UCLA
UCLA Previously Published Works

Title

Quantifying the Evolutionary Constraints and Potential of Hepatitis C Virus NS5A Protein.

Permalink

https://escholarship.org/uc/item/96r5x26v

Journal

mSystems, 6(2)

ISSN

2379-5077

Authors

Dai, Lei

Du, Yushen

Qi, Hangfei

et al.

Publication Date

2021-04-13

DOI

10.1128/mSystems.01111-20

Copyright Information

Peer reviewed

Quantifying the Evolutionary Constraints and Potential of Hepatitis C Virus NS5A Protein

Lei Dai,[a,d] Yushen Du,[b,c] Hangfei Qi,[c] Christian D. Huber,[d] Dongdong Chen,[b] Tian-Hao Zhang,[c] Nicholas C. Wu,[e] Ergang Wang,[c] James O. Lloyd-Smith,[d] Ren Sun[b]

[a]CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[b]Cancer Institute, ZJU-UCLA Joint Center for Medical Education and Research, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
[c]Department of Molecular and Medical Pharmacology, University of California Los Angeles, Los Angeles, California, USA
[d]Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, California, USA
[e]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Lei Dai and Yushen Du contributed equally to this work. Author order was determined alphabetically.

ABSTRACT RNA viruses, such as hepatitis C virus (HCV), influenza virus, and SARS-CoV-2, are notorious for their ability to evolve rapidly under selection in novel environments. It is known that the high mutation rate of RNA viruses can generate huge genetic diversity to facilitate viral adaptation. However, less attention has been paid to the underlying fitness landscape that represents the selection forces on viral genomes, especially under different selection conditions. Here, we systematically quantified the distribution of fitness effects of about 1,600 single amino acid substitutions in the drug-targeted region of NS5A protein of HCV. We found that the majority of nonsynonymous substitutions incur large fitness costs, suggesting that NS5A protein is highly optimized. The replication fitness of viruses is correlated with the pattern of sequence conservation in nature, and viral evolution is constrained by the need to maintain protein stability. We characterized the adaptive potential of HCV by subjecting the mutant viruses to selection by the antiviral drug daclatasvir at multiple concentrations. Both the relative fitness values and the number of beneficial mutations were found to increase with the increasing concentrations of daclatasvir. The changes in the spectrum of beneficial mutations in NS5A protein can be explained by a pharmacodynamics model describing viral fitness as a function of drug concentration. Overall, our results show that the distribution of fitness effects of mutations is modulated by both the constraints on the biophysical properties of proteins (i.e., selection pressure for protein stability) and the level of environmental stress (i.e., selection pressure for drug resistance).

IMPORTANCE Many viruses adapt rapidly to novel selection pressures, such as antiviral drugs. Understanding how pathogens evolve under drug selection is critical for the success of antiviral therapy against human pathogens. By combining deep sequencing with selection experiments in cell culture, we have quantified the distribution of fitness effects of mutations in hepatitis C virus (HCV) NS5A protein. Our results indicate that the majority of single amino acid substitutions in NS5A protein incur large fitness costs. Simulation of protein stability suggests viral evolution is constrained by the need to maintain protein stability. By subjecting the mutant viruses to selection under an antiviral drug, we find that the adaptive potential of viral proteins in a novel environment is modulated by the level of environmental stress, which can be explained by a pharmacodynamics model. Our comprehensive characterization of the fitness landscapes of NS5A can potentially guide the design of effective strategies to limit viral evolution.

In our evolutionary battles with microbial pathogens, RNA viruses are among the most formidable foes. HIV-1 and hepatitis C virus (HCV) acquire drug resistance in patients under antiviral therapies. Influenza virus, Ebola virus, and SARS-CoV2 cross the species barrier to infect human hosts. Understanding the evolution of RNA viruses is therefore of paramount importance for developing antivirals and vaccines and assessing the risk of future emergence events (1–3). Comprehensive characterization of viral fitness landscapes, and the principles underpinning them, will provide us with a map of evolutionary pathways accessible to viruses and guide our design of effective strategies to limit antiviral resistance, immune escape, and cross-species transmission (4–6).

Although the concept of fitness landscapes has been around for a long time (7), their properties in real biological systems are still under active investigation. Previous empirical studies of fitness landscapes have been constrained by limited sampling of sequence space. In a typical study, mutants are generated by site-directed mutagenesis and assayed for growth rate individually. We and others have recently utilized a high-throughput technique, often referred to as "deep mutational scanning" or "quantitative high-resolution genetics," to profile the fitness effect of mutations by integrating deep sequencing with selection experiments *in vitro* or *in vivo* (8–14). This application of next-generation sequencing has raised the exciting prospect of large-scale fitness measurements (15–18) and a revolution in our understanding of molecular evolution (19).

The distribution of fitness effects (DFE) of mutations is a fundamental entity in genetics and reveals the local structure of a fitness landscape (12, 20–29). Deleterious mutations are usually abundant and impose severe constraints on the accessibility of fitness landscapes. In contrast, beneficial mutations are rare and provide the raw materials of adaptation. Quantifying the DFE of viruses is crucial for understanding how these pathogens evolve to acquire drug resistance and surmount other evolutionary challenges.

Previously, most empirical studies of the DFE have been performed in a single, static environment (20, 21). A central challenge is to characterize the DFE, and its determinants, in fluctuating or heterogeneous environments where evolution typically occurs (e.g., fluctuating drug concentrations or a gradient across space). More attention has been paid to this area recently. For bacteria, the fitness effects of mutations at different drug concentrations, or under physical and chemical stress, have been studied (30–32). One study has demonstrated that drug concentration modulates the shape of the DFE and determines the evolvability under new environments (33). In another study, the implications of differing drug concentrations on the adaptive landscape have been examined in the context of resistance evolution (34). For viruses, the fitness effects of mutations have been measured across different hosts (35–37). The shape of the DFE of viruses was inferred from experimentally passaged populations (38) and from patient data (39), but not quantified systematically. Combining quantitative high-resolution genetics with different selection conditions will provide a more comprehensive investigation of the DFE under varying levels of positive selection.

In this study, we profile the DFE of ~1,600 single amino acid substitutions in a drug-targeted viral protein by coupling a selection experiment of a mutant library and deep sequencing. We show that the replication fitness of virus mutants is correlated with the pattern of conservation in patient-derived HCV sequences, suggesting that amino acid sites with high fitness costs are often highly conserved. Combined with simulations of protein stability, we confirm that protein stability is a major determinant of the deleterious effect of mutants and imposes a strong constraint to viral evolution. Furthermore, we examine the changes in DFE under varying levels of environmental stress by tuning the concentration of an antiviral drug. The distribution of beneficial

fitness effects of mutations shifts with the increase of environmental stress, in accordance with theoretical predictions (40).

RESULTS

Profiling the fitness landscape of the drug-interacting domain of HCV NS5A protein. The system used in our study is hepatitis C virus (HCV; genotype 2a. J6/JFH1 chimera), a positive-sense single-stranded RNA virus with a genome of ~9.6 kb. HCV has been studied extensively in the past 2 decades in patients and in the laboratory and provides an excellent model system to study viral evolution. Previously, we constructed a saturation mutagenesis library with all single amino acid substitutions in domain IA (amino acids 18 to 103) of HCV NS5A protein (11). This domain is the target of several directly acting antiviral drugs, including the potent HCV NS5A inhibitor daclatasvir (DCV) (41). Here, we utilized the same plasmid library to further study the DFE of mutations and examine its adaptive potential under various drug selection pressures through a series of new selection experiments. We observed 2,520 nonsynonymous mutations in the plasmid library, as well as 105 synonymous mutations. After transfection to reconstitute mutant viruses, we performed selection in an HCV cell culture system (42, 43). The relative fitness (RF) of a mutant virus was calculated based on the changes in frequency of the mutant virus and the wild-type virus after one round of selection in cell culture (Fig. S1A). In our selection experiment, we grew 5 small sublibraries (~500 mutants each) separately to reduce the noise in fitness measurements (see Materials and Methods). The fitness data reported in this study are highly correlated with the previously reported independent experiment (Fig. S1B and C) (11).

Our experiment provides a comprehensive profiling of the fitness effect of single amino acid substitutions (1,565 out of 1,634 possible substitutions, after filtering out low-frequency mutants in the plasmid library). We grouped together nonsynonymous mutations leading to the same amino acid substitution (Data set S1). As expected, the fitness effects of synonymous mutations were nearly neutral, while most nonsynonymous mutations were deleterious (Fig. 1A and B). The RF of all mutations is shown with the heatmap in Fig. 1C. We found that the majority of single amino acid mutations had fitness costs, and more than half of them were found to be significantly deleterious, or "lethal" (shown at −8 for Fig. 1A; Materials and Methods). The fraction of lethal mutations is 59.5% (932/1,565) for single amino acid substitutions and 95.1% (77/81) for nonsense mutations with known RF. As NS5A is essential for viral replication, the nonsense mutations should be detrimental. The four nonsense mutations (4/81) that were not identified to be lethal in our profile may due to an experimental artifact that is inevitable in high-throughput genetic screening studies (14, 44). The low tolerance of nonsynonymous mutations in HCV NS5A, which is an essential protein for viral replication, is consistent with previous small-scale mutagenesis studies of RNA viruses (45). Our data support the view that RNA viruses are very sensitive to the effect of deleterious mutations, possibly due to the compactness of their genomes (46, 47).

Using the distribution of fitness effects of synonymous mutations as a benchmark for neutrality, we identified that only 2.4% (37/1,565) of single amino acid mutations are beneficial (Materials and Methods). The estimated fraction of beneficial mutations is consistent with previous small-scale mutagenesis studies of viruses, including bacteriophages, vesicular stomatitis virus, etc. (20, 45, 48, 49). Our results indicate that HCV NS5A protein is under strong purifying selection, suggesting that viral proteins are highly optimized in their natural conditions.

Deleterious mutations as evolutionary constraints. Mutations that severely reduce replication fitness impose constraints on the evolution of viruses and are less likely to contribute to adaptation through gain of function. We analyzed the sequence diversity of HCV sequences identified in patients from the HCV sequence database of Los Alamos National Lab and the European HCV Database (euHCVdb) (Materials and Methods). To avoid biases toward specific genotypes, we included ~2,600 sequences from all HCV genotypes in analysis. The sequence diversity at each site was highly correlated with the replication fitness (the mean fitness of observed mutants at each site)
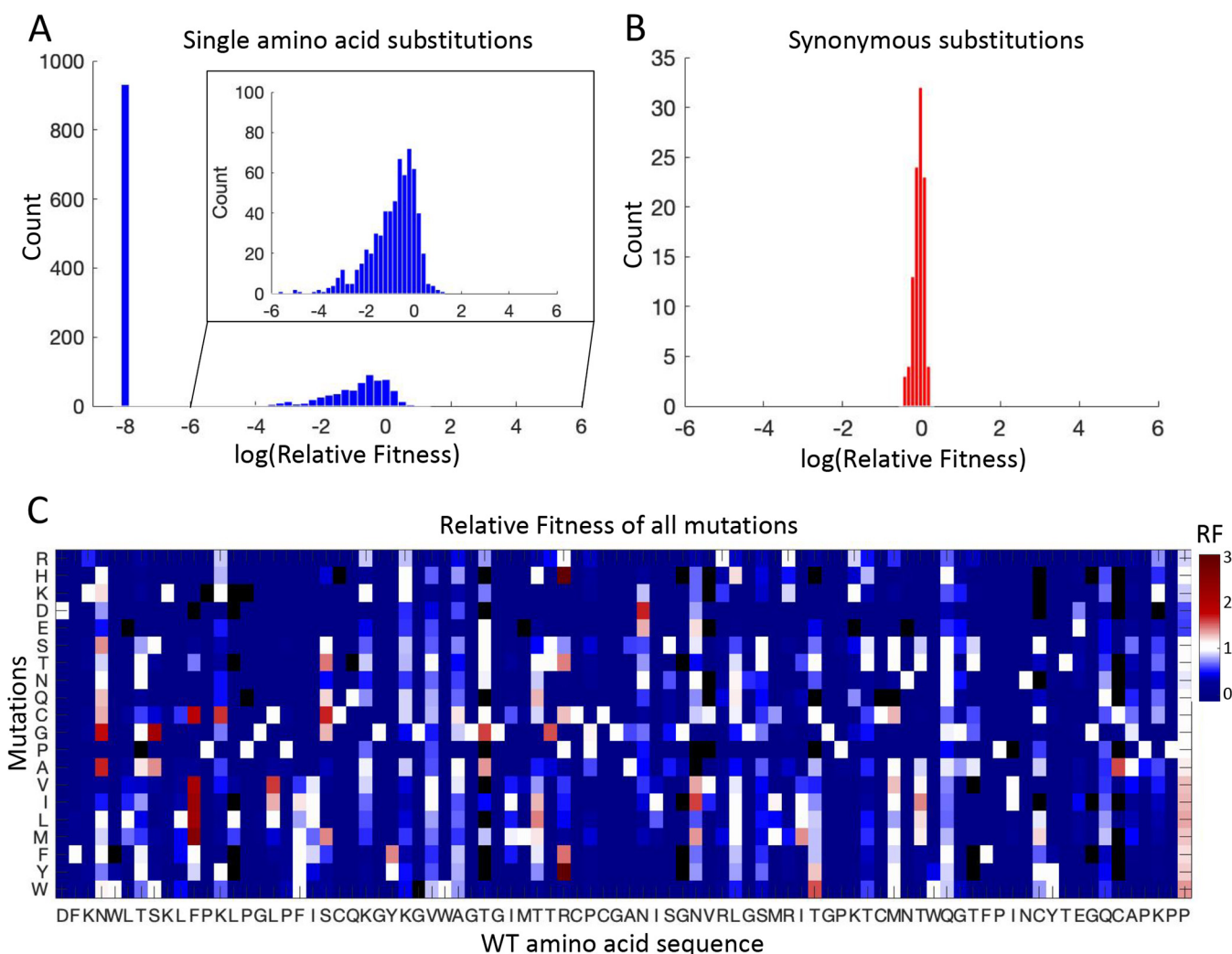
FIG 1 Distribution of fitness effects (DFE) of single amino acid substitutions in domain IA of HCV NS5A protein without drug selection. (A) DFE of single amino acid substitutions. The x axis shows the log transformed relative fitness. Lethal mutations are shown at −8. A zoom-in view shows the nonlethal portion of substitutions. (B) DFE of synonymous substitutions, which is centered at 0 for log transformed relative fitness. (C) The Heatmap shows the relative fitness of all mutations. Lethal mutations are shown in dark blue (relative fitness = 0). Mutations that were filtered due to low frequency in the plasmid library (unknown relative fitness) are shown in black.

measured in our study (Fig. 2A; Spearman's $\rho = 0.81$, $P = 1.75 \times 10^{-21}$). The amino acid sites with high fitness costs were often highly conserved, such as residues 32, 33, 39, 57, 59, 60, 76, 88, 91, 94, et al. We also calculated the frequency of natural occurrence for all mutations and noticed that the majority of mutations with a frequency of >0.1 were relatively neutral in replication fitness (Fig. 2B). Conversely, mutations that do not occur in nature (frequency of 0) may not be lethal for replication fitness, pointing to the potential limited sampling of natural sequences.

To understand the biophysical basis of mutational effects (50), we took advantage of the available structural information. The crystal structure of NS5A domain I is available, excluding the amphipathic helix at the N terminus (51, 52). We calculated the relative solvent accessibility of all residues and found that the fitness effects of deleterious mutations at buried sites (i.e., with lower solvent accessibility) were more pronounced than those at surface-exposed sites (Fig. 2C, Fig. S2A) (53). Residues with average fitness of <0.2 showed a lower relative solvent accessibility (Fig. S2B). Moreover, we performed simulations of protein stability for individual mutants using PyRosetta (Materials and Methods) (54, 55). A mutation with ΔΔG of >0, i.e., shifting the free energy difference to favor the unfolded state, is expected to destabilize the
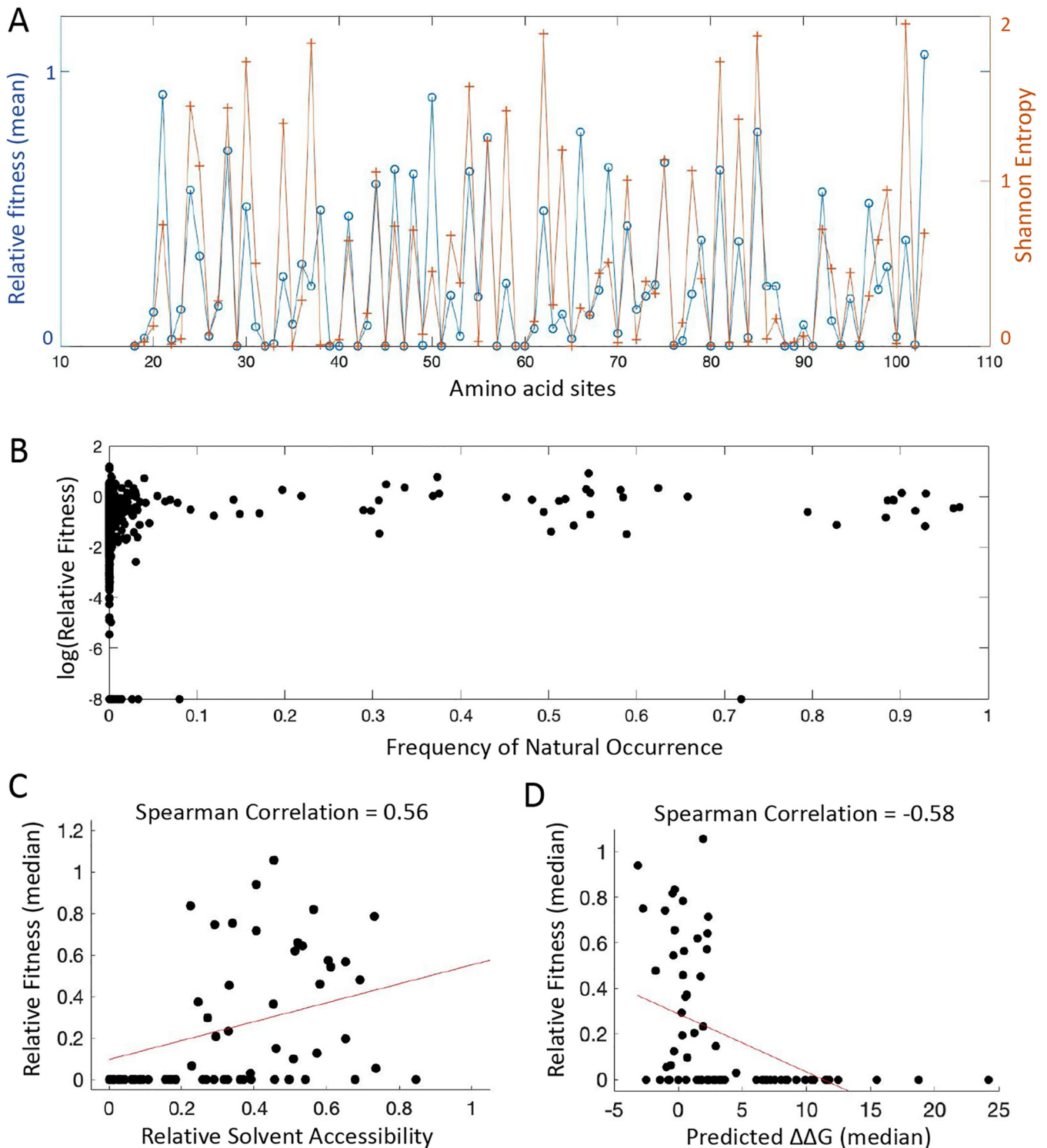
FIG 2 Mutations with deleterious fitness effects reveal constraints of protein evolution. (A) The pattern of sequence conservation observed in patient sequences is highly correlated with the replication fitness measured in cell culture. The blue line shows the average relative fitness for each residue considering all mutations; the orange line shows the Shannon entropy. (B) The Scatterplot shows the frequency of natural occurrence and the log transformed relative fitness for individual mutants. (C) Mutations at amino acid sites with lower solvent accessibility tend to incur larger fitness costs. The relative solvent accessibility for each residue is significantly correlated with median relative fitness (Spearman's $\rho = 0.56$, $P = 3.4 \times 10^{-7}$). (D) Mutations at amino acid sites with larger effects on destabilizing protein stability (predicted $\Delta\Delta G > 0$) tend to reduce the viral replication fitness. Changes in folding free energy $\Delta\Delta G$ (Rosetta energy unit) of the NS5A monomer were predicted by PyRosetta (PDB: 3FQM). The median predicted $\Delta\Delta G$ at each amino acid site is shown. The median fitness of observed mutants at each amino acid site is shown. In panels C and D, red lines represent the fits by linear regression and are only used to guide the eye.

protein. Three protein structures were utilized. First, we performed protein stability prediction based on the 3FQM structure, which has the closest reference sequence to the NS5A sequence we used in our experiments but still differs by 20 amino acid substitutions. At the residue level, we found mutations that decreased protein stability (median predicted $\Delta\Delta G$ change for each residue) led to reduced viral fitness (the median and mean fitness of observed mutants at each site, $P = 7.7 \times 10^{-8}$ and $P = 2.3 \times 10^{-6}$, respectively; Fig. 2D, Fig. S2C). For example, mutations at a stretch of highly conserved residues (F88 to N91) that run through the core of NS5A protein tended to destabilize the protein and significantly reduced the viral fitness. Mutations that increase $\Delta\Delta G$ beyond a threshold ($\sim$5 Rosetta energy units) were mostly lethal. This is consistent with the threshold robustness model, which predicts that proteins become unfolded after using up the stability margin (15, 56, 57). The negative correlation between protein stability and viral fitness was confirmed by predicting $\Delta\Delta G$ using a different Protein Data Bank (PDB) model (4CL1; Fig. S2D to F). Although the sequence of 3FQM and 4CL1 has 29 amino acid differences (83.4% identity), the protein structures are highly similar to each other, and the predicted $\Delta\Delta G$s are highly consistent among residues (Fig. S2D and E). Furthermore, we performed homology structural modeling using SWISS-MODEL (58) and predicted the protein structure based on the NS5A sequence we used in the experiments (Fig. S2G to I). With the same amino acid sequence, the predicted structure allowed us to compare $\Delta\Delta G$ and viral fitness for each individual mutant. Consistent with the result at the residue level (Fig. S2H), the negative correlation and the protein stability threshold exist for all the mutants (Fig. S2I). We also note that mutations can be deleterious because they impair protein function rather than destabilize the protein, so the correlation between protein stability and fitness is not expected to be perfect. The level of correlation between $\Delta\Delta G$ and fitness that we observed is similar to that from previous studies of other proteins (13, 30, 59).

Adaptive potential as a function of environmental stress. Beneficial mutations are the raw materials of protein evolution (20). We aimed to study the role of environmental stress in modulating the adaptive potential of drug-targeted viral proteins. In an independent study (11), the mutant library of HCV NS5A protein was selected under a single drug concentration ([DCV] = 20 pM) to profile the effects of mutations on drug resistance. In this study, we selected the mutant library at 10, 40, and 100 pM of DCV. The drug concentrations were chosen based on the *in vitro* 50% inhibitory concentration (IC_{50}) of wild-type HCV virus ($\sim$20 pM) to represent different levels of environmental stress (mild, intermediate, and strong).

By tuning the concentration of DCV, we observed a change in the DFE (Table S1), particularly of beneficial mutations (Fig. 3A). At higher drug concentrations, we observed an increase in the total number of beneficial mutations (Fig. 3B, Table S2). Furthermore, the cumulative distribution function (CDF) of beneficial mutations also shows an increase in the median and maximum relative fitness (Fig. 3C). We further tested whether the shape of this distribution changed under drug selection. Previous empirical studies supported the hypothesis that the DFE of beneficial mutations is exponential or bounded on the right (40, 45, 48, 60–69). Following a maximum likelihood approach, we fit the DFE of beneficial mutations to the generalized Pareto distribution (Fig. S3; Materials and Methods). The fitted distribution is described by two parameters, a scale parameter (τ) and a shape parameter (κ) that determines the behavior of the distribution's tail. Using a likelihood-ratio test (70), we found that our data are consistent with the null hypothesis that the DFE of beneficial mutations is exponential ($\kappa = 0$) (Table S2).

Furthermore, we used a maximum-likelihood approach to fit a displaced-gamma distribution to the DFE to estimate the distance to the phenotypic optimum in Fisher's geometric model (FGM) (71, 72) (Fig. S4). The displaced-gamma distribution has the shape of a negative gamma distribution, shifted by a parameter s_0 that indicates the distance of the initial genotype (i.e., wild type) to the optimum (Materials and
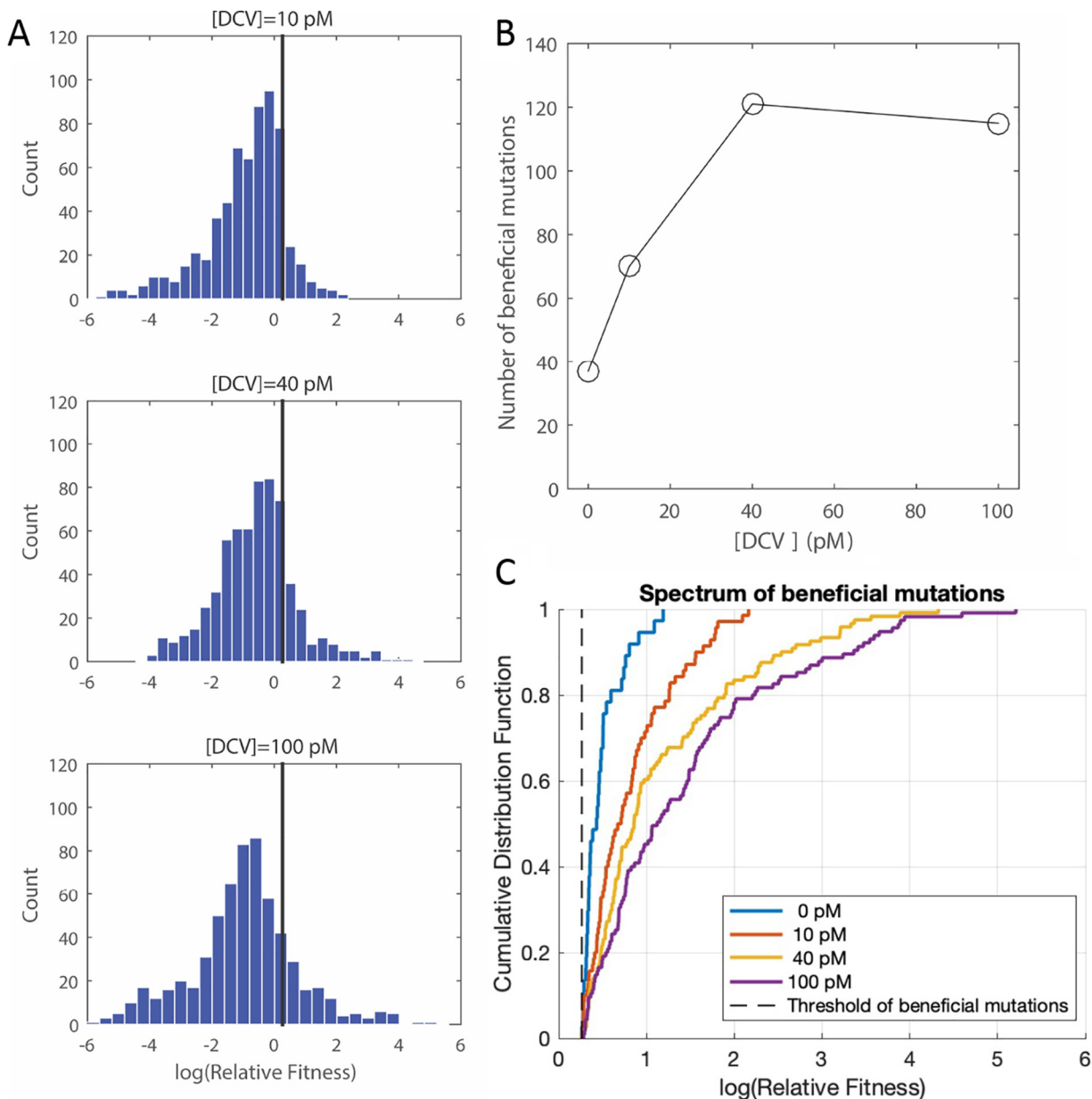
FIG 3 The spectrum of beneficial mutations changes under increasing environmental stress imposed by the antiviral drug daclatasvir. (A) DFE of single amino acid substitutions in domain IA of HCV NS5A protein under increasing environmental stress by daclatasvir. The black line indicates the threshold used for classifying beneficial mutations (Materials and Methods). (B) The number of beneficial mutations as a function of environmental stress imposed by daclatasvir. (C) The cumulative distribution function (CDF) of the fitness effect of beneficial mutations. The dashed black line indicates the threshold used for classifying beneficial mutations.

Methods). Estimated distances to the phenotypic optimum under different conditions are summarized in Table S3. In accordance with theoretical expectations, we found that the distance to the phenotypic optimum increased as the level of environmental stress increased (i.e., increasing drug concentration).

A pharmacodynamics model explains the shift of DFE with increased drug concentration. Our results show that the adaptive potential of proteins is modulated by the strength of environmental stress. To explain the changing spectra of beneficial mutations upon drug treatment, we employed a pharmacodynamics model describing viral fitness as a function of drug concentration (i.e., phenotype-fitness mapping) (Fig. 4A).
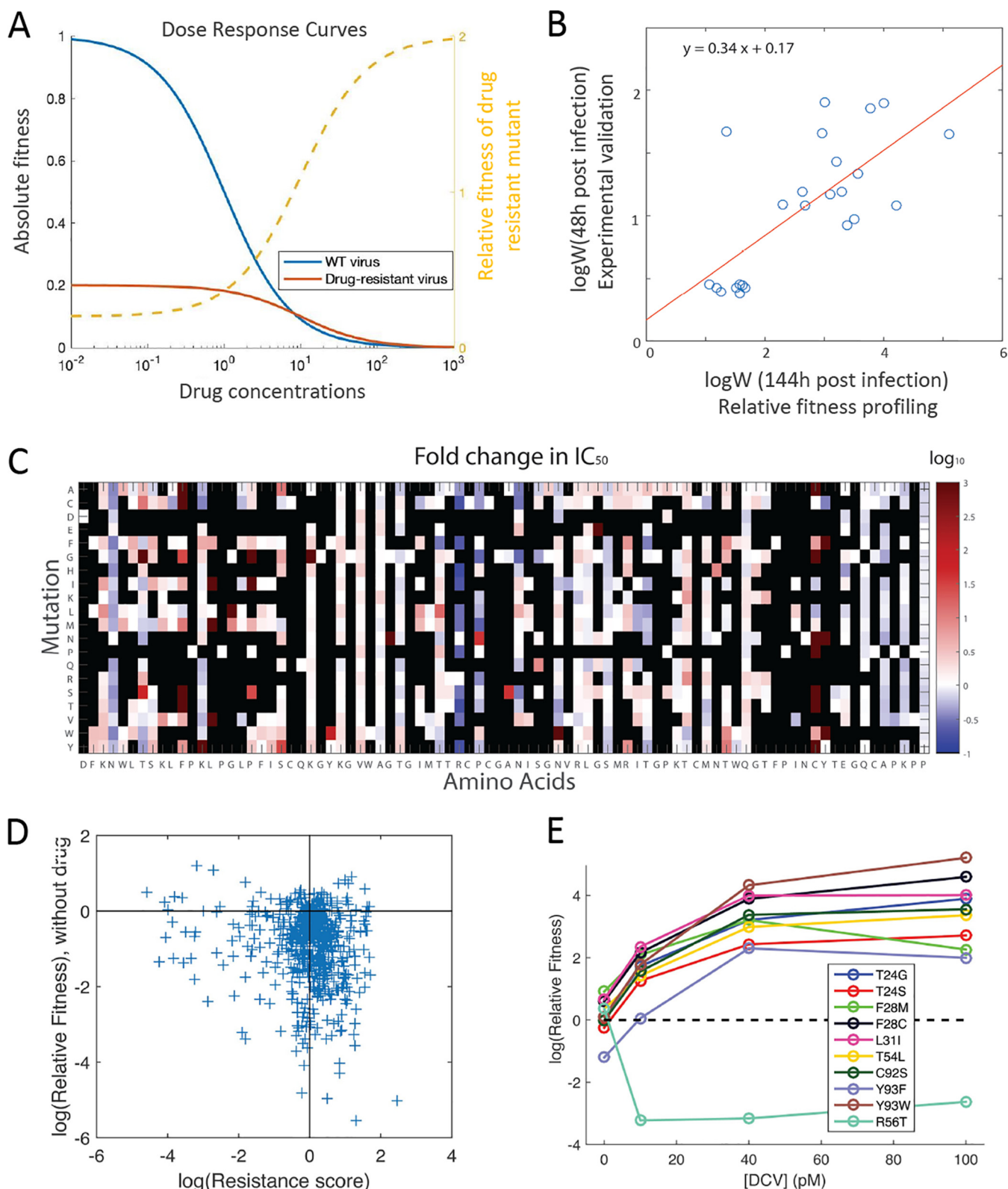
FIG 4 The adaptive potential under drug selection is determined by the effects of mutations on replication fitness and drug resistance. (A) Hypothetical dose response curves of the wild-type (WT) virus and a drug-resistant mutant virus. The blue and red lines represent the absolute fitness of WT virus and drug-resistant mutant virus, respectively. The yellow dashed line represents the relative fitness of the mutant virus to WT virus. The absolute fitness decreases with drug concentration [*drug*] following $f = f_0 \frac{IC_{50}}{IC_{50} + [drug]}$, where f_0 is the fitness without drug selection and IC_{50} is the half inhibitory concentration. Compared to the wild-type virus, the hypothetical drug-resistant mutant carries a fitness cost (smaller f_0) but is less sensitive to drug inhibition (larger IC_{50}). The relative fitness of the drug-resistant mutant is expected to increase with drug concentration. When drug concentration $\rightarrow \infty$,

$$f = f_0 \frac{IC_{50}}{IC_{50} + [drug]}$$

where f_0 is the fitness without drug selection and IC_{50} is the half inhibitory concentration. The absolute fitness f decreases with drug concentration ([drug]). In this model, the fitness of each mutant under drug selection is contributed by two traits, the fitness without drug selection (f_0) and the effect on drug resistance (IC_{50}). We define the drug resistance score (W) of a mutant as the ratio of the relative fitness under drug selection to that without drug selection.

$$W_{mut,[drug]} = \left(\frac{f_{mut}}{f_{wt}} \right) / \frac{f_{0\,mut}}{f_{0\,wt}}$$

Based on the above pharmacodynamics model, W is proportional to the IC_{50} of the mutant. To examine the accuracy of W using an experimentally validated dose response curve, we utilized a set of mutants previously constructed by site-directed mutagenesis (Fig. S5) (11). The drug dose response curves were experimentally measured for each individual mutant. We found that the effects of mutations on drug resistance (W) estimated from the fitness data were generally consistent with estimates based on the measured dose response curves (Fig. 4B and Fig. S6; Materials and Methods), suggesting that the drug resistance score W is accurate and can be used to estimate IC_{50}. Thus, we estimated the IC_{50} value of all profiled mutations (Fig. 4C). We found that residues 28, 31, 92, and 93 are enriched with drug resistance mutations with high IC_{50} values, consistent with a previous experimental study (11). These positions were also reported to be hot spots for DCV drug resistance in multiple HCV genotypes (73–75).

This pharmacodynamics model can help explain the change of DFE with the increase of drug concentration. The mutations that reduce a protein's binding affinity to drug molecules (i.e., less inhibited by the drug) may come with a fitness cost (i.e., smaller f_0 than the wild type). Among all the nonlethal single amino acid substitutions profiled in our HCV NS5A protein library, we found that roughly half of the mutations increased resistance to DCV (i.e., improved new function) at the expense of replication fitness without drug (Fig. 4D; Spearman's $\rho = -0.13$, $P = 8.3 \times 10^{-4}$). This group of resistance mutations (lower-right section in Fig. 4D) can become beneficial when the environmental stress imposed by the antiviral drug is strong, leading to an increase in the proportion of beneficial mutations at higher drug concentrations. Moreover, as the wild-type virus moves further away from the phenotypic optimum, the relative fitness of the drug-resistant mutant is expected to increase with environmental stress (Fig. 4A, dashed line). Indeed, we found that the relative fitness of validated drug-resistant mutants increased at higher drug concentration (Fig. 4E).

DISCUSSION

Site-directed mutagenesis and experimental evolution are traditional approaches to examine the DFE (76–79). Both methods provide pivotal insights into the shape of the DFE, yet with limitations. The site-directed mutagenesis approach requires fitness assays for each individual mutant and can only provide a sparse sampling of

FIG 4 Legend (Continued)

the $RF_{mut[drug]} \rightarrow f_{0\,Mut} IC_{50}(mut) IC_{50(mut)} / f_{0\,WT} IC_{50(WT)}$. In the hypothetical curve, we set $f_{0\,WT} = 1$, $IC_{50(WT)} = 1$; $f_{0\,Mut} = 0.2$, $IC_{50(Mut)} = 10$. Then $RF_{mut[drug]}$ would approach $0.2 \cdot 10 = 2$ when drug concentration $\rightarrow \infty$. The hypothetical curves explain the increase of beneficial mutations upon drug treatment. (B) The drug resistance score W estimated from validation experiments of individual mutants is consistent with the estimates based on the pharmacodynamics modeling of the screening result (Pearson correlation = 0.71, $P = 1.1 \times 10^{-4}$). As the experiment collected virus at 48 h postinfection while the screening cultured for 144 h, the ratio between log($W_{experimental\ validation}$) and log($W_{fitness\ profiling}$) is expected to be 48 h/144 h = 0.33 under exponential growth. The fitted linear curve (red line) gives a ratio of 0.34, which is consistent with the expectation. (C) The heatmap shows the predicted IC_{50} value of all mutants. Lethal mutations are marked with black. (D) The effects of mutations on replication fitness (i.e., fitness without drug) and drug resistance score W at [DCV] = 40 pM are shown by the scatterplot. (E) Relative fitness of the validated drug-resistant and drug-sensitive mutants (Fig. S5) as a function of [DCV]. With the increase of drug concentration, the relative fitness of the drug-resistant mutant is increased.

mutations. In experimental evolution, the sampling of sequence space via *de novo* mutations is biased toward large-effect beneficial mutations, as they are more likely to fix in the population. In contrast, the deep mutational scanning approach (9), which utilizes high-throughput sequencing to simultaneously assay the fitness or phenotype of a library of mutants, allows for unbiased and large-scale sampling of fitness land-scapes and thus is ideal for studying the characteristics of empirical DFE. The downside of this high-throughput approach is that the fitness measurements can be noisy, espe-cially for large mutant libraries (80). In our experiment, we divided the mutant library into smaller sublibraries (~500 mutants) in selection experiments. We compared the data to an independent experiment and found that the fitness estimates were largely reproducible (Fig. S2). We also showed that the observed change in the DFE under dif-ferent conditions was consistent with validation experiments (Fig. 3). Since this study is focused on the properties of the entire distribution of mutations rather than the effects of specific mutations, our findings on the general patterns of DFE are robust to the errors in fitness estimates. Our study quantified the fitness effects of single amino acid substitutions in the drug-targeted region of an essential viral protein. In general, the empirical DFE of HCV NS5A was consistent with previous findings that viral proteins were highly optimized in the natural condition and very sensitive to the effects of dele-terious mutations.

One crucial point is that DFE will vary as a function of the environment (33, 35, 81). In the study by Stiffler, the level of environmental stress is controlled by ampicillin con-centration (33). Because TEM-1's function is to degrade ampicillin, deleterious muta-tions that impair the enzyme function ("loss-of-function") would become more delete-rious at higher dose of ampicillin. In our system, we expect that the function of HCV NS5A protein for viral replication and drug resistance to daclatasvir are two relatively independent traits; thus, the dose of daclatasvir should not alter the strength of purify-ing selection on maintaining protein stability and viral replication. Indeed, we do not find much difference on the deleterious side of DFE across different environments. Instead, we have observed significant changes on the beneficial side of DFE as a func-tion of the drug dose. Because HCV NS5A protein is not well adapted in the novel envi-ronment of daclatasvir selection, the effect of drug resistance mutations ("gain-of-func-tion") becomes more beneficial at higher drug dose. Moreover, due to the pleiotropic effect of mutations on drug resistance and replication fitness (Fig. 4), there is an increasing supply of beneficial mutations at higher drug dose.

Although different systems have distinct protein-drug interactions that lead to dif-ferent resistance profiles (82), the results in our study provide a general framework to study the DFE of drug-targeted proteins. Future studies along this line will further our understanding of how proteins evolve new functions under the constraint of maintain-ing their original function (83), as exemplified in the evolution of resistance to directly acting antiviral drugs (84). Quantifying the characteristics of the DFE of drug-targeted proteins in different environments (e.g., varying levels of environmental stress or con-flicting selection pressures) would allow us to assess repeatability in the outcomes of viral evolution (85) and guide the design of therapies to minimize drug resistance (34).

MATERIALS AND METHODS

Mutagenesis. The mutant library of HCV NS5A protein domain IA (86 amino acids) was constructed using saturation mutagenesis as previously described (11). In brief, the entire region was divided into five sublibraries, each containing 17 to 18 amino acids (~500 mutants in each sublibrary). NNK (N: A/T/C/G, K: T/G) was used to replace each amino acid. The oligos, each of which contains one random codon, were synthesized by Integrated DNA Technologies (IDT). The mutated region was ligated to the flanking constant regions, subcloned into the pFNX-HCV plasmid, and then transformed into bacteria. The pFNX-HCV plasmid carrying the viral genome was synthesized in Ren Sun's lab based on the chimeric sequence of genotype 2a HCV strains J6/JFH1.

Cell culture. The human hepatoma cell line (Huh-7.5.1) was provided by Francis Chisari from the Scripps Research Institute, La Jolla, California. The cells were cultured in T-75 tissue culture flasks (Genesee Scientific) at 37°C with 5% CO_2. The complete growth medium contained Dulbecco's modified Eagle's medium (Corning Cellgro), 10% heat-inactivated fetal bovine serum (Omega Scientific), 10 mM

HEPES (Life Technologies), 1× minimal essential medium (MEM) nonessential amino acids solution (Life Technologies) and 1× penicillin-streptomycin-glutamine (Life Technologies).

Selection of mutant viruses. The plasmid mutant library was transcribed *in vitro* using a T7 RiboMAX Express large scale RNA production system (Promega) and purified using a PureLink RNA minikit (Life Technologies). Then, 10 μg of *in vitro* transcribed RNA was used to transfect 4 million Huh-7.5.1 cells via electroporation using Bio-Rad Gene Pulser (246 V, 950 μF). The supernatant was collected 6 days posttransfection, and virus titer was determined by immunofluorescence assay. The viruses collected after transfection were used to infect ~2 million Huh-7.5.1 cells with an multiplicity of infection (MOI) of around 0.1 to 0.2. The five sublibraries were passaged for selection separately. For the three different levels of selection pressure, the growth medium was supplemented with 10 pM, 40 pM, and 100 pM HCV NS5A inhibitor daclatasvir (BMS-790052), respectively. The supernatant was collected at 6 days postinfection.

Preparation of Illumina sequencing samples. For each sample, viral RNA was extracted from 700 μl supernatant collected after transfection and after selection using a QIAamp viral RNA minikit (Qiagen). Extracted RNA was reverse transcribed into cDNA with a SuperScript III reverse transcriptase kit (Life Technologies). The targeted region in NS5A (51 to 54 nucleotides [nt]) was PCR amplified using KOD Hot Start DNA polymerase (Novagen). The Eppendorf thermocycler was set as follows: 2 min at 95°C; 25 to 35 three-step cycles of 20 s at 95°C, 15 s at 52 to 56°C (sublibrary 1, 52°C; 2, 52°C; 3, 52°C; 4, 56°C; 5, 54°C), and 25s at 68°C; 1 min at 68°C. The number of PCR cycles was chosen based on the copy number of cDNA templates as determined by quantitative PCR (qPCR) (Bio-Rad). The PCR products were purified using a PureLink PCR purification kit (Life Technologies) and prepared for Illumina HiSeq 2000 sequencing (paired-end, 100 bp) following 5'-phosphorylation using T4 polynucleotide kinase (New England BioLabs), 3' dA-tailing using a dA-tailing module (New England BioLabs), and TA ligation of the adapter using T4 DNA ligase (Life Technologies). Each sample was tagged with unique 3-bp customized barcodes, which were part of the adapter sequence and were sequenced as the first three nucleotides in both the forward and reverse reads (59).

Analysis of Illumina sequencing data. The sequencing data were parsed using the SeqIO function of BioPython. The reads from different samples were demultiplexed by the barcodes and mapped to the entire mutated region in NS5A by allowing, at maximum, 5 mismatches with the reference genome (11). Since both forward and reverse reads cover the whole amplicon, we used paired reads to correct for sequencing errors. A mutation was called only if it was observed in both reads and the quality score at the corresponding position was at least 30. Sequencing reads containing mutations not supposed to appear in our single-codon mutant library were excluded from downstream analysis. The sequencing depth for each sublibrary is at least ~10^5 and 2 orders of magnitude higher than the library complexity.

Calculation of relative fitness. For each condition of selection experiments (i.e., different concentration of daclatasvir [DCV]), the relative fitness (RF) of a mutant virus to the wild-type virus was calculated by the relative changes in frequency after selection,

$$RF_{mut}([DCV]) = \left(\frac{f_{mut}^{T=2}}{f_{mut}^{T=1}}\right) / \left(\frac{f_{WT}^{T=2}}{f_{WT}^{T=1}}\right)$$

where $f_{mut}^{T=round}$ and $f_{WT}^{T=round}$ is the frequency of the mutant virus and the wild-type virus at round 1 (after transfection) or round 2 (after infection). The fitness of wild-type virus is normalized to 1. The fitness values estimated from one round (round 1 to round 2) have been shown to be highly consistent with estimates based on round 0 to round 1 (Fig. S2) and estimates from multiple rounds of selection (11). A mutant was labeled as "missing" if the mutant's frequency in the plasmid library was less than 0.0005. A mutant was labeled as "lethal" if the mutant's frequency after transfection was less than 0.0005 or its frequency after infection was 0 (RF = 0) (11).

The threshold for beneficial mutations was chosen as $2\sigma_{silent}$, where $2\sigma_{silent}$ is the standard deviation of the log transformed RF of synonymous mutations (Fig. 1). The fitness effects of nonsynonymous mutations leading to the same amino acid substitution were averaged to estimate the fitness effect of the given single amino acid substitution.

Fitting the distribution of fitness effects of beneficial mutations. The distribution of log transformed RF of beneficial mutations was fitted to a generalized Pareto distribution following a maximum likelihood approach (70):>):

$$F(x|\kappa,\tau) = \begin{cases} 1-(1+\frac{\kappa}{\tau}x)^{-1/\kappa}, x\geq 0, \text{ if } \kappa > 0 & \text{(Frechet)} \\ 1-(1+\frac{\kappa}{\tau}x)^{-1/\kappa}, 0\leq x < -\frac{\tau}{\kappa}, \text{ if } \kappa < 0 & \text{(Weibull)} \\ 1-e^{-x/\tau}, x\geq 0, \text{ if } \kappa = 0 & \text{(Gumbel)} \end{cases}$$

Only mutations with RF higher than the beneficial threshold $2\sigma_{silent}$ were included in the distribution of beneficial mutations. The RFs were normalized to the beneficial threshold. The shape parameter κ determines the tail behavior of the distribution, which can be divided into three domains of attraction, Gumbel domain (exponential tail, κ = 0), Weibull domain (truncated tail, κ < 0), and Fréchet domain (heavy tail, κ > 0). For each selection condition, a likelihood ratio test was performed to evaluate whether the null hypothesis κ = 0 (exponential distribution) can be rejected.

Fitting the distribution of fitness effects to Fisher's geometrical model. Fisher's geometrical model predicts that the distribution of fitness effects of mutations is distributed according to a negative displaced gamma distribution (71, 72). This distribution has a shape parameter (α), a scale parameter (β), and a displacement parameter (s_o). We assume that RFs are measured with a normally distributed measurement error with standard deviation σ_{silent}. Thus, the observed distribution of RFs is modeled as the sum of a gamma and normally distributed random variable. We used the NormalGamma package in R to numerically compute the normal-gamma density function (86). Maximum likelihood estimates of the parameters of the negative displaced gamma distribution were obtained with L-BFGS-B optimization implemented in the R function optim.

Inferring drug resistance from fitness data. We can quantify the drug resistance of each mutant in the library by computing its fold change in relative fitness,

$$W([DCV]) = \frac{RF_{mut}([DCV])}{RF_{mut}}$$

Here, RF_{mut} is the relative fitness of a mutant under the natural condition (i.e., no drug). W is the fold change in relative fitness and represents the level of drug resistance relative to the wild type. $W > 1$ indicates drug resistance, and $W < 1$ indicates drug sensitivity.

This empirical measure of drug resistance can be directly linked to a simple pharmacodynamics model (84), where the viral replicative fitness is modeled as a function of drug dose,

$$W_{predict}([DCV]) = \left(\frac{IC_{mut}}{[DCV] + IC_{mut}} \right) / \left(\frac{IC_{wt}}{[DCV] + IC_{wt}} \right)$$

Here, IC denotes the half-inhibitory concentration. The Hill coefficient describing the sigmoidal shape of the dose response curve is fixed to 1, as used in fitting the dose response curves of wild-type virus and validated mutant viruses. The drug resistance score W inferred from fitness data is consistent with the drug resistance score $W_{predict}$ predicted from dose response curves of validated mutants (Fig. S6).

Calculation of relative solvent accessibility. DSSP (https://swift.cmbi.umcn.nl/gv/dssp/DSSP_3 .html) was used to compute the solvent accessible surface area (SASA) (87) from the HCV NS5A protein structure (PDB: 3FQM) (52). The SASA was then normalized to relative solvent accessibility (RSA) using the empirical scale reported in reference 88.

Predictions of protein stability. Homology modeling based on our NS5A sequence was performed using the SWISS-MODEL server (58) (https://swissmodel.expasy.org/).

The $\Delta\Delta G$ (in Rosetta energy units) of HCV NS5A mutants was predicted using PyRosetta (version PyRosetta4.conda.mac.python37.Release r242) as the difference in scores between the monomer structure of mutants (single amino acid mutations from sites 32 to 103) and the reference (PDB: 3FQM; 4CL1 or the homology model). The score is designed to capture the change in thermodynamic stability caused by the mutation ($\Delta\Delta G$).

The PDB file of the NS5A dimer was cleaned and trimmed to a monomer (chain A). Next, all side chains were repacked (sampling from the 2010 Dunbrack rotamer library [88]) and minimized for the reference structure using the "ddg_monomer" scoring function. After an amino acid mutation was introduced, the mutated residue was repacked, followed by line minimization of the backbone and all side chains (algorithm, "linmin"). This procedure was performed 10 times, and the predicted $\Delta\Delta G$ of a mutant structure is the average of all the scoring structures.

We note that predictions based on the NS5A monomer structure were only meant to provide a crude profile of how mutations at each site may impact protein stability. Potential structural constraints at the dimer interface have been ignored, which is further complicated by the observations of two different NS5A dimer structures (51, 52).

Diversity of HCV sequences identified in patients. Aligned nucleotide sequences of HCV NS5A protein were downloaded from the Los Alamos National Lab database (89) (all HCV genotypes, ~2,600 sequences total) and clipped to the region of interest (amino acids 18 to 103 of NS5A). Sequences that caused gaps in the alignment of the H77 reference genome were manually removed. After translation to amino acid sequences, sequences with ambiguous amino acids were removed (~2,300 amino acid sequences after filtering). The sequence diversity at each amino acid site was quantified by Shannon entropy. The frequency of amino acid on each site that differs from our NS5A sequence was calculated.

Data and reagent availability. All research materials are available upon request. Raw sequencing data have been submitted to the NIH Sequence Read Archive (SRA) under BioProject number PRJNA395730. All scripts have been deposited at https://github.com/leidai-evolution/DFE-HCV.

Ethics statement. The use of human cell lines and infectious agents in this paper is approved by the Institutional Biosafety Committee at the University of California, Los Angeles (IBC no. 40.10.2-f).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, TXT file, 0.03 MB.

FIG S1, TIF file, 1.2 MB.

FIG S2, TIF file, 1.2 MB.

FIG S3, TIF file, 1.2 MB.
FIG S4, TIF file, 1.2 MB.
FIG S5, TIF file, 1.2 MB.
FIG S6, TIF file, 1.2 MB.
TABLE S1, PDF file, 0.02 MB.
TABLE S2, PDF file, 0.03 MB.
TABLE S3, PDF file, 0.03 MB.

REFERENCES

1. Domingo E, Sheldon J, Perales C. 2012. Viral quasispecies evolution. Microbiol Mol Biol Rev 76:159–216. https://doi.org/10.1128/MMBR.05023-11.
2. Goldberg DE, Siliciano RF, Jacobs WR. 2012. Outwitting evolution: fighting drug-resistant TB, malaria, and HIV. Cell 148:1271–1283. https://doi.org/10.1016/j.cell.2012.02.021.
3. Metcalf CJE, Birger RB, Funk S, Kouyos RD, Lloyd-Smith JO, Jansen VAA. 2015. Five challenges in evolution and infectious diseases. Epidemics 10:40–44. https://doi.org/10.1016/j.epidem.2014.12.003.
4. Ke R, Loverdo C, Qi H, Sun R, Lloyd-Smith JO. 2015. Rational design and adaptive management of combination therapies for hepatitis C virus infection. PLoS Comput Biol 11:e1004040. https://doi.org/10.1371/journal.pcbi.1004040.
5. Barton JP, Goonetilleke N, Butler TC, Walker BD, McMichael AJ, Chakraborty AK. 2016. Relative rate and location of intra-host HIV evolution to evade cellular immunity are predictable. Nat Commun 7:11660. https://doi.org/10.1038/ncomms11660.
6. Turner PE, Elena SF. 2000. Cost of host radiation in an RNA virus. Genetics 156:1465–1470.
7. Wright S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution, p 356–366. In Proceedings of the Sixth International Congress of Genetics. http://www.esp.org/books/6th-congress/facsimile/contents/6th-cong-p356-wright.pdf.
8. Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. Elife 3:e03300. https://doi.org/10.7554/eLife.03300.
9. Fowler DM, Fields S. 2014. Deep mutational scanning: a new style of protein science. Nat Methods 11:801–807. https://doi.org/10.1038/nmeth.3027.
10. Wu NC, Young AP, Dandekar S, Wijersuriya H, Al-Mawsawi LQ, Wu T-T, Sun R. 2013. Systematic identification of H274Y compensatory mutations in influenza A virus neuraminidase by high-throughput screening. J Virol 87:1193–1199. https://doi.org/10.1128/JVI.01658-12.
11. Qi H, Olson CA, Wu NC, Ke R, Loverdo C, Chu V, Truong S, Remenyi R, Chen Z, Du Y, Su S-Y, Al-Mawsawi LQ, Wu T-T, Chen S-H, Lin C-Y, Zhong W, Lloyd-Smith JO, Sun R. 2014. A quantitative high-resolution genetic profile rapidly identifies sequence determinants of hepatitis C viral fitness and drug sensitivity. PLoS Pathog 10:e1004064. https://doi.org/10.1371/journal.ppat.1004064.
12. Hietpas RT, Jensen JD, Bolon DNA. 2011. Experimental illumination of a fitness landscape. Proc Natl Acad Sci U S A 108:7896–7901. https://doi.org/10.1073/pnas.1016024108.
13. Du Y, Wu NC, Jiang L, Zhang T, Gong D, Shu S, Wu T-T, Sun R. 2016. Annotating protein functional residues by coupling high-throughput fitness profile and homologous-structure analysis. mBio 7:e01801-16 https://doi.org/10.1128/mBio.01801-16.
14. Du Y, Xin L, Shi Y, Zhang T-H, Wu NC, Dai L, Gong D, Brar G, Shu S, Luo J, Reiley W, Tseng Y-W, Bai H, Wu T-T, Wang J, Shu Y, Sun R. 2018. Genome-wide identification of interferon-sensitive mutations enables influenza vaccine design. Science 359:290–296. https://doi.org/10.1126/science.aan8806.
15. Olson CA, Wu NC, Sun R. 2014. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. Curr Biol 24:2643–2651. https://doi.org/10.1016/j.cub.2014.09.072.
16. Li C, Qian W, Maclean CJ, Zhang J. 2016. The fitness landscape of a tRNA gene. Science 352:837–840. https://doi.org/10.1126/science.aae0568.
17. Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016. Network of epistatic interactions within a yeast snoRNA. Science 352:840–844. https://doi.org/10.1126/science.aaf0965.
18. Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R. 2016. Adaptation in protein fitness landscapes is facilitated by indirect paths. Elife 5:e16965. https://doi.org/10.7554/eLife.16965.
19. He X, Liu L. 2016. Toward a prospective molecular evolution. Science 352:769–770. https://doi.org/10.1126/science.aaf7543.
20. Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat Rev Genet 8:610–618. https://doi.org/10.1038/nrg2146.
21. Bataillon T, Bailey S. 2014. Effects of new mutations on fitness: insights from models and data. Ann N Y Acad Sci 1320:76–92. https://doi.org/10.1111/nyas.12460.
22. Burch CL, Chao L. 2000. Evolvability of an RNA virus is determined by its mutational neighbourhood. Nature 406:625–628. https://doi.org/10.1038/35020564.
23. Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit E, Poulain J, Barnaud G, Gros P-A, Tenaillon O. 2013. Capturing the mutational landscape of the beta-lactamase TEM-1. Proc Natl Acad Sci U S A 110:13067–13072. https://doi.org/10.1073/pnas.1215206110.
24. Chevereau G, Dravecká M, Batur T, Guvenek A, Ayhan DH, Toprak E, Bollenbach T. 2015. Quantifying the determinants of evolutionary dynamics leading to drug resistance. PLoS Biol 13:e1002299. https://doi.org/10.1371/journal.pbio.1002299.
25. Bank C, Hietpas RT, Jensen JD, Bolon DNA. 2015. A systematic survey of an intragenic epistatic landscape. Mol Biol Evol 32:229–238. https://doi.org/10.1093/molbev/msu301.
26. Desai MM. 2013. Statistical questions in experimental evolution. J Stat Mech 2013:P01003. https://doi.org/10.1088/1742-5468/2013/01/P01003.
27. Pressman A, Moretti JE, Campbell GW, Müller UF, Chen IA. 2017. Analysis of in vitro evolution reveals the underlying distribution of catalytic

activity among random sequences. Nucleic Acids Res 45:8167–8179. https://doi.org/10.1093/nar/gkx540.

28. Gonzalez R, Wu B, Li X, Martinez F, Elena SF. 2019. Mutagenesis scanning uncovers evolutionary constraints on tobacco etch potyvirus membrane-associated 6K2 protein. Genome Biol Evol 11:1207–1222. https://doi.org/10.1093/gbe/evz069.

29. Kemble H, Nghe P, Tenaillon O. 2019. Recent insights into the genotype-phenotype relationship from massively parallel genetic assays. Evol Appl 12:1721–1742. https://doi.org/10.1111/eva.12846.

30. Firnberg E, Labonte JW, Gray JJ, Ostermeier M. 2014. A comprehensive, high-resolution map of a gene's fitness landscape. Mol Biol Evol 31:1581–1592. https://doi.org/10.1093/molbev/msu081.

31. Dandage R, Pandey R, Jayaraj G, Rai M, Berger D, Chakraborty K. 2018. Differential strengths of molecular determinants guide environment specific mutational fates. PLoS Genet 14:e1007419-20. https://doi.org/10.1371/journal.pgen.1007419.

32. Kraemer SA, Morgan AD, Ness RW, Keightley PD, Colegrave N. 2016. Fitness effects of new mutations in Chlamydomonas reinhardtii across two stress gradients. J Evol Biol 29:583–593. https://doi.org/10.1111/jeb.12807.

33. Stiffler MA, Hekstra DR, Ranganathan R. 2015. Evolvability as a function of purifying selection in TEM-1 β-lactamase. Cell 160:882–892. https://doi.org/10.1016/j.cell.2015.01.035.

34. Ogbunugafor CB, Wylie CS, Diakite I, Weinreich DM, Hartl DL. 2016. Adaptive landscape by environment interactions dictate evolutionary dynamics in models of drug resistance. PLoS Comput Biol 12:e1004710. https://doi.org/10.1371/journal.pcbi.1004710.

35. Lalić J, Cuevas JM, Elena SF. 2011. Effect of host species on the distribution of mutational fitness effects for an RNA virus. PLoS Genet 7:e1002378. https://doi.org/10.1371/journal.pgen.1002378.

36. Vale PF, Choisy M, Froissart R, Sanjuán R, Gandon S. 2012. The distribution of mutational fitness effects of phage φX174 on different hosts. Evolution 66:3495–3507. https://doi.org/10.1111/j.1558-5646.2012.01691.x.

37. Cervera H, Lalić J, Elena SF. 2016. Effect of host species on topography of the fitness landscape for a plant RNA virus. J Virol 90:10160–10169. https://doi.org/10.1128/JVI.01243-16.

38. Foll M, Poh Y-P, Renzette N, Ferrer-Admetlla A, Bank C, Shim H, Malaspinas A-S, Ewing G, Liu P, Wegmann D, Caffrey DR, Zeldovich KB, Bolon DN, Wang JP, Kowalik TF, Schiffer CA, Finberg RW, Jensen JD. 2014. Influenza virus drug resistance: a time-sampled population genetics perspective. PLoS Genet 10:e1004185. https://doi.org/10.1371/journal.pgen.1004185.

39. Renzette N, Pfeifer SP, Matuszewski S, Kowalik TF, Jensen JD. 2017. On the analysis of intrahost and interhost viral populations: human cytomegalovirus as a case study of pitfalls and expectations. J Virol 91:e01976-16. https://doi.org/10.1128/JVI.01976-16.

40. Orr H. 2003. The distribution of fitness effects among beneficial mutations. Genetics 1526:1519–1526.

41. Gao M, Nettles RE, Belema M, Snyder LB, Nguyen VN, Fridell RA, Serrano-Wu MH, Langley DR, Sun J-H, O'Boyle DR, Lemm JA, Wang C, Knipe JO, Chien C, Colonno RJ, Grasela DM, Meanwell NA, Hamann LG. 2010. Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. Nature 465:96–100. https://doi.org/10.1038/nature08960.

42. Lindenbach BD, Evans MJ, Syder AJ, Wölk B, Tellinghuisen TL, Liu CC, Maruyama T, Hynes RO, Burton DR, McKeating JA, Rice CM. 2005. Complete replication of hepatitis C virus in cell culture. Science 309:623–626. https://doi.org/10.1126/science.1114016.

43. Wakita T, Pietschmann T, Kato T, Date T, Miyamoto M, Zhao Z, Murthy K, Habermann A, Kräusslich H-G, Mizokami M, Bartenschlager R, Liang TJ. 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. Nat Med 11:791–796. https://doi.org/10.1038/nm1268.

44. Doud MB, Bloom JD. 2016. Accurate measurement of the effects of all amino-acid mutations on influenza hemagglutinin. Viruses 8:155–117. https://doi.org/10.3390/v8060155.

45. Sanjuan R, Moya A, Elena SF. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc Natl Acad Sci U S A 101:8396–8401. https://doi.org/10.1073/pnas.0400146101.

46. Elena SF, Carrasco P, Daròs J-A, Sanjuán R. 2006. Mechanisms of genetic robustness in RNA viruses. EMBO Rep 7:168–173. https://doi.org/10.1038/sj.embor.7400636.

47. Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, Bhella D, Gifford RJ, Rixon FJ, Bieniasz PD. 2013. Extreme genetic fragility of the HIV-1 capsid. PLoS Pathog 9:e1003461. https://doi.org/10.1371/journal.ppat.1003461.

48. Burch C, Guyader S, Samarov D, Shen H. 2007. Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus ϕ6. Genetics 176:467–476. https://doi.org/10.1534/genetics.106.067199.

49. Silander O, Tenaillon O, Chao L. 2007. Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. PLoS Biol 5:e94. https://doi.org/10.1371/journal.pbio.0050094.

50. Liberles DA, Teichmann SA, Bahar I, Bastolla U, Bloom J, Bornberg-Bauer E, Colwell LJ, de Koning APJ, Dokholyan NV, Echave J, Elofsson A, Gerloff DL, Goldstein RA, Grahnen JA, Holder MT, Lakner C, Lartillot N, Lovell SC, Naylor G, Perica T, Pollock DD, Pupko T, Regan L, Roger A, Rubinstein N, Shakhnovich E, Sjölander K, Sunyaev S, Teufel AI, Thorne JL, Thornton JW, Weinreich DM, Whelan S. 2012. The interface of protein structure, protein biophysics, and molecular evolution. Protein Sci 21:769–785. https://doi.org/10.1002/pro.2071.

51. Tellinghuisen TL, Marcotrigiano J, Rice CM. 2005. Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. Nature 435:374–379. https://doi.org/10.1038/nature03580.

52. Love RA, Brodsky O, Hickey MJ, Wells PA, Cronin CN. 2009. Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. J Virol 83:4395–4403. https://doi.org/10.1128/JVI.02352-08.

53. Ramsey DC, Scherrer MP, Zhou T, Wilke CO. 2011. The relationship between relative solvent accessibility and evolutionary rate in protein evolution. Genetics 188:479–488. https://doi.org/10.1534/genetics.111.128025.

54. Das R, Baker D. 2008. Macromolecular modeling with Rosetta. Annu Rev Biochem 77:363–382. https://doi.org/10.1146/annurev.biochem.77.062906.171838.

55. Chaudhury S, Lyskov S, Gray JJ. 2010. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics 26:689–691. https://doi.org/10.1093/bioinformatics/btq007.

56. Wylie CS, Shakhnovich EI. 2011. A biophysical protein folding model accounts for most mutational fitness effects in viruses. Proc Natl Acad Sci U S A 108:9916–9921. https://doi.org/10.1073/pnas.1017572108.

57. Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, Arnold FH. 2005. Thermodynamic prediction of protein neutrality. Proc Natl Acad Sci U S A 102:606–611. https://doi.org/10.1073/pnas.0406744102.

58. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303. https://doi.org/10.1093/nar/gky427.

59. Wu NC, Olson CA, Du Y, Le S, Tran K, Remenyi R, Gong D, Al-Mawsawi LQ, Qi H, Wu T-T, Sun R. 2015. Functional constraint profiling of a viral protein reveals discordance of evolutionary conservation and functionality. PLoS Genet 11:e1005310. https://doi.org/10.1371/journal.pgen.1005310.

60. MacLean RC, Buckling A. 2009. The distribution of fitness effects of beneficial mutations in Pseudomonas aeruginosa. PLoS Genet 5:e1000406. https://doi.org/10.1371/journal.pgen.1000406.

61. Bataillon T, Zhang T, Kassen R. 2011. Cost of adaptation and fitness effects of beneficial mutations in Pseudomonas fluorescens. Genetics 189:939–949. https://doi.org/10.1534/genetics.111.130468.

62. Carrasco P, de la Iglesia F. d l, Elena SF. 2007. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco etch virus. J Virol 81:12979–12984. https://doi.org/10.1128/JVI.00524-07.

63. Cowperthwaite MC, Bull JJ, Meyers LA. 2005. Distributions of beneficial fitness effects in RNA. Genetics 170:1449–1457. https://doi.org/10.1534/genetics.104.039248.

64. Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuán R. 2010. Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. Genetics 185:603–609. https://doi.org/10.1534/genetics.110.115162.

65. Imhof M, Schlötterer C. 2001. Fitness effects of advantageous mutations in evolving Escherichia coli populations. Proc Natl Acad Sci U S A 98:1113–1117. https://doi.org/10.1073/pnas.98.3.1113.

66. Kassen R, Bataillon T. 2006. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat Genet 38:484–488. https://doi.org/10.1038/ng1751.

67. Rokyta D, Joyce P, Caudle S, Wichman H. 2005. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat Genet 37:441–444. https://doi.org/10.1038/ng1535.

68. Orr H. 2006. The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. J Theor Biol 238:279–285. https://doi.org/10.1016/j.jtbi.2005.05.001.

69. Orr HA. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution (N Y) 52:935–949. https://doi.org/10.2307/2411226.

70. Beisel CJ, Rokyta DR, Wichman HA, Joyce P. 2007. Testing the extreme value domain of attraction for distributions of beneficial fitness effects. Genetics 176:2441–2449. https://doi.org/10.1534/genetics.106.068585.

71. Martin G, Lenormand T. 2006. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. Evolution 60:893–907. https://doi.org/10.1554/05-412.1.

72. Bank C, Hietpas RT, Wong A, Bolon DN, Jensen JD. 2014. A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. Genetics 196:841–852. https://doi.org/10.1534/genetics.113.156190.

73. Nakamoto S, Kanda T, Wu S, Shirasawa H, Yokosuka O. 2014. Hepatitis C virus NS5A inhibitors and drug resistance mutations. World J Gastroenterol 20:2902–2912. https://doi.org/10.3748/wjg.v20.i11.2902.

74. Kliemann DA, Tovo CV, Da Veiga ABG, De Mattos AA, Wood C. 2016. Polymorphisms and resistance mutations of hepatitis C virus on sequences in the European hepatitis C virus database. World J Gastroenterol 22:8910–8917. https://doi.org/10.3748/wjg.v22.i40.8910.

75. Patiño-Galindo JÁ, Salvatierra K, González-Candelas F, López-Labrador FX. 2016. Comprehensive screening for naturally occurring hepatitis c virus resistance to direct-acting antivirals in the NS3, NS5A, and NS5B genes in worldwide isolates of viral genotypes 1 to 6. Antimicrob Agents Chemother 60:2402–2416. https://doi.org/10.1128/AAC.02776-15.

76. Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. 2016. The mutational robustness of influenza A virus. PLoS Pathog 12:e1005856. https://doi.org/10.1371/journal.ppat.1005856.

77. Domingo-Calap P, Cuevas JM, Sanjuán R. 2009. The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. PLoS Genet 5:e1000742. https://doi.org/10.1371/journal.pgen.1000742.

78. Sanjuán R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Philos Trans R Soc Lond B Biol Sci 365:1975–1982. https://doi.org/10.1098/rstb.2010.0063.

79. Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature 519:181–186. https://doi.org/10.1038/nature14279.

80. Matuszewski S, Hildebrandt ME, Ghenu A-H, Jensen JD, Bank C. 2016. A statistical guide to the design of deep mutational scanning experiments. Genetics 204:77–87. https://doi.org/10.1534/genetics.116.190462.

81. Martin G, Lenormand T. 2006. The fitness effect of mutations across environments: a survey in light of fitness landscape models. Evolution 60:2413–2427. https://doi.org/10.1111/j.0014-3820.2006.tb01878.x.

82. Robinson M, Tian Y, Delaney WE, Greenstein AE. 2011. Preexisting drug-resistance mutations reveal unique barriers to resistance for distinct antivirals. Proc Natl Acad Sci U S A 108:10290–10295. https://doi.org/10.1073/pnas.1101515108.

83. Soskine M, Tawfik DS. 2010. Mutational effects and the evolution of new protein functions. Nat Rev Genet 11:572–582. https://doi.org/10.1038/nrg2808.

84. Rosenbloom DIS, Hill AL, Rabi SA, Siliciano RF, Nowak MA. 2012. Antiretroviral dynamics determines HIV evolution and predicts therapy outcome. Nat Med 18:1378–1385. https://doi.org/10.1038/nm.2892.

85. de Visser JAGM, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. Nat Rev Genet 15:480–490. https://doi.org/10.1038/nrg3744.

86. Plancade S, Rozenholc Y, Lund E. 2012. Generalization of the normal-exponential model: exploration of a more accurate parametrisation for the signal distribution on Illumina BeadArrays. BMC Bioinformatics 13:329. https://doi.org/10.1186/1471-2105-13-329.

87. Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637. https://doi.org/10.1002/bip.360221211.

88. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. 2013. Maximum allowed solvent accessibilites of residues in proteins. PLoS One 8:e80635. https://doi.org/10.1371/journal.pone.0080635.

89. Kuiken C, Yusim K, Boykin L, Richardson R. 2005. The Los Alamos hepatitis C sequence database. Bioinformatics 21:379–384. https://doi.org/10.1093/bioinformatics/bth485.