. Evolution of phenotypic drug susceptibility and viral replication capacity during long-term virologic failure of protease inhibitor therapy in human immunodeficiency virus-infected adults. *Journal of Virology*. 2002; 76(21):11104–11112. <https://doi.org/10.1128/JVI.76.21.11104-11112.2002> PMID: 12368352
54. Stoddart CA, Liegler TJ, Mammano F, Linquist-Stepps VD, Hayden MS, Deeks SG, et al. Impaired replication of protease inhibitor-resistant HIV-1 in human thymus. *Nature Medicine*. 2001; 7(6):712. <https://doi.org/10.1038/89090> PMID: 11385509
55. Bangsberg DR, Moss AR, Deeks SG. Paradoxes of adherence and drug resistance to HIV antiretroviral therapy. *Journal of Antimicrobial Chemotherapy*. 2004; 53(5):696–699. <https://doi.org/10.1093/jac/dkh162> PMID: 15044425
56. Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, Quintero J, et al. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors; 1995.
57. Dam E, Quercia R, Glass B, Descamps D, Launay O, Duval X, 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). <https://doi.org/10.1371/journal.ppat.1000345>
58. Chang MW, Torbett BE. Accessory mutations maintain stability in drug-resistant HIV-1 protease. *Journal of molecular biology*. 2011; 410(4):756–760. <https://doi.org/10.1016/j.jmb.2011.03.038> PMID: 21762813
59. Robinson LH, Myers RE, Snowden BW, Tisdale M, Blair ED. HIV type 1 protease cleavage site mutations and viral fitness: implications for drug susceptibility phenotyping assays. *AIDS research and human retroviruses*. 2000; 16(12):1149–1156. <https://doi.org/10.1089/088922200414992> PMID: 10954890
60. Flynn WF, Chang MW, Tan Z, Oliveira G, Yuan J, Okulicz JF, et al. Deep sequencing of protease inhibitor resistant HIV patient isolates reveals patterns of correlated mutations in Gag and protease. *PLoS computational biology*. 2015; 11(4). <https://doi.org/10.1371/journal.pcbi.1004249> PMID: 25894830
61. Rhee SY, Taylor J, Fessel WJ, Kaufman D, Towner W, Troia P, et al. HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrobial Agents and Chemotherapy*. 2010; 54(10):4253–4261. <https://doi.org/10.1128/AAC.00574-10> PMID: 20660676
62. Brenner BG, Routy JP, Petrella M, Moisi D, Oliveira M, Detorio M, et al. Persistence and fitness of multidrug-resistant human immunodeficiency virus type 1 acquired in primary infection. *Journal of Virology*. 2002; 76(4):1753–1761. <https://doi.org/10.1128/JVI.76.4.1753-1761.2002> PMID: 11799170
63. Johnson JA, Li JF, Wei X, Lipscomb J, Irlbeck D, Craig C, et al. Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naïve populations and associate with reduced treatment efficacy. *PLoS Medicine*. 2008; 5(7):e158. <https://doi.org/10.1371/journal.pmed.0050158> PMID: 18666824

64. Barbour JD, Hecht FM, Wrin T, Liegler TJ, Ramstead CA, Busch MP, et al. Persistence of primary drug resistance among recently HIV-1 infected adults. *Aids*. 2004; 18(12):1683–1689. <https://doi.org/10.1097/01.aids.0000131391.91468.ff> PMID: 15280779
65. Flynn WF, Haldane A, Torbett BE, Levy RM. Inference of epistatic effects leading to entrenchment and drug resistance in HIV-1 protease. *Molecular biology and evolution*. 2017; 34(6):1291–1306. <https://doi.org/10.1093/molbev/msx095> PMID: 28369521
66. Biswas A, Haldane A, Arnold E, Levy RM. Epistasis and entrenchment of drug resistance in HIV-1 subtype B. *eLife*. 2019; 8. <https://doi.org/10.7554/eLife.50524>
67. Qi H, Olson CA, Wu NC, Ke R, Loverdo C, Chu V, et al. A quantitative high-resolution genetic profile rapidly identifies sequence determinants of hepatitis C viral fitness and drug sensitivity. *PLoS Pathogens*. 2014; 10(4):e1004064. <https://doi.org/10.1371/journal.ppat.1004064> PMID: 24722365
68. Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *Elife*. 2016; 5:e16965. <https://doi.org/10.7554/eLife.16965> PMID: 27391790
69. Rhee SY, Liu T, Ravela J, Gonzales MJ, Shafer RW. Distribution of human immunodeficiency virus type 1 protease and reverse transcriptase mutation patterns in 4,183 persons undergoing genotypic resistance testing. *Antimicrobial Agents and Chemotherapy*. 2004; 48(8):3122–3126. <https://doi.org/10.1128/AAC.48.8.3122-3126.2004> PMID: 15273130
70. HIV Databases;. Available from: <http://www.hiv.lanl.gov/>.
71. Boucher JI, Whitfield TW, Dauphin A, Nachum G, Hollins C III, Zeldovich KB, et al. Constrained mutational sampling of amino acids in HIV-1 protease evolution. *Molecular biology and evolution*. 2019; 36(4):798–810. <https://doi.org/10.1093/molbev/msz022> PMID: 30721995
72. Parera M, Fernandez G, Clotet B, Martinez MA. HIV-1 protease catalytic efficiency effects caused by random single amino acid substitutions. *Molecular Biology and Evolution*. 2006; 24(2):382–387. <https://doi.org/10.1093/molbev/msl168> PMID: 17090696
73. Du Y, Zhang TH, Dai L, Zheng X, Gorin AM, Oishi J, et al. Effects of mutations on replicative fitness and major histocompatibility complex class I binding affinity are among the determinants underlying cytotoxic-T-lymphocyte escape of HIV-1 gag epitopes. *mBio*. 2017; 8(6):e01050–17. <https://doi.org/10.1128/mBio.01050-17> PMID: 29184023
74. Al-Mawsawi LQ, Wu NC, Olson CA, Shi VC, Qi H, Zheng X, et al. High-throughput profiling of point mutations across the HIV-1 genome. *Retrovirology*. 2014; 11(1):124. <https://doi.org/10.1186/s12977-014-0124-6> PMID: 25522661
75. Sanjuan R, Moya A, Elena SF. The contribution of epistasis to the architecture of fitness in an RNA virus. *Proceedings of the National Academy of Sciences*. 2004; 101(43):15376–15379. <https://doi.org/10.1073/pnas.0404125101>
76. Silander OK, Tenaillon O, Chao L. Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biology*. 2007; 5(4):e94. <https://doi.org/10.1371/journal.pbio.0050094> PMID: 17407380
77. Dai L, Du Y, Qi H, Wu NC, Lloyd-Smith JO, Sun R. Quantifying the evolutionary potential and constraints of a drug-targeted viral protein. *bioRxiv*. 2016; p. 078428.
78. Parera M, Martinez MA. Strong epistatic interactions within a single protein. *Molecular Biology and Evolution*. 2014; 31(6):1546–1553. <https://doi.org/10.1093/molbev/msu113> PMID: 24682281
79. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*. 2011; 332(6034):1193–1196. <https://doi.org/10.1126/science.1203801> PMID: 21636772
80. Elena SF. Little evidence for synergism among deleterious mutations in a nonsegmented RNA virus. *Journal of Molecular Evolution*. 1999; 49(5):703–707. <https://doi.org/10.1007/PL00000082> PMID: 10552052
81. Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, Carlson JM, et al. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. *Journal of Experimental Medicine*. 2008; 205(5):1009–1017. <https://doi.org/10.1084/jem.20072457> PMID: 18426987
82. Sierra S, Dávila M, Lowenstein PR, Domingo E. Response of foot-and-mouth disease virus to increased mutagenesis: influence of viral load and fitness in loss of infectivity. *Journal of Virology*. 2000; 74(18):8316–8323. <https://doi.org/10.1128/jvi.74.18.8316-8323.2000> PMID: 10954530
83. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature*. 2003; 424(6944):94. <https://doi.org/10.1038/nature01707> PMID: 12808465

84. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, et al. DNA deamination mediates innate immunity to retroviral infection. *Cell*. 2003; 113(6):803–809. [https://doi.org/10.1016/S0092-8674\(03\)00423-9](https://doi.org/10.1016/S0092-8674(03)00423-9) PMID: 12809610
85. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*. 2003; 424(6944):99. <https://doi.org/10.1038/nature01709> PMID: 12808466
86. Crotty S, Cameron CE, Andino R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proceedings of the National Academy of Sciences*. 2001; 98(12):6895–6900. <https://doi.org/10.1073/pnas.111085598>
87. Ferguson AL, Mann JK, Omarjee S, Ndung'u T, Walker BD, Chakraborty AK. Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. *Immunity*. 2013; 38(3):606–617. <https://doi.org/10.1016/j.immuni.2012.11.022> PMID: 23521886
88. Mann JK, Barton JP, Ferguson AL, Omarjee S, Walker BD, Chakraborty A, et al. The Fitness Landscape of HIV-1 Gag: Advanced Modeling Approaches and Validation of Model Predictions by In Vitro Testing. *PLOS Computational Biology*. 2014; 10(8):1–11. <https://doi.org/10.1371/journal.pcbi.1003776>
89. Butler TC, Barton JP, Kardar M, Chakraborty AK. Identification of drug resistance mutations in HIV from constraints on natural evolution. *Physical Review E*. 2016; 93(2):022412. <https://doi.org/10.1103/PhysRevE.93.022412> PMID: 26986367
90. Louie RH, Kaczorowski KJ, Barton JP, Chakraborty AK, McKay MR. Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies. *Proceedings of the National Academy of Sciences*. 2018; 115(4):E564–E573.
91. Barton JP, De Leonardis E, Coucke A, Cocco S. ACE: adaptive cluster expansion for maximum entropy graphical model inference. *Bioinformatics*. 2016; 32(20):3089–3097. <https://doi.org/10.1093/bioinformatics/btw328> PMID: 27329863
92. Solis M, Nakhaei P, Jalalirad M, Lacoste J, Douville R, Arguello M, et al. RIG-I-mediated antiviral signaling is inhibited in HIV-1 infection by a protease-mediated sequestration of RIG-I. *Journal of Virology*. 2011; 85(3):1224–1236. <https://doi.org/10.1128/JVI.01635-10> PMID: 21084468
93. Shah P, McCandlish DM, Plotkin JB. Contingency and entrenchment in protein evolution under purifying selection. *Proceedings of the National Academy of Sciences*. 2015; 112(25):E3226–E3235. <https://doi.org/10.1073/pnas.1412933112>
94. Draghi JA, Plotkin JB. Selection biases the prevalence and type of epistasis along adaptive trajectories. *Evolution*. 2013; 67(11):3120–3131. <https://doi.org/10.1111/evo.12192> PMID: 24151997
95. Kitayimbwa JM, Mugisha JYT, Saenz RA. Estimation of the HIV-1 backward mutation rate from transmitted drug-resistant strains. *Theoretical Population Biology*. 2016; 112:33–42. <https://doi.org/10.1016/j.tpb.2016.08.001> PMID: 27553875
96. Wensing AJ, van de Vijver D, Angarano G, Åsjö B, Balotta C, Boeri E, et al. Prevalence of Drug-Resistant HIV-1 Variants in Untreated Individuals in Europe: Implications for Clinical Management. *The Journal of Infectious Diseases*. 2005; 192(6):958–966. <https://doi.org/10.1086/432916> PMID: 16107947
97. Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science*. 1988; 242(4882):1171–1173. PMID: 2460925
98. Cuevas JM, Geller R, Garijo R, López-Aldeguer J, Sanjuán R. Extremely high mutation rate of HIV-1 in vivo. *PLoS Biology*. 2015; 13(9):e1002251. <https://doi.org/10.1371/journal.pbio.1002251> PMID: 26375597
99. Han TX, Xu XY, Zhang MJ, Peng X, Du LL. Global fitness profiling of fission yeast deletion strains by barcode sequencing. *Genome biology*. 2010; 11(6):R60. <https://doi.org/10.1186/gb-2010-11-6-r60> PMID: 20537132
100. Van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature methods*. 2009; 6(10):767. <https://doi.org/10.1038/nmeth.1377> PMID: 19767758
101. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*. 1996; 271(5255):1582–1586. <https://doi.org/10.1126/science.271.5255.1582> PMID: 8599114
102. Fernandes JD, Faust TB, Strauli NB, Smith C, Crosby DC, Nakamura RL, et al. Functional segregation of overlapping genes in HIV. *Cell*. 2016; 167(7):1762–1773. <https://doi.org/10.1016/j.cell.2016.11.031> PMID: 27984726

103. Weinberger ED. Fourier and Taylor series on fitness landscapes. *Biological cybernetics*. 1991; 65(5):321–330. <https://doi.org/10.1007/BF00216965>
104. Uguzzoni G, Lovis SJ, Oteri F, Schug A, Szurmant H, Weigt M. Large-scale identification of coevolution signals across homo-oligomeric protein interfaces by direct coupling analysis. *Proceedings of the National Academy of Sciences*. 2017; 114(13):E2662–E2671. <https://doi.org/10.1073/pnas.1615068114>
105. Levy RM, Haldane A, Flynn WF. Potts Hamiltonian models of protein co-variation, free energy landscapes, and evolutionary fitness. *Current opinion in structural biology*. 2017; 43:55–62. <https://doi.org/10.1016/j.sbi.2016.11.004> PMID: 27870991
106. Chen L, Perlina A, Lee CJ. Positive selection detection in 40,000 human immunodeficiency virus (HIV) type 1 sequences automatically identifies drug resistance and positive fitness mutations in HIV protease and reverse transcriptase. *Journal of Virology*. 2004; 78(7):3722–3732. <https://doi.org/10.1128/JVI.78.7.3722-3732.2004> PMID: 15016892
107. He X, Qian W, Wang Z, Li Y, Zhang J. Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. *Nature genetics*. 2010; 42(3):272. <https://doi.org/10.1038/ng.524> PMID: 20101242
108. Yerly S, Kaiser L, Race E, Bru JP, Clavel F, Perrin L. Transmission of antiretroviral-drug-resistant HIV-1 variants. *The Lancet*. 1999; 354(9180):729–733. [https://doi.org/10.1016/S0140-6736\(98\)12262-6](https://doi.org/10.1016/S0140-6736(98)12262-6).
109. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature*. 2006; 444(7121):929. <https://doi.org/10.1038/nature05385> PMID: 17122770
110. Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins: Structure, Function, and Bioinformatics*. 2011; 79(3):830–838.
111. Barton JP, Goonetilleke N, Butler TC, Walker BD, McMichael AJ, Chakraborty AK. Relative rate and location of intra-host HIV evolution to evade cellular immunity are predictable. *Nature Communications*. 2016; 7:11660. <https://doi.org/10.1038/ncomms11660> PMID: 27212475