Environmental changes dictate selection and nonspecific epistasis in an empirical phenotype-environment-fitness landscape

Chen, J.Z.[1], Fowler, D.M.[2,3], and Tokuriki, N.[1 *]

[1]Michael Smith Laboratories, University of British Columbia

[2]Department of Genome Sciences, University of Washington

[3]Department of Bioengineering, University of Washington

*Correspondence: tokuriki@msl.ubc.ca

Summary

The fitness landscape, a function that maps genotypic and phenotypic changes to their effects on fitness, is an invaluable concept in evolutionary biochemistry. Though widely discussed, measurements of phenotype-fitness landscapes in proteins remain scarce. Here, we quantify all single mutational effects on fitness and phenotype (antibiotic resistance level) of VIM-2 β-lactamase (5600 variants) across a 64-fold range of ampicillin concentrations by deep mutational scanning. We then construct a phenotype-fitness landscape that takes variations in environmental selection pressure into account (a phenotype-environment-fitness landscape). We found that a simple, empirical landscape accurately models the ~39,000 mutational data points, which suggests the evolution of VIM-2 can be predicted based on the selection environment. Our landscape provides new quantitative knowledge on the evolution of the β-lactamases and proteins in general, particularly their evolutionary dynamics under sub-inhibitory antibiotic concentrations, as well as the mechanisms and environmental dependence of nonspecific epistasis.

Introduction

The quantitative understanding of genotype-phenotype-fitness relationships of proteins is central in evolutionary biochemistry (Dalziel et al., 2009; Kaltenbach and Tokuriki, 2014; Yi and Dean, 2019). Genotype, a protein's sequence information, largely dictates its phenotypes, which are the measurable *in vitro* physio-chemical traits of proteins such as protein activity and stability. Nonetheless, *in vivo*, multiple factors could modulate protein function and behavior in the cell (*e.g.*, the level of protein expression). The fitness of the organism carrying a specific protein is greatly influenced by the environment where the organism lives, and the selection pressure to a particular protein's function vary significantly with environmental factors. Thus, deciphering the 'phenotype-fitness' relationship across environmental variation is essential to understanding and potentially predicting a protein's evolutionary dynamics (Bershtein et al., 2017; DePristo et al., 2005; Kemble et al., 2019; Pál et al., 2006). The phenotype-fitness landscapes typically manifest as a non-linear relationship, in the form of a 'threshold' or 'mesa-like' model (Dasmeh and Serohijos, 2018; DePristo et al., 2005; Hartl et al., 1985; Lundin et al., 2018; Lunzer et al., 2005; Sarkisyan et al., 2016; Tokuriki and Tawfik, 2009). In this model, the fitness conferred by variants with biochemical traits above a certain selection 'threshold' would saturate or plateau, while fitness rapidly decreases for variants below the threshold of the trait. Such models have been widely recognized conceptually, and provide molecular explanations for important evolutionary phenomenon such as nonspecific epistasis (Bershtein et al., 2006; Dasmeh and Serohijos, 2018; Diss and Lehner, 2018; Kemble et al., 2019; Pokusaeva et al., 2019), mutational robustness (Bershtein et al., 2013, 2006; Hartl et al., 1985; Lundin et al., 2018; Tokuriki and Tawfik, 2009) and evolvability (Dean, 1995; DePristo et al., 2005; Lunzer et al., 2005; Mehlhoff et al., 2020; Sarkisyan et al., 2016; Stiffler et al., 2015; Tokuriki and Tawfik, 2009). However, quantitative and experimental studies of the threshold model are rare, as most reports contain just a handful of data points and each landscape is confined to the scope of a single level of selection pressure (Bershtein et al., 2013; Dean, 1995; Hartl et al., 1985; Lunzer et al., 2005; Pokusaeva et al., 2019). Consequently, detailed empirical models that can recapitulate the fitness landscapes of proteins, let alone the phenotype-fitness relationship across varying environmental conditions, have never been established. Such models will enable the quantitative evaluation of the evolutionary dynamics in proteins, and the development of predictive models for protein evolution in the response to a given environmental selection pressure.

To describe a realistic phenotype-fitness landscape, we require a quantitative mapping of the phenotype-fitness relationship with careful examinations in the context of environmental factors. Recent advances in largescale mutational characterization experiments, such as deep mutational scanning (DMS)(Fowler et al., 2014, 2010), have enabled the high-throughput measurement of protein fitness (growth of host organism carrying encoded protein variants) and phenotype (*e.g.*, antibiotic resistance level, protein binding and expression) for several thousands of mutational variants (Firnberg et al., 2014; Hietpas et al., 2011; Jacquier et al., 2013; Klesmith et al., 2017; Mavor et al., 2016; Melnikov et al., 2014; Noda-García et al., 2019; Pokusaeva et al., 2019; Starr et al., 2020, 2017; Stiffler et al., 2015; Wrenbeck et al., 2017). DMS across different conditions or genetic backgrounds can also determine higher dimensional characteristics of protein phenotype and/or fitness, such as antibiotic concentration dependent fitness(Firnberg et al., 2014; Melnikov et al., 2014; Stiffler et al., 2015), substrate specificity(Chen et al., 2020; Meer et al., 2016; Melnikov et al., 2014) and inter/intra-protein and protein-environment interactions (Diss and Lehner, 2018; Flynn et al., 2020; Horovitz et al., 2019; Kemble et al., 2020; Mavor et al., 2018,

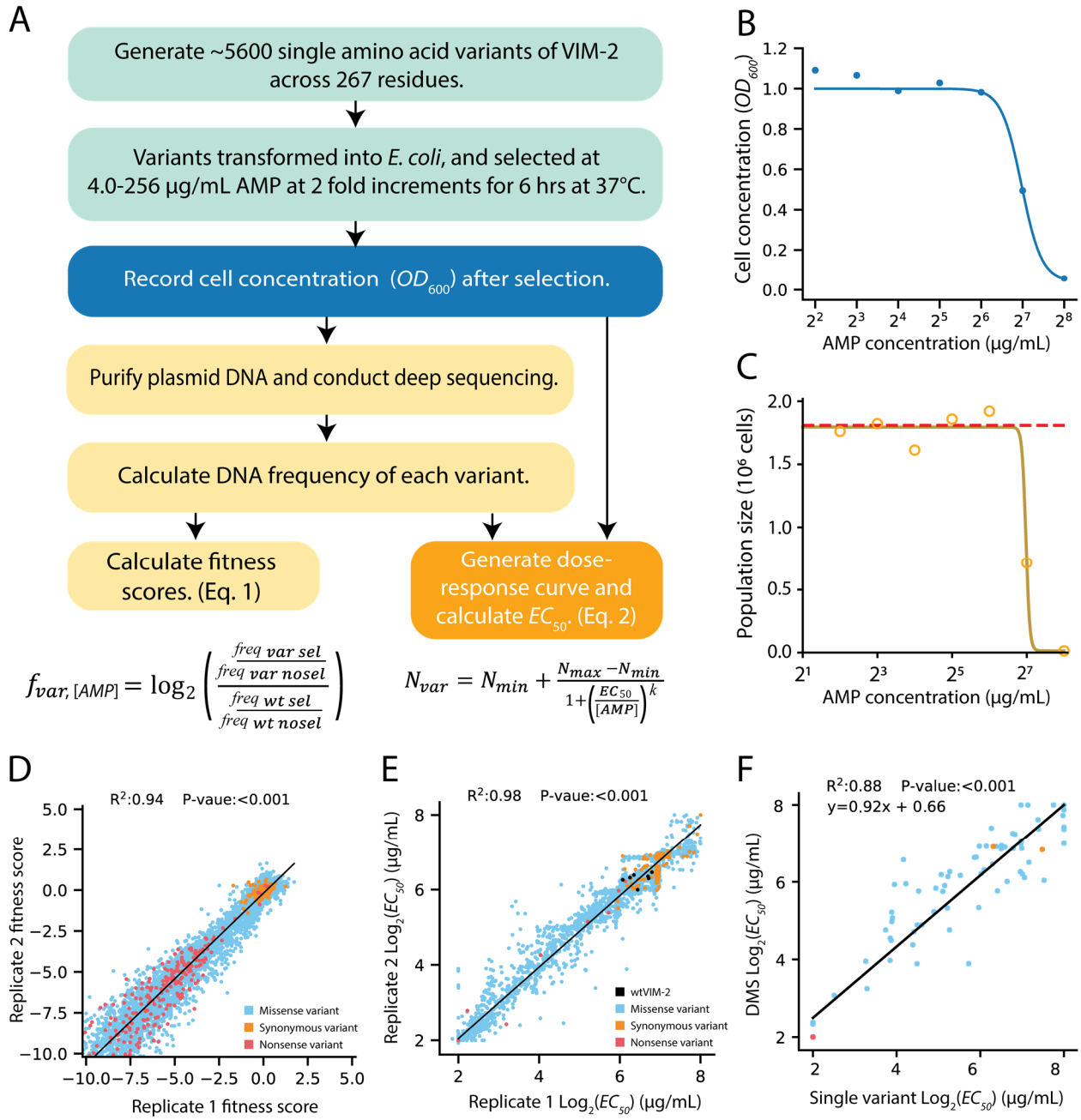
2016; Noda-García et al., 2019; Salinas and Ranganathan, 2018; Starita et al., 2013; Thompson et al., 2020).

In this work, we seek to obtain a global phenotype-environment-fitness landscape using the VIM-2 β-lactamase as a model system, for which we previously generated all single point mutants (~5600 variants) (Chen et al., 2020). To this end, we determined the fitness of an *Escherichia coli* strain carrying VIM-2 variants across different environments dictated by seven different ampicillin (AMP) concentrations (~39,000 data points). We also measured the phenotype, specifically the antibiotic resistance level (EC_{50}), of each variant by integrating the experimental data across all tested AMP concentrations into dose-response curves. We show that the phenotype-fitness relationships exhibit a universal behavior according to AMP concentration. Taking advantage of this result, we generate an empirical model for generalizing the phenotype-environment-fitness landscape of VIM-2. We then demonstrate that this empirical and quantitative model can explain the evolutionary behavior of the resistance gene at low antibiotic concentrations, and the molecular mechanisms underlying the prevalence of nonspecific epistasis in mutational datasets.

Results

Selection of VIM-2 variant libraries in a range of ampicillin concentrations

The overall workflow for our antibiotic selection experiment is presented in **Figure 1A**. We previously generated a library containing ~5600 single amino acid variants of wild-type (*wt*) VIM-2 in a vector that expresses VIM-2 through a constitutive ampR promoter (Chen et al., 2020). The library was transformed into *E. coli* and subsequently grown in LB liquid media containing a range of AMP concentrations (4.0-256 µg/mL AMP in two-fold increments), or in the absence of AMP, for 6 hours at 37°C. The AMP concentrations used in this study covers the full dynamic range of VIM-2's conferred resistance, from just above 2 µg/mL AMP where *E. coli* can grow without VIM-2 expression (Chen et al., 2020), to more than double the EC_{50} of 84 µg/mL where *E. coli* expressing *wt*VIM-2 can no longer grow. The concentration of cells in each culture after the selection was measured via OD_{600} to track the total population size (**Figure 1B**), and the mutated region of VIM-2 gene was amplified by PCR for deep sequencing (Chen et al., 2020). Two replicates of the experiment with antibiotic selection and three replicates of growth without selection were conducted.



$$f_{var,[AMP]} = \log_2\left(\frac{\frac{freq\ var\ sel}{freq\ var\ nosel}}{\frac{freq\ wt\ sel}{freq\ wt\ nosel}}\right) \qquad N_{var} = N_{min} + \frac{N_{max} - N_{min}}{1+\left(\frac{EC_{50}}{[AMP]}\right)^k}$$

Figure 1. Workflow for generating fitness scores and EC_{50} values using DMS. (**A**) Flowchart of the DMS process. A library of protein variants is first generated and transformed into host *E. coli*, which are then placed under selection with a range of antibiotic concentrations. The OD_{600} of the selected culture is recorded after selection and the selected culture is processed for deep sequencing. For each variant, the DNA frequency can be used to directly calculate the fitness scores, as well as the population size of each variant (N_{var}) after selection (using the OD_{600}). The population size can be used to generate a dose-response curve for each variant. (**B**) Cell density of a culture pool encompassing variants in the first 39 positions of VIM-2, measured after selection for each AMP concentration. (**C**) Example of a dose-response curve generated for the I19A variant. The solid line represents a fit to a sigmoidal curve. The red dashed line indicates the average non-selected population size of the replicate. Additional examples of dose-response curves are available in **Figure S2**. (**D**) Correlation of variant fitness scores between two replicates under selection with 64 µg/mL AMP. Correlation plots for other AMP concentrations are presented in **Figure S1**. For panels **D-F**, variants are colored by mutation type; the same color scheme applies to all panels. (**E**)

Correlation of EC_{50} values fitted from two replicates. (**F**) Correlation between EC_{50} determined by using deep sequencing (y-axis) and or from isolated variants grown in separate liquid cultures (x-axis). See **Table S1** for the EC_{50} values of isolated variants.

Calculation of fitness scores and EC_{50}

The output of the DMS experiments generates sequencing counts for each variant in the library, which can be used to calculate the fitness scores (reflecting relative growth) at each AMP concentration as well as the phenotype (EC_{50}) for each variant. The crux of our approach is that the EC_{50} values are calculated independently from the fitness score in each condition, such that their values are not interdependent, even when using the same set of deep sequencing data. Thus, this allows us to examine the relationship between phenotype and fitness as separate factors under each selection pressure.

The fitness (f) of a variant under each AMP concentration is calculated as the relative growth of a variant compared to *wt*VIM-2 as we described previously (**Figure 1A**, **Table S2** for all fitness scores)(Chen et al., 2020). For each variant, and at each level of AMP selection, an enrichment score was calculated as the ratio of the frequency of a variant after selection to the frequency without selection. Next, the fitness score of each variant was calculated by normalizing the enrichment score of a variant to that of *wt*VIM-2, generating a measure of fitness relative to *wt*VIM-2 using **equation (1)**.

$$f_{var,[AMP]} = \log_2\left(\frac{\frac{freq_{\,var\,sel}}{freq_{\,var\,nosel}}}{\frac{freq_{\,wt\,sel}}{freq_{\,wt\,nosel}}}\right) \quad (1)$$

The antibiotic resistance level of each variant was determined via a dose-response curve by integrating the deep sequencing data from all seven different conditions (**Figure 1A-1C, Table S3** for all EC_{50} values). To this end, we first calculated the population size of cells expressing each variant (N_{var}) at each antibiotic concentration by multiplying the frequency with the cell concentration of the culture after antibiotic selection. Then we generated a dose-response curve for each variant by plotting N_{var} against the antibiotic concentration, and calculated the EC_{50} by fitting the sigmoidal relationship in **equation (2)**

$$N_{var} = N_{min} + \frac{N_{max} - N_{min}}{1+\left(\frac{EC_{50}}{[AMP]}\right)^{k}} \quad (2)$$

Where N_{max} and N_{min} correspond to the maximum and minimum value of the variant population size, N_{var}, [AMP] refers to the AMP concentration in µg/mL and k represents the Hill coefficient, which determines the rate of change of the curve.

Quality of fitness scores and EC_{50} values

Next, we confirmed that our experiments yielded reliable fitness and EC_{50} values for the ~5600 variants across a range of AMP conditions. The fitness scores show a strong linear correlation between replicates, with an R^2 of at least 0.92 for all conditions except at the highest AMP concentration (256 µg/mL), which displayed an R^2 of 0.45 (**Figure 1D** and **S1**). The observed noise arising at 256 µg/mL between replicates likely results from the use of an AMP concentration much higher than the $EC_{50\text{-wt}}$ (84 µg/mL). At this high concentration, the frequency of the *wt*VIM-2 after selection, which is the denominator in **equation (1)**, is drastically reduced and introduces greater error in this condition. The fitness scores show distinct separation between functional and non-functional variants. Variants with synonymous mutations are centered at a fitness score of 0

152 and the variants containing a stop codon mutation are centered near a fitness of -4 (**Figure 1D** and
153 **S1**). Hence, a fitness score of -4 or lower denotes inactive variants, and scores that are below -4
154 are set to -4 in subsequent analyses.

155 We also found the EC_{50} values to strongly correlate between replicates, with a linear
156 regression R^2 of 0.98 (**Figure 1E**). Moreover, the EC_{50} values derived from DMS are comparable
157 with those measured on individual variants by a conventional dose-response curve resistance assay
158 using the same selection method (87 unique variants), with a strong linear regression R^2 of 0.88
159 (**Figure 1F**).

160 ***Dependence of fitness on phenotype and selection pressure***

161 The fitness scores we obtained across a range of AMP concentrations allows us to examine how
162 the selection pressure influences the distribution of fitness effects (DFE). The fitness scores reflect
163 relative growth of *E. coli* expressing a given VIM variant relative to *wt*VIM-2, and by extension,
164 the resistance conferred to *E. coli* by a VIM variant in each AMP concentration. The resistance
165 conferred by each variant is further dependent the EC_{50}, which reflects the variant's enzymatic
166 efficiency (k_{cat}/K_M) of and/or the amount of enzyme expressed in the periplasm fraction of the cell.
167 Thus, the DFE is inherently affected by both protein phenotype as well as the external selection
168 environment.

169 Overall examination shows the DFEs and the distribution of EC_{50} values are bimodal
170 (**Figure 2**), as typically observed in other proteins (Diss and Lehner, 2018; Firnberg et al., 2014;
171 Fowler et al., 2010; Gray et al., 2018; Hietpas et al., 2011; Jacquier et al., 2013; Mavor et al., 2016;
172 Pokusaeva et al., 2019; Rockah-Shmuel et al., 2015; Stiffler et al., 2015; Wrenbeck et al., 2017).
173 In each distribution, one peak is formed by variants of near-neutral effects and a second peak is
174 formed from variants with strongly deleterious effects, while only a small fraction of mutations
175 exhibit higher values than *wt*. To further quantify and classify the variants in the distributions as
176 having 'negative', 'neutral' or 'positive' effects with respect to the *wt*, we performed two-tailed z-
177 tests between the fitness (or EC_{50}) of the non-synonymous variants against the distribution of the
178 synonymous variants (a null model with N=251); test results are controlled for FDR by the

179 Benjamini-Hochberg procedure with $\alpha = 0.05$. The distribution of EC_{50} exhibits the most distinct
180 separation between neutral and non-functional variants, where 63% are negative ($EC_{50} < 72$
181 μg/mL), 35% of mutations are neutral (72 μg/mL $\leq EC_{50} \geq$ 171 μg/mL) and 2% are positive (EC_{50}
182 >171 μg/mL) (**Figure 2A**).

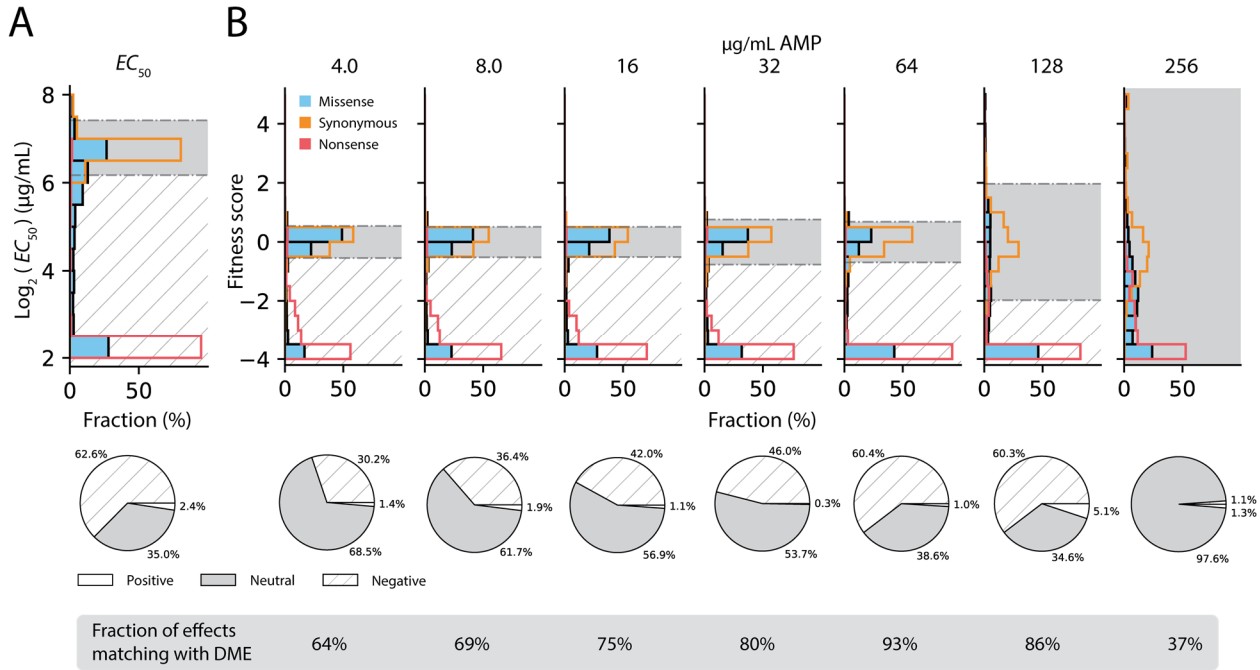


Figure 2. Distributions of fitness and EC_{50} effects. (**A**) The distribution of EC_{50} for all variants. (**B**) The DFEs of all variants under selection across all AMP selection concentrations. In all plots, the distributions are shown as a percentage of all variants within a mutation type (synonymous, missense or nonsense), with the color legend in **B**). The background colors indicate regions of positive (white), neutral (gray) and negative (hatched) effects, with the legend and a pie chart summary of the proportion of positive/neutral/negative effects below each histogram. For DFEs, the proportion of variants where positive/neutral/negative effects match the effects in the EC_{50} are displayed as percentages at the bottom. See **Table S2** for all fitness scores and **Table S3** for all EC_{50} values.

Naturally, fitness in the DFEs decline with increasing AMP concentration (**Figure 2B**), as can be inferred from the overall growth of the *E. coli* transformed with the library (**Figure 1B**). At lower AMP concentrations of 4.0-32 μg/mL, the majority (54-69%) of variants exhibit neutral effects. When the AMP concentration approaches two-fold of the $EC_{50\text{-wt}}$ of 84 μg/mL, the majority of variants become negative, up to 60%. At concentrations greater than $EC_{50\text{-wt}}$ (128-256 μg/mL), the distribution of synonymous mutations becomes broader as the *wt* resistance is also no longer sufficient to sustain high growth, and the data becomes noisier. In particular, at 256 μg/mL almost all variants are neutral again relative to *wt*VIM-2.

The shifting of the DFEs illuminates how mutational responses can be substantially altered depending on the selection pressure. For example, when comparing variant effects (negative/neutral/positive) across the DFEs and the distribution of EC_{50} values, only 52.3% of mutations exhibit the same effect across all distributions, with 26.7% consistently deleterious, 25.6% consistently neutral and none that are consistently positive. The remaining 47.7% of mutations exhibit distinct effects dependent on the selection pressure, where 39.5% switch between neutral and deleterious, 5.9% switch between positive and neutral, and 2.1% switch between all three states. We find that effects in the DFEs at 64-128 μg/mL mostly closely resemble the effects in EC_{50}, where 93% and 86% of variants have the same effect across fitness and EC_{50}, respectively (**Figure 2**); as these DFEs are measured at the closest concentrations to the $EC_{50\text{-wt}}$ of 84 μg/mL, the full extent of negative mutational effects is captured by the fitness scores without affecting the *wt* fitness.

Phenotype-fitness relationships across different AMP concentrations

Next, we seek to quantify the relationship between the EC_{50} values and the fitness scores obtained at each antibiotic concentration. We fit their relationship with a 'threshold' model (Bershtein et al., 2013; Chen et al., 2020; Dean, 1995; Hartl et al., 1985; Lunzer et al., 2005; Stiffler et al., 2015), *i.e.*, the four-parameter sigmoidal function in **equation (3)** (**Figure 3**).

$$f(EC_{50}) = f_{min} + \frac{f_{max} - f_{min}}{1+\left(\frac{X_{50}}{EC_{50}}\right)^{n}} \quad (3)$$

The f_{min} and f_{max} denote the minimum and maximum fitness scores, respectively, while n represents the Hill coefficient. The inflection point of the curve, also the concentration of AMP where fitness is half-way between f_{min} and f_{max}, is denoted by X_{50}.

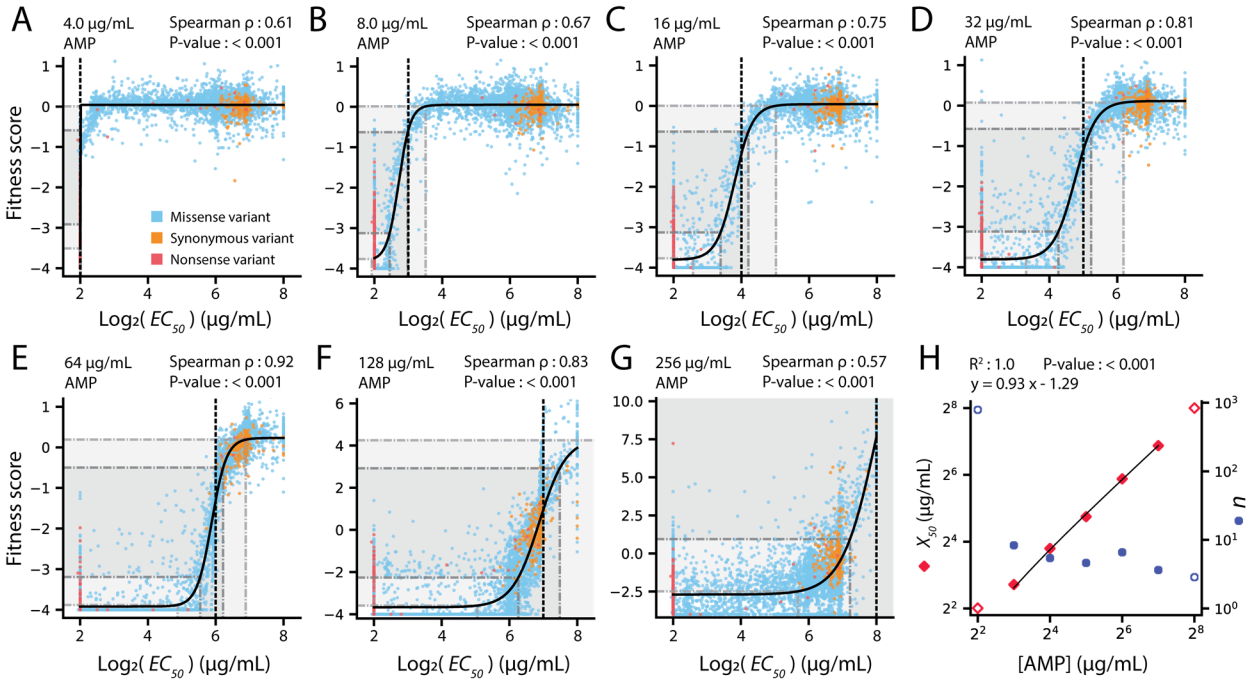


Figure 3. Phenotype-fitness landscapes across different AMP concentrations. Panels **A**-**G** show the comparison of DMS fitness scores with EC_{50} of all variants under selection in (**A**) 4.0 (N=5517), (**B**) 8.0 (N=5516), (**C**) 16 (N=5503), (**D**) 32 (N=5517), (**E**) 64 (N=5517), (**F**) 128 (N=5517) and (**G**) 256 (N=5517) µg/mL AMP. In each panel, variants are colored according to mutation type with the color legend in **A**). The solid black curve indicates the line of best fit for a sigmoidal curve fit (**equation 3**), while the black vertical dashed line indicates the AMP concentration used during selection. The dark grey regions extending to the x and y-axes indicates the region of linear association in the sigmoidal curve. The light grey regions extending to both axes indicate the range between 1% - 99% fitness in the fitted curve. The correlation coefficient and P-value of a Spearman rank-order correlation between the fitness score and EC_{50} is shown above each plot. (**H**) Sigmoidal fit parameters (**equation 3**) for the curves shown in panels **A**-**G**. The X_{50} values are shown in orange diamonds and the Hill coefficients (n) are shown in blue circles. The filled symbols indicate the conditions where the region of linear association in the sigmoidal curve (dark grey) was fully captured by the AMP concentrations of the experiment, while open symbols indicate the opposite. The R^2, P-value and equation for the linear regression of X_{50} with [AMP] is shown above the plot. Plots of all parameters can be found in **Figure S3**.

Overall, the threshold model recapitulates the phenotype-fitness relationship across all AMP concentrations, with Spearman rank-order correlation between fitness and EC_{50} ranging from

0.57-0.92 (**Figure 3A-3G**). Outside the sigmoidal transition range, fitness scores plateau at high and low fitness for variants with higher and lower EC_{50}, respectively. In particular, at AMP concentrations of 4.0-64 µg/mL (**Figure 3A-3E**), f_{min} and f_{max} maintain steady values in the fit at an average of -3.8 and 0.1, respectively (**Table 1**, **Figure S3A** and **S3B**). Since, *wt*VIM-2 can fully survive in these conditions, higher EC_{50} variants do not exhibit any fitness advantage over *wt*VIM-2. The existence of a fitness plateau for variants with an EC_{50} higher than the selection threshold also suggests that there are no systematic trade-offs between EC_{50} and fitness in our experimental system; higher EC_{50} variants do not exhibit low fitness in the presence of low AMP. At concentrations above $EC_{50\text{-wt}}$ (128 and 256 µg/mL) the f_{max} increases as *wt*VIM-2 becomes challenged and exhibits decreased growth, while higher EC_{50} variants continue to grow (**Figure 3F** and **3G**).

The quantification of the phenotype-fitness relationships reveals some notable traits. First, the transitions of fitness between the lower and upper fitness plateaus (X_1- X_{99}, points in the curve at 1% and 99% fitness) (**Table 1**) occur in a relatively small range of EC_{50} differences (3.0-7.3 fold range) when considering conditions where the transition can be fully resolved within experimental conditions (8.0-64 µg/mL AMP). This confirms AMP selects the phenotype of VIM-2 variants in a discrete, threshold-like manner. Second, where the linear range can be fully captured (8.0-128 µg/mL AMP), the shape of the transitions occur in a very similar fashion across AMP concentrations; the Hill coefficient remaining steady with an average of 5.7 ± 1.8 S.D, leading to linear transition ranges that spans a 1.4 - 2.3 fold change in EC_{50}. (**Figure 3H, S3C** and **S3D**). This suggests that the mode of selection by AMP against VIM phenotypes is universal across a wide range of AMP concentrations. Third, the inflection points (X_{50}) in the phenotype-fitness relationship are a nearly perfect function of AMP concentration as described by the linear relationship in **equation (4)** (**Figure 3H**).

$$X_{50}([AMP]) = 0.93 \times [AMP] - 1.29 \qquad (4)$$

Taken together, the selection of VIM-2 phenotypes is strongly dictated by the concentration of AMP in the environment, and the range of phenotypic selection is largely predictable across a defined range of concentrations.

Table 1. Fitted parameters of sigmoidal phenotype-fitness relationships

	Equation fit [a]				Linear range [a]				Transition range [a]			
[AMP] (µg/mL)	f_{max}	f_{min}	n	X_{50} (µg/mL)	EC_{50} lower (µg/mL)	EC_{50} upper (µg/mL)	f lower	f upper	X_1 (µg/mL)	X_{99} (µg/mL)	$f_{1\%}$	$f_{99\%}$
4	0.0	-3.5	788	4.0	4.0	4.0	-2.9	-0.6	4.0	4.0	-3.5	0.0
8	0.0	-3.8	8.3	6.6	5.5	7.9	-3.1	-0.6	3.8	11.4	-3.8	0.0
16	0.0	-3.8	5.4	13.9	10.5	18.5	-3.1	-0.6	6.0	32.4	-3.8	0.0
32	0.1	-3.8	4.6	26.9	19.2	37.5	-3.1	-0.6	9.9	72.7	-3.8	0.1
64	0.2	-3.9	6.6	58.9	46.6	74.5	-3.2	-0.5	29.3	118	-3.9	0
128	4.3	-3.7	3.6	117	76.7	179	-2.3	2.9	33.2	415	-3.6	4
256	18.0	-2.7	2.9	256	149	439	0.9	14.4	51.4	1280	-2.5	18

[a] All parameters shown are calculated from a fit to equation (3).

A phenotype-environment-fitness landscape of VIM-2

Utilizing the universal behavior of the Hill coefficient (n) and the inflection points (X_{50}) in the phenotype-fitness relationships, we propose a simple, empirical model to describe the complete phenotype-fitness landscape of VIM-2 under AMP selection in **equation (5)**.

$$f(EC_{50}, [AMP]) = \frac{1}{1+\left(\frac{X_{50}([AMP])}{EC_{50}}\right)^{5.7}} \quad (5)$$

This equation describes the fitness (f) of a variant as a function of its EC_{50} and the AMP concentration. All parameters are functionally the same as **equation (3)**, but n = 5.7, and the inflection point of the curve (X_{50}) is now a linear function of [AMP] as shown in **equation (4)**. Furthermore, we normalize the observed fitness scores to the f_{min} and f_{max} determined from the sigmoidal fits at each AMP concentration in **Table 1**, such that a fitness of 0 represents the minimum fitness and a fitness of 1 represents the maximum fitness (**Figure 4A**). The model enables us to depict the entire phenotype-environment-fitness landscape, which shows agreement with ~39,000 data points across seven different selection conditions (**Figure 4B**). Thus, the universal description by a simple mathematical model suggests that the selection of VIM variants is highly dictated by the environmental AMP concentration, and that the evolutionary response of any given variant can be rather predictable.

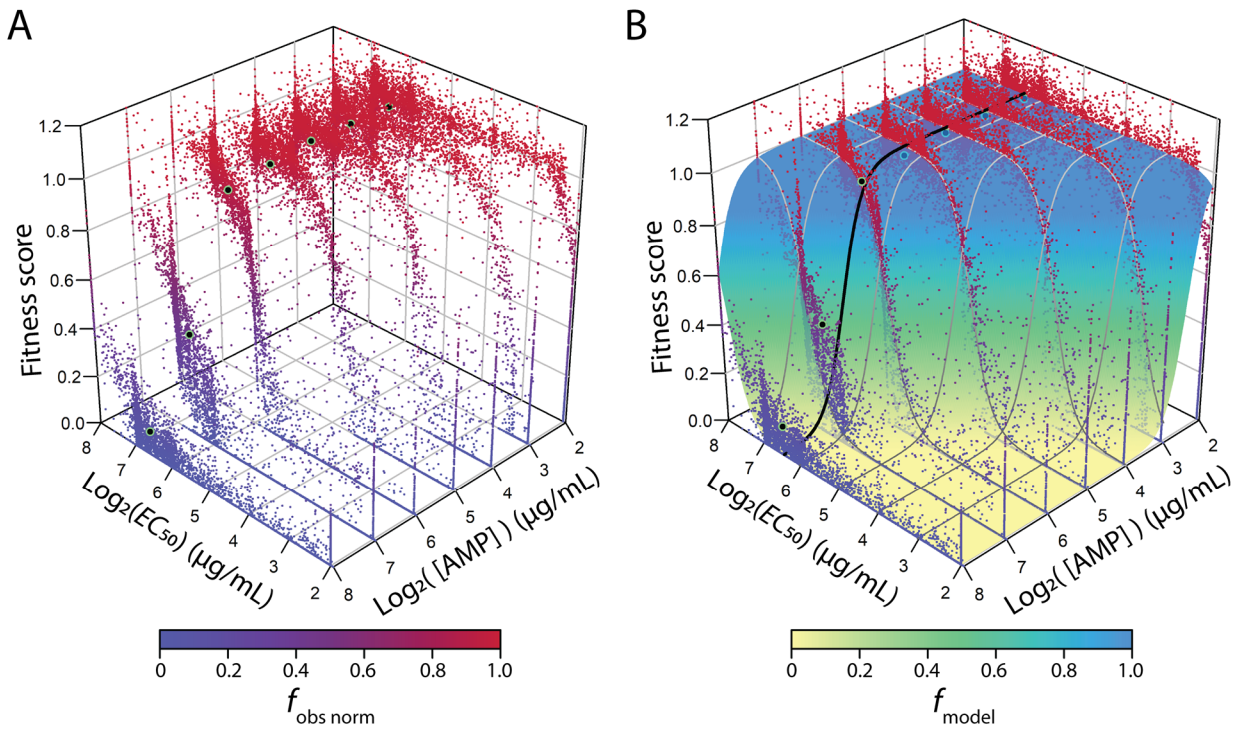


Figure 4. The VIM-2 Phenotype-fitness landscape as a function of EC_{50} and AMP concentration. (**A**) For each variant, the observed fitness score is plotted in relation to the AMP concentration during selection, and its respective EC_{50}. The fitness scores are normalized at each concentration according to **equation (13)** in the Methods. The black dots represent the fitness of *wt*VIM-2 after normalization. (**B**) The modelled behavior of fitness scores according to **equation (5)**. The dots are the same data as panel **A**) while the surface depicts the predicted fitness scores according to the model. The black line indicates the model fitness at the $EC_{50\text{-wt}}$. Cross sections of the model at two-fold intervals of EC_{50} or AMP concentrations are marked by grey lines.

Insight into the evolutionary dynamics of sub-MIC selection

Our global phenotype-environment-fitness landscape allows us to gain quantitative insights into an important question regarding the evolution of antibiotic resistance: a phenomenon known as sub-MIC selection. Antibiotic concentrations well below the 'minimal inhibitory concentration' (MIC) of the *wt* in the population can still promote the emergence of higher resistance variants (Fröhlich et al., 2020; Gullberg et al., 2011; Jasinska et al., 2020; Liu et al., 2011; Westhoff et al., 2017). Typically, any resistant variant with MIC_{var} higher than MIC_{wt} will have an absolute resistance and fitness advantage when the antibiotic concentration is over *MICwt* (*i.e.*, when MIC_{wt} < [AMP] < MIC_{var}) (Andersson and Hughes, 2014; Baquero et al., 1998; Blondeau, 2009; Drlica and Zhao, 2007; Gullberg et al., 2011) (**Figure 5A**). In sub-MIC selection where the antibiotic concentration is below *MICwt*, the *wt* is still viable, but the resistance difference could still lead to the selection of the variant over the *wt*, provided the benefit of resistance is high enough to offset any trade-offs due to fitness costs (Andersson and Hughes, 2014; Gullberg et al., 2011). The lowest antibiotic concentration where the variant can overtake the *wt* in fitness is designated the 'minimal selective concentration' or MSC. Quantitative knowledge supporting existing models of sub-MIC selection remains rare, however (Das et al., 2020; Gullberg et al., 2011).

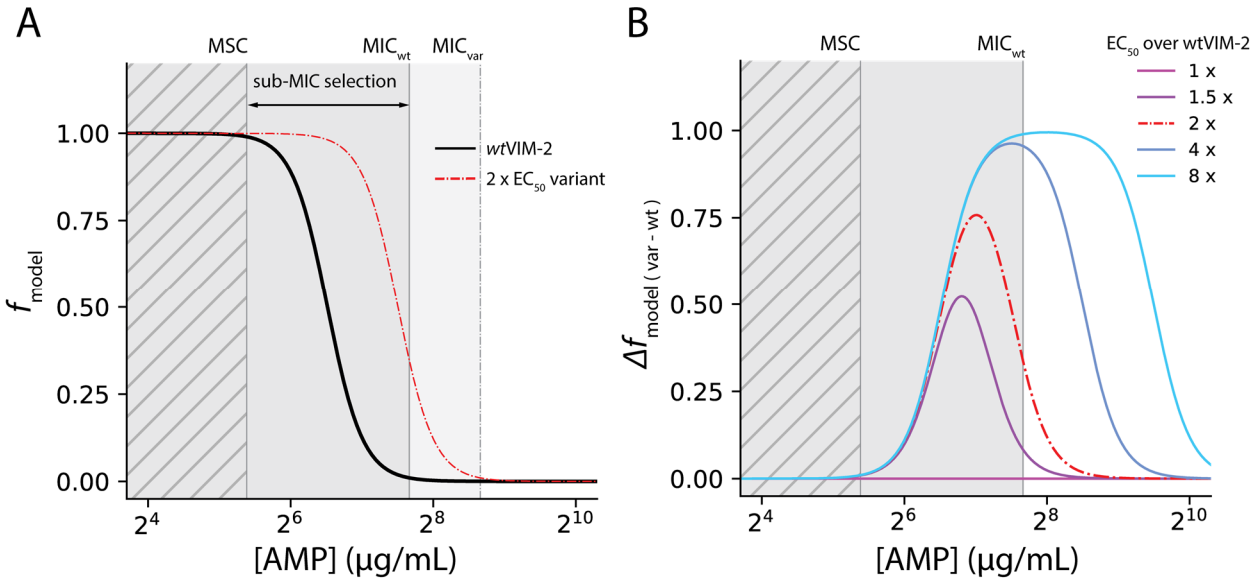


Figure 5. Sub-MIC selection behavior of VIM-2 variants. (**A**) The fitness of *wt*VIM-2 and a hypothetical variant with 2 × $EC_{50\text{-}wt}$ as a function of AMP concentration at a fixed EC_{50}. The MIC (f = 0.01) of both curves and the MSC (f_{wt} = 0.99) are marked as vertical lines. The region in light grey is the 'mutant selection window', the region in dark grey is the 'sub-MIC selection window' and the hatched region indicates no selection of the variant over wt. (**B**) The difference in fitness between *wt*VIM-2 and higher resistance variants across different AMP concentrations. Each curve indicates a variant with an EC_{50} of 1-8 fold relative to *wt*VIM-2. The coloring of the sub-MIC selection range in the background is the same as in **A**).

Our model can bridge this gap by providing a quantitative depiction of the mechanism underlying sub-MIC selection. For example, *wt*VIM-2, with EC_{50} of 84 µg/mL, would exhibit an MIC (1% of f_{max}) of 203 µg/mL AMP, and an MSC (99% of f_{max}) of 42 µg/mL (see '**Calculation of MIC and MSC**' in methods). Hence, for *wt*VIM-2, the range for sub-MIC selection can extend to ~5-fold below the MIC*wt*. Furthermore, we can calculate the fitness difference between *wt*VIM-2 and variants exhibiting higher resistance (**Figure 5B**). We observe that selection can still strongly favor the higher resistance variant (f_{var} - f_{wt} > 0.1) until the AMP concentration drops to ~2-fold

above the MSC (**Figure 5B**). This is the case even when the variant is only 1.5 to 2-fold more resistant than the *wt*, with the fitness difference being more pronounced for variants with higher resistance. Thus, AMP concentrations as low as 5-fold below the *wt* MIC can still select for variants with higher resistance than *wt*VIM-2, which is largely consistent with studies of sub-MIC selection using other distinct systems (Fröhlich et al., 2020; Gullberg et al., 2011; Jasinska et al., 2020; Liu et al., 2011; Westhoff et al., 2017).

The molecular basis for nonspecific epistasis

Our empirical model can also provide a quantitative understanding of the mechanisms underlying mutational epistasis, *i.e.*, the non-additive interaction between mutational effects (Domingo et al., 2019; Starr and Thornton, 2016). Epistasis is a key phenomenon in protein evolution that dictates the shape and ruggedness of fitness landscapes, which subsequently affect important evolutionary phenomenon such as robustness (Bershtein et al., 2013, 2006; Hartl et al., 1985; Lundin et al., 2018; Tokuriki and Tawfik, 2009), evolvability (Dean, 1995; DePristo et al., 2005; Lunzer et al., 2005; Mehlhoff et al., 2020; Sarkisyan et al., 2016; Stiffler et al., 2015; Tokuriki and Tawfik, 2009; Yang et al., 2019) and evolutionary contingency (Baier et al., 2019; Starr et al., 2017; Starr and Thornton, 2016). One form of epistasis, defined as 'nonspecific epistasis', is known to be highly prevalent and strongly influences the accessibility of mutations. Nonspecific epistasis typically stems from the non-linear relationship between fitness and protein phenotype, such as we observed between fitness and EC_{50} in this study (**Figure 3**). The key distinction of nonspecific (as opposed to specific) epistasis is that epistasis can still be observed at the level of fitness even if the mutational effects of two mutations on a phenotype, *i.e.*, EC_{50} (and thus enzymatic efficiency or protein expression) are non-epistatic (changes relative to *wt* are log-additive, or more simply described as 'additive') (Bank et al., 2015; Bershtein et al., 2006; Dasmeh and Serohijos, 2018; DePristo et al., 2005; Diss and Lehner, 2018; Domingo et al., 2019; Starr and Thornton, 2016). While the existence of nonspecific epistasis is widely appreciated, we lack quantitative understanding of the extents to which different types (magnitude/sign) of epistasis can arise depending on the selective environment.

To this end, we simulated the prevalence and types of nonspecific epistasis in VIM-2 fitness using the empirically measured distribution of EC_{50} and the phenotype-environment-fitness landscape (**Figure 6**). Since each variant only contains a single mutation, we can treat the variants' effects directly as mutational effects. Specifically, we randomly sampled 200 single point missense mutations, calculated the ΔEC_{50} values of a given mutation from *wt* (**Figure 6B**), then combined all pairwise mutations based on a null model (**Figure 6C**)—a double mutational effect on EC_{50} is equal to the additive effect of two single mutations. All pairwise combinations of 200 mutations leads to ${}_{200}C_2$ = 19900 mutant pairs. Then, using the model in **equation (5)**, we computed the change in fitness (Δf) for all individual and pairwise combinations at each AMP concentration using the single and pairwise EC_{50}. We quantified epistasis by comparing the observed fitness change (Δf_{obs}) calculated from the pairwise EC_{50}, and the expected fitness change (Δf_{exp}), which is the sum of the Δf of both individual mutations (**Figure 6D**, See methods for '**Nonspecific epistasis of VIM-2 variants**'). Finally, we categorized the observed epistasis into different types, based on the magnitude and sign of the observed fitness change compared to the expected additive change: First, magnitude epistasis, where Δf_{obs} has the same sign as Δf_{exp} but has significantly greater (positive magnitude epistasis) or lesser (negative magnitude epistasis) magnitude. Second, sign epistasis, where Δf_{obs} has an opposite sign to Δf_{exp} (**Figure 6A**).

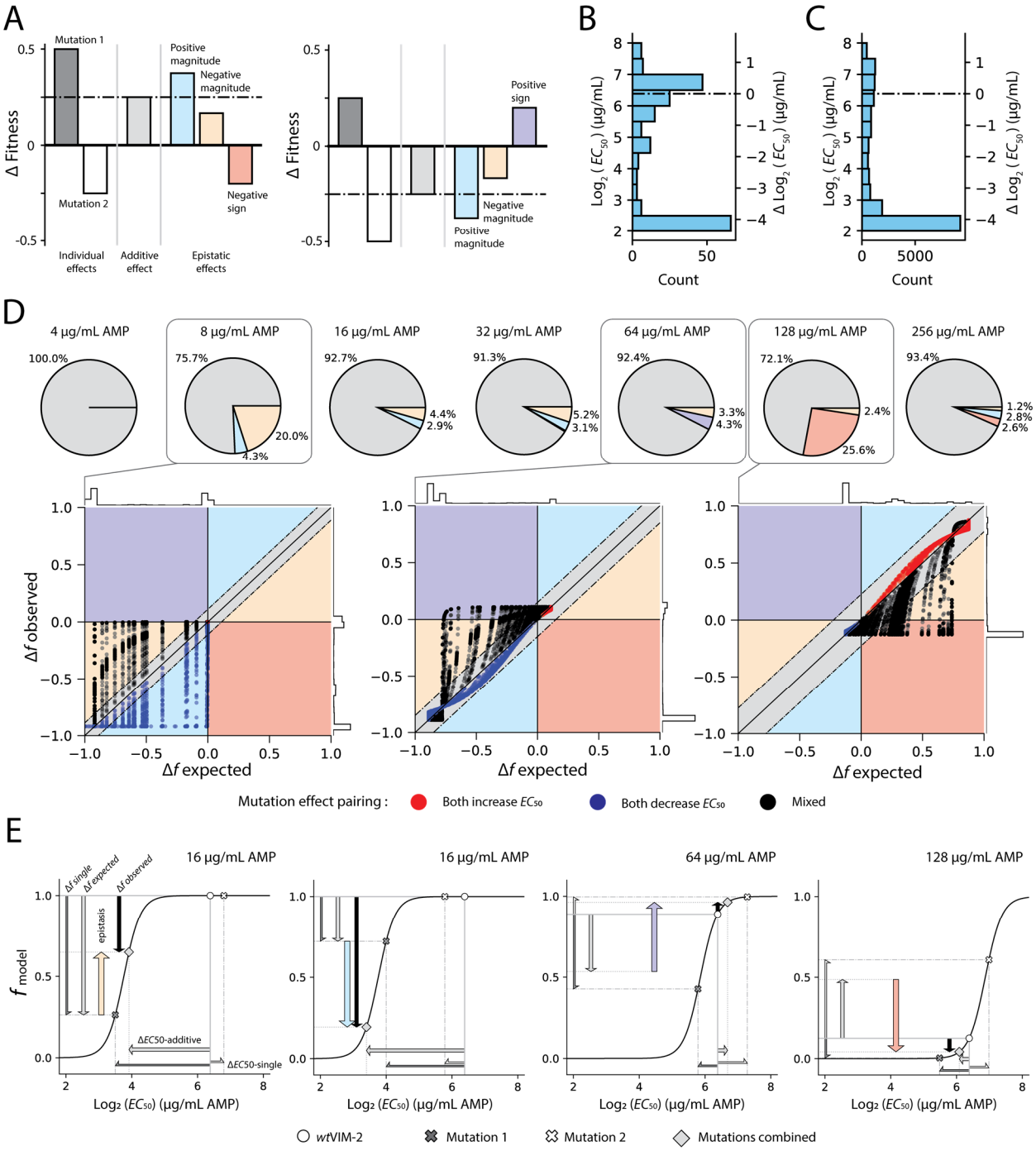


Figure 6. Nonspecific epistasis varies across selection conditions. (**A**) Conceptual illustration of each type of epistasis considering an overall positive (left) or negative (right) additive effect. This given color scheme for epistasis types is maintained throughout the figure. (**B**) Distribution of the 200 randomly sampled EC_{50} values, with the change in EC_{50} relative to *wt*VIM-2 displayed on the right y-axis. (**C**) Distribution of the 19900 pairwise additive EC_{50} values, with the change in EC_{50} relative to *wt*VIM-2 displayed on the right y-axis. (**D**) The proportion of mutational pairs exhibiting epistatic effects under each condition; only proportions higher than 0.5% are labelled. Below the pie charts are scatterplots from key conditions showing the difference between expected and observed fitness changes. The backgrounds of the scatterplots are also colored according to the type of epistatic effect. A gray background indicates no significant epistasis, or additivity, and is defined as a range of 1.96 × S.D. of the synonymous mutant fitness

distribution after normalization by **Equation (13)** for each AMP concentration. In each scatterplot, points representing the combined effect of a pair of mutations are colored according to the effect of each single mutation (red when both increase EC_{50} over *wt*VIM-2, blue when both decrease EC_{50} and black when one increases while the other decreases EC_{50}). Histograms of all data points in each scatterplot are shown for each axis. (**E**) Visual depictions of each type of epistasis at distinct AMP concentrations, as a result of a non-linear relationship between fitness and EC_{50}.

Intriguingly, all types of epistasis can be found in our sample set, and the proportion of each type is drastically dependent on the environment (0 - 26% epistasis). At the lowest AMP concentration (4 µg/mL), no epistasis can be observed as most mutations simply exhibit neutral effects. At 8 µg/mL AMP, the greatest proportion of magnitude epistasis is observed, with 20% of mutational combinations exhibiting negative magnitude epistasis and about 4% of combinations showing positive magnitude epistasis. At 16 and 32 µg/mL AMP, smaller fractions of positive and negative magnitude epistasis are observed (7.3% and 8.3% in total, respectively). However, from 64-256 µg/mL AMP, a substantial fraction of mutational combinations exhibit sign-epistasis. A small fraction of positive sign epistasis are observed at 64 µg/mL AMP, and negative sign epistasis is observed at 128 and 256 µg/mL AMP. In particular, a striking 26% of mutational combinations exhibit negative sign epistasis at 128 µg/mL AMP. It should be emphasized that these substantially different epistasis patterns observed in different environments are based on the same set of mutational combinations within the experimentally obtained distribution of EC_{50}, even with non-epistatic effects on the phenotypic level (EC_{50}); such patterns are the results of a non-linear mapping between phenotype and fitness.

An examination of the non-linear relationship can provide the mechanistic explanation underlying the constraints (**Figure 6E**). Positive magnitude epistasis only occurs in combinations of mutants that both increase EC_{50} or decrease EC_{50} over wtVIM-2, where either mutation alone exhibits a weak effect but the combined additive change in EC_{50} can traverse the steep fitness transition in the landscape and cause a much stronger effect. In contrast, negative magnitude epistasis only occurs in mutant combinations of opposite effects (one increases while the other decreases EC_{50}), as only mutations of opposing effects can diminish the combined additive EC_{50} change. Sign epistasis is only present near AMP concentrations close to the $EC_{50\text{-wt}}$ of 84 µg/mL, and occurs only between mutant combinations of opposite effects; sign epistasis cannot occur when the *wt*VIM-2 is on a high or low fitness plateau, where fitness changes can only occur in one direction (*e.g.*, fitness can only decrease when on the high fitness plateau). Positive sign epistasis is only observed when [AMP] is below $EC_{50\text{-wt}}$ while negative sign epistasis only occurs when [AMP] is above $EC_{50\text{-wt}}$ (**Figure 6D**).

Taken together, using the empirical distribution of mutational effects and the threshold model that we have established, we demonstrated that the selective environment strongly affects the resulting epistasis. We have also illustrated how epistasis can be prevalent in largescale mutational studies using fitness measurement (Araya et al., 2012; Bank et al., 2015; Diss and Lehner, 2018; Fowler et al., 2010; Olson et al., 2014; Palmer et al., 2015; Pokusaeva et al., 2019; Sarkisyan et al., 2016), as epistasis can still arise even when the underlying phenotype simply experience an additive change (Dasmeh and Serohijos, 2018). As the trend of apparent epistasis can be drastically different according to the environment, an understanding of the environmental factors is essential to decipher the evolutionary dynamics of proteins. Note that, in this model, we assumed additive effects on EC_{50}, while empirical mutational combinations are likely to exhibit non-additive effects with respect to protein function and stability due to specific epistasis (Domingo et al., 2019; Miton and Tokuriki, 2016; Starr and Thornton, 2016). In such cases,

428 epistasis will be even more frequently observed, and a rigorous model such as ours may help to
429 distinguish between nonspecific and specific epistasis within high-throughput mutational
430 experiments.

431 **Discussion**

432 In this study, we have presented a phenotype-environment-fitness landscape of VIM-2 β-lactamase,
433 constructed using a comprehensive set of ~39,000 data points, across a 64-fold range of AMP
434 concentrations. Our observations provide a long-sought, quantitative verification of an important
435 model for protein evolution, the 'threshold' model. With this comprehensive and empirical dataset,
436 we unveil how the phenotype-fitness relationship can be shaped by environmental variation,
437 particularly the selection strength. Importantly, we demonstrate that a simple quantitative model
438 can recapitulate the behavior of VIM-2 β-lactamase within the landscape. Our findings provide
439 evidence that the evolution of resistance in a population follows a relatively simple and predictable
440 rule, dictated through the selection pressure exerted by β-lactam antibiotics. Moreover, we provide
441 a mechanistic understanding of how different environments can cause different types of epistasis
442 to arise, solely from knowledge of the threshold relationship between protein phenotype and fitness.
443 Our observations suggest that the influence of the environment on the phenotype-fitness
444 relationship is extremely important to understand the evolutionary dynamics of proteins (Das et
445 al., 2020; Steinberg and Ostermeier, 2016). Hence, strong understanding of the effect of
446 environmental factors may allow us to accurately predict the evolution of antibiotic resistance
447 genes and other proteins.

448 Since we employed VIM-2, a class B β-lactamase, as a model for this study and measured
449 changes in EC_{50}—an important trait commonly measured trait for β-lactamases as well as other
450 antibiotic resistance genes—our model may be applicable to other β-lactamases, even in other
451 classes (A, C and D classes). Our findings could prove useful in developing a more general
452 understanding of β-lactamase evolution. Indeed, our observations agree with several behaviors
453 observed in β-lactamases, or other antibiotic resistance genes, beyond just the threshold-like
454 selection behavior. The ~5-fold range of sub-MIC selection estimated from our model largely
455 matches with other studies of the sub-MIC selection range in other systems (Fröhlich et al., 2020;
456 Gullberg et al., 2011; Jasinska et al., 2020; Liu et al., 2011; Westhoff et al., 2017). We do not
457 observe any notable, systematic trends that indicate trade-offs between high EC_{50} and fitness at
458 lower AMP concentrations, supporting observations that higher resistance does not necessarily
459 convey a large fitness cost, and that it can be maintained even under low selection pressure
460 (Gullberg et al., 2011; Westhoff et al., 2017). This also matches previous modelling where a flat
461 maximum fitness plateau is found for the mutants of TEM-1 irrespective of the selection pressure
462 (Stiffler et al., 2015). The lack of a fitness cost for more resistant variants is likely a result of the
463 non-essential nature of VIM-2 in our model system, coupled with a fairly moderate protein
464 expression level. Of course, when present, fitness trade-offs add greater complexity to the
465 dynamics of sub-MIC selection (Das et al., 2020; Gullberg et al., 2011). Regardless, our
466 quantitative descriptions for those phenomena are invaluable to further deepening our
467 understanding of the behaviors of these antibiotic resistance genes.

468 We note our relatively simple experimental system poses limits on the generalization of
469 our findings, and additional complexities may be studied in the future. *i*) We only mutated within
470 the VIM-2 gene, and mutations in other genetic context such as the promoter and origin of
471 replication of the plasmid might trigger different dynamics. *ii*) We conducted selection over 6
472 hours of continuous growth, up to the point where the cells are reaching the early phases of growth

saturation. Different culturing conditions may also expose different dynamics. *iii*) The results of AMP selection in *E. coli* may not reflect the behavior of VIM-2 with other classes of antibiotics, such as cephalosporins or carbapenems, as well as other clinically important bacterial hosts, *e.g.*, *P. aeruginosa* and *A. baumannii*. *iv*) These findings from a simple laboratory environment may only be partly applicable to behavior in natural environments, which will have additional variation outside of antibiotic concentration. *v*) Though we observed no systematic fitness-resistance trade-offs, higher resolution measurements of individual variants, and comparison to a condition without any selection may reveal trade-offs that are not as pronounced in our system (Gullberg et al., 2011; Jasinska et al., 2020; Mehlhoff et al., 2020; Westhoff et al., 2017; Zwart et al., 2018).

Nonetheless, we demonstrated that a quantitative mapping of the fitness landscape using deep sequencing is an extremely powerful method to study protein evolution. Similar studies for more complex conditions, and of different model proteins, will be key for further developing our understanding of protein evolution, in particular, the molecular mechanisms underlying key evolutionary phenomenon like mutational robustness, epistasis and evolvability. Ultimately, better knowledge of phenotype-fitness relationships will also improve our ability to predict the evolution of proteins, such as antibiotic resistance genes, in a complex, natural environment. Such knowledge should also lead to better design and engineering strategies to generate novel proteins and metabolic pathways.

Acknowledgements

We thank Adrian Serohijos, Pouria Dasmeh and members of the Tokuriki lab for insightful comments on the manuscript. We thank the Canadian Institute of Health Research (CIHR) Foundation Grant (FDN-148437) for the financial support. Nobuhiko Tokuriki is a Michael Smith Foundation of Health Research (MSFHR) career investigator.

Author Contributions

John Z Chen: Conceptualization, Software, Formal analysis, Investigation, Methodology, Writing - original draft, Writing - review and editing; Douglas M Fowler: Software, Validation, Methodology, Writing - review and editing; Nobuhiko Tokuriki: Conceptualization, Supervision, Funding acquisition, Project administration, Writing - review and editing

Declaration of Interest

The authors declare no competing interests.

503 **Methods**

504 ***Deep mutational scanning of VIM-2***

505 The variant library of VIM-2 was generated in a previous study, and all selection experiments,
506 deep sequencing and quality control for sequencing data were conducted in the same manner as
507 before(Chen et al., 2020). Briefly, the plasmid based wild-type VIM-2 gene (*wt*VIM-2) gene was
508 mutated into a library containing all single amino acid substitutions through PCR mutagenesis
509 using a primer that replaces each codon position with a degenerate 'NNN' sequence. The library
510 of all 267 codon positions was divided into 7 groups of 39 amino acids (31 in the last group),
511 forming 117nt 'tiles' that can be sequenced with full overlapping reads using Illumina Next-seq
512 550. The libraries were transformed into *E. coli* E. cloni 10G electrocompetent cells (Lucigen)
513 using the accompanying protocol, and grown in liquid culture (Miller LB broth) without selection,
514 or selected with 4.0-256 µg/mL AMP in two-fold increments for 6 hours at 37°C. After 6 hours
515 the cell concentration of the *E. coli* culture was measured by optical density at 600 nm (OD_{600}).
516 The culture was then removed from selection by centrifugation and aspiration of the selection
517 media, then resuspended in fresh LB; this procedure was repeated for a total of three washes. The
518 washed cells were grown overnight, then the plasmid DNA was isolated using a QiaPrep 96 DNA
519 extraction kit (Qiagen). One 'sample' in this experiment encompasses a single replicate of one
520 library group of variants grown under a single condition. This process was conducted in two
521 replicates on separate days, while an additional replicate (three total) was performed for the non-
522 selected library. The third non-selected library was initially intended as a control to detect if
523 variants in the library are lost during the 6 hours of growth without selection, and was grown
524 overnight but was not subject to the 6 hours of growth at 37°C. However, we observed no notable
525 differences in variants present between the three replicates of the non-selected library and the third
526 replicate was treated the same as the other two replicates in subsequent analysis.

527 Following selection and DNA extraction, plasmid DNA from each library group was
528 amplified by PCR for 15 cycles, using primers that amplify only the 117 bp region of each group
529 and attach the universal Nextera adaptor sequence to each end of the amplicon. The amplicons are
530 extended by PCR again to include the sample indices (i7 and i5) and flow cell binding sequence,
531 then sequenced using a NextSeq 550 sequencing system (Illumina, Inc.). The *wt*VIM-2 plasmid
532 was also included in the sequencing run to measure deep sequencing error rates. The raw
533 sequencing data can be found in the NCBI Sequencing Read Archive (SRA) (BioProject accession:
534 **PRJNA634597**). Subsequent processing of deep sequencing data and calculation of fitness scores
535 are conducted using our previously published in-house Python scripts
536 (https://github.com/johnchen93/DMS-FastQ-processing). Using these scripts after sequencing, the
537 forward and reverse paired-end sequencing reads are merged and the posterior quality score of
538 each position is calculate from the quality scores of the two reads (Edgar and Flyvbjerg, 2015).
539 Sequences that have more than 20 mismatches between forward and reverse reads, or contains any
540 position with a posterior quality score less than 10 were discarded. The median number of reads
541 that pass the merging process is ~660,000. Only one of the samples from the non-selected library
542 and 7 samples from the selected conditions had less than 100,000 reads after filtering (all limited
543 to replicate 1), and these low read samples were handled accordingly in the subsequent applications
544 (see sections below on '**Calculation of fitness scores**' and '**Population size estimates using deep**
545 **sequencing data**').

546 The merged sequences were used to identify mutations, and sequences with more than one
547 codon mutated were discarded. To further ensure the quality of the variants we identified, the

548 sequencing errors observed when sequencing *wt*VIM-2 was used to calculate the expected
549 frequency of single amino acid variants arising from sequencing errors alone. Then, the non-
550 selected library was filtered so that variants with frequencies lower than 2 × of the expected
551 frequency from sequencing errors or have a read count less than 5 are excluded from further
552 analysis. Only the variants that pass all of the above quality control measures are used for further
553 calculations. The frequencies of each amino acid variant were calculated using **equation (6)** for
554 each variant. Variants that passed quality filtering in the non-selected library but were no longer
555 present in the selected libraries were given a dummy count of 1 to estimate the lowest possible
556 frequency.

557 $$freq_{var} = read\ count_{var} / read\ count_{total} \quad (6)$$

558 ***Calculation of fitness scores***

559 The variant frequencies were used to calculate fitness scores for selection conducted at each AMP
560 concentration using **equation (1)**, where each selected library is matched with the non-selected
561 replicate that was grown in the same replicate for calculations; the extra non-selected library
562 replicate was not considered. Fitness scores were averaged by taking the arithmetic mean and
563 averaged scores below -4.0 were set to a minimum of -4.0 for further analysis. The non-selected
564 library in group 1 of replicate 1 contained less than 10,000 reads after merging and only ~25% of
565 the variants were detected after filtering, meaning that most variants in positions 1-39 only have
566 fitness scores from replicate 2. Additionally, of the 7 samples selected with AMP that contained
567 less than 100,000 reads after FastQ merging, 2 samples were retained after scoring as they share
568 correlation R^2 greater than 0.8 with the other replicate while the scores from the other 5 samples
569 were excluded; this includes library group 2 (selected with 16 and 128 µg/mL AMP), 4 (8 µg/mL
570 AMP), 6 (16 µg/mL AMP) and 7 (128 µg/mL AMP) from replicate 1.

571 ***Population size estimates using deep sequencing data***

572 Population sizes were calculated using the variant frequencies and the OD_{600} of the library after
573 selection. The code used to calculate the population size and EC_{50} can be found at
574 https://github.com/johnchen93/DMS-EC50. To consolidate the 3 replicates of the non-selected
575 library, the frequency of each variant was averaged between any replicates where the variant
576 passed quality filtering; frequencies from selected libraries were kept as separate replicates. The
577 frequencies of each variant were then used to estimate the population size of cells containing that
578 variant by multiplying the frequency with the OD_{600} observed immediately after the 6 hours of
579 selection, then multiplied again with an estimated cell count of 10^9 cells per 1 unit of OD_{600} using
580 **equation (7)**; all cultures were selected at the same total volume of 1mL.

581 $$N_{var} = OD_{600} \times 10^9 \text{ cells per } OD_{600} \times freq_{var} \quad (7)$$

582 It is possible for the frequency to represent an unrealistic cell count, but in our case even purged
583 variants that have been given a dummy count of 1 will still have population estimates of at least
584 300-500 cells. Samples with less than 100,000 reads after merging were excluded from this
585 analysis, which consist of library groups 2 (at 16 and 128 µg/mL AMP), 4 (8 µg/mL AMP), 6 (8,
586 16 and 32 µg/mL AMP) and 7 (128 µg/mL AMP) in replicate 1.

587 ***Calculation of EC_{50} for each variant***

588 The estimated population size of each variant (y-axis) is plotted against the AMP concentration
589 (x-axis) to generate a dose-response curve. A sigmoidal fit is applied to the curve using the

590 scipy.optimize package's 'curve_fit' function to calculate the EC_{50} using **equation (2)**. In the
591 initial fit, $\underline{N}_{max}$ is estimated as the average population size of the variant in the non-selected libraries
592 or the library selected at the lowest AMP concentration, whichever is the higher value. N_{min}, k and
593 EC_{50} were estimated as 0, -1 and 32µg/mL AMP, respectively. The fitting was performed
594 separately for each replicate.

595 All dose-response curves were plotted regardless of whether a successful fit was made and
596 each curve fit for each replicate was manually curated. Curves where the EC_{50} lies in a region with
597 data missing in two consecutive concentrations are discarded. Some variants exhibit EC_{50}
598 transitions outside the measured range of concentrations (4 - 256 µg/mL AMP), in which case the
599 EC_{50} is estimated as the lowest (4 µg/mL AMP) and highest AMP concentration (256 µg/mL
600 AMP) if the EC_{50} transition occurs past the lowest and highest AMP concentration, respectively.
601 To merge the EC_{50} between replicates, we took the arithmetic average of the $\log_2$ (EC_{50}) (**Table**
602 **S1**).

603 ***Measuring EC_{50} using conventional drug-response assays***

604 Previously, we measured the EC_{50} for AMP of 39 amino acid variants that were isolated and
605 assayed individually(Chen et al., 2020). In this study, we determined the EC_{50} of an additional 48
606 amino acid variants (63 unique codons) and measured their EC_{50} using the same method. Variant
607 libraries at positions across the length of the VIM-2 sequence (positions 5, 16, 42, 73, 96, 153, 162,
608 173, 201, 209 and 251) were transformed into *E. coli*, plated on LB-agar plates and single colonies
609 were picked, cultured and stored in glycerol stocks in a 96 well plate; for each position, we picked
610 8 colonies. Two colonies each of *E. coli* transformed with *wt*VIM-2 or empty vector were also
611 isolated in the same manner and placed in the same plate as controls. The identity of each isolated
612 variant was confirmed through Sanger sequencing.

613 Antibiotic selection was conducted using the same procedure as the DMS selection
614 presented above, where a liquid culture of the E. coli containing each isolated variant is grown for
615 6 hrs under selection with 2.0-256 ug/mL AMP at 2-fold increments, at 37°C. The OD_{600} of each
616 culture is measured after selection and plotted against the AMP concentration to generate a dose-
617 response curve. The curve was fitted with the sigmoidal **equation (2)** to calculate the EC_{50}, where
618 OD_{600} was used directly in place of N_{var}, while N_{min} and N_{max} were estimated as 0 and the OD_{600}
619 from growth without AMP, respectively. Final EC_{50} values of the variants picked in this study can
620 be found in **Table S3**.

621 ***Phenotype-fitness relationship at single AMP concentrations***

622 The fitness scores (y-axis) at each AMP concentration were plotted against the EC_{50} (x-
623 axis). The relationship between fitness scores and EC_{50} at each AMP concentration was fitted to
624 the sigmoidal curve in **equation (3)**. The initial estimates for f_{max}, f_{min} and n are 2.0, -4.0 and 1.0
625 respectively for all AMP concentrations. The initial estimate for X_{50}, the inflection point of the
626 curve, was set to be the same as the AMP concentration from which the dataset was measured.

627 The boundaries of the transition between the lower and upper fitness plateaus (X_1-X_{99}) were
628 calculated as the EC_{50} values at 1% and 99% fitness relative to the range between the fitted f_{max}
629 and f_{min} of each AMP concentration.

630 The linear region of the sigmoidal curve for the DMS fitness scores was calculated using
631 **equations (8-11)**, based on the final fitted values for each AMP concentration (Sebaugh and
632 McCray, 2003).

633 $Y_{lower} = f_{max} + (\frac{f_{min}-f_{max}}{1+1/4.6805})$ (8) $Y_{upper} = f_{max} + (\frac{f_{min}-f_{max}}{1+4.6805})$ (9)

634 $X_{lower} = X_{50}(\frac{f_{min}-Y_{lower}}{Y_{lower}-f_{max}})^{\frac{1}{n}}$ (10) $X_{upper} = X_{50}(\frac{f_{min}-Y_{upper}}{Y_{upper}-f_{max}})^{\frac{1}{n}}$ (11)

635 In this case, X_{lower} and X_{upper} indicate the start and the end of the linear range in terms of EC_{50} while
636 Y_{lower} and Y_{upper} indicate the start and end of the linear range in terms of fitness score.

637 ***General model for the phenotype-fitness landscape of VIM-2***

638 To generate the model, we start by connecting the sigmoidal behavior of the phenotype-fitness
639 relationships across different AMP concentrations into the general form in **equation (12)**.

640 $$f(EC_{50}, [AMP]) = f_{min} + \frac{f_{max} - f_{min}}{1+\left(\frac{X_{50}([AMP])}{EC_{50}}\right)^{n}} \quad (12)$$

641 All parameters are functionally the same as **equation (3)** except the inflection point of the curve
642 (X_{50}), which is now a linear function of AMP as shown in **equation (4)**. The Hill coefficient (n) is
643 5.7, the average where the linear range of the sigmoidal transition was not limited by the selection
644 concentration (8-128 µg/mL). Furthermore, we normalize the observed fitness scores to the f_{min}
645 and f_{max} determined from the sigmoidal fits at each AMP concentration in **Table 1** using **equation**
646 **(13)**, such that a fitness of 0 represents the minimum fitness and a fitness of 1 represents the
647 maximum fitness.

648 $$f_{obs\,norm} = \frac{f_{obs}-f_{min}}{f_{max}-f_{min}} \quad (13)$$

649 This allows us to arrive at the final form of the model presented in **equation (5)**.

650 ***Calculation of MIC and MSC from the general model***

651 We calculated the minimal inhibitory concentration (MIC) of *wt*VIM-2 as the AMP concentration
652 at which fitness is reduced to a threshold $t = 0.01$ (1% of f_{max}), according to **equation (14)** which
653 is simply a rearrangement of **equation (5)**.

654 $$[AMP] = \frac{\left[\left(\frac{1}{t}-1\right)^{\frac{1}{5.7}} \times EC_{50} + 1.29\right]}{0.93} \quad (14)$$

655 Similarly, we calculated the minimal selective concentration (MSC) as the AMP concentration at
656 99% fitness (t = 0.99) for wtVIM-2, such that no variant can exhibit more than a 1% increase in
657 fitness over *wt*VIM-2 at this concentration of AMP or lower.

658 ***Nonspecific epistasis of VIM-2 variants***

659 We randomly sampled 200 missense variants, extracted their EC_{50} values, and generated all
660 pairwise mutational combinations (${}_{200}C_2$ = 19900 mutant pairs). Since the analysis was focused on
661 the numerical behavior rather than the physical behavior of the enzyme, mutations from the same
662 residue positions were allowed to be combined, though this rarely occurred as only 88 out of 19900
663 combinations (~0.4%) were from combining the same positions. From these mutant pairs, we used
664 **equation (5)** to calculate the expected fitness effect due to additive changes in fitness relative to
665 *wt*VIM-2. First, the relative change in fitness for each single variant in the pair is calculated
666 according to **equation (15)**, then added together as shown in **equation (16)**.

667 $$\Delta f_i = f(EC_{50-i}, [AMP]) - f(EC_{50-wt}, [AMP]) \qquad i = 1,2 \qquad (15)$$

668 $$\Delta f_{exp} = \Delta f_1 + \Delta f_2 \qquad (16)$$

669 Since the model explicitly defines 0 as the minimum fitness and 1 as the maximum fitness, $\Delta f\ exp$
670 is limited to the maximum change that can occur within this range (-1 to 1).

671 We also calculate the observed fitness effect according to **equation (5)** as an additive effect
672 in the $\log_2(EC_{50})$ change relative to *wt*VIM-2. First, the relative change in EC_{50} is calculated in
673 **equation (17)** and the additive EC_{50} effects are used to calculate the new EC_{50} in **equation (18)**.

674 $$\Delta \log_2(EC_{50-i}) = \log_2(EC_{50-i}) - \log_2(EC_{50-wt}) \qquad i = 1,2 \qquad (17)$$

675 $$EC_{50-1,2} = 2^{(\log_2(EC_{50-wt}) + \Delta\log_2(EC_{50-1}) + \Delta\log_2(EC_{50-2}))} \qquad (18)$$

676 We limit $EC_{50\text{-}1,2}$ to be no lower than 2.0 (4.0 µg/mL untransformed), as an [AMP] of 2.0 µg/mL
677 is not selective and is not accounted for by **equation (5)**. The observed fitness effect is then
678 calculated according to **equation (19)**.

679 $$\Delta f_{obs} = f\left(EC_{50-1,2}, [AMP]\right) - f(EC_{50-wt}, [AMP]) \qquad (19)$$

680 To identify significant epistatic effects, we use the synonymous fitness distribution at each [AMP]
681 as a measure of expected measurement noise. When $|\Delta f_{obs} - \Delta f_{exp}| < 1.96 \times$ S.D. of the synonymous
682 fitness (after normalization by **equation (13)**), the effect of the mutational pair is considered not
683 significantly epistatic. For mutational pairs with significant epistasis, they are classified as
684 exhibiting magnitude epstasis if the Δf_{obs} and Δf_{exp} have the same sign, or classified as exhibiting
685 sign epistasis if Δf_{obs} and Δf_{exp} have opposite signs. For magnitude epistasis, positive magnitude
686 epistasis is defined as when $\Delta f_{obs} > \Delta f_{exp}$ and $\Delta f_{exp} > 0$ or $\Delta f_{obs} < \Delta f_{exp}$ and $\Delta f_{exp} < 0$, while negative
687 magnitude epistasis is defined when $\Delta f_{obs} < \Delta f_{exp}$ and $\Delta f_{exp} > 0$ or $\Delta f_{obs} > \Delta f_{exp}$ and $\Delta f_{exp} < 0$. For
688 sign epistasis, positive sign epistasis is defined as when $\Delta f_{obs} > 0$ and $\Delta f_{exp} < 0$, while negative sign
689 epistasis is defined as when $\Delta f_{obs} < 0$ and $\Delta f_{exp} > 0$.

690 **References**

691 Andersson, D.I., Hughes, D., 2014. Microbiological effects of sublethal levels of antibiotics.
692 Nature Reviews Microbiology 12, 465–478. https://doi.org/10.1038/nrmicro3270
693 Araya, C.L., Fowler, D.M., Chen, W., Muniez, I., Kelly, J.W., Fields, S., 2012. A fundamental
694 protein property, thermodynamic stability, revealed solely from large-scale measurements
695 of protein function. Proceedings of the National Academy of Sciences 109, 16858–
696 16863. https://doi.org/10.1073/pnas.1209751109
697 Baier, F., Hong, N., Yang, G., Pabis, A., Miton, C.M., Barrozo, A., Carr, P.D., Kamerlin, S.C.,
698 Jackson, C.J., Tokuriki, N., 2019. Cryptic genetic variation shapes the adaptive
699 evolutionary potential of enzymes. Elife 8. https://doi.org/10.7554/eLife.40789
700 Bank, C., Hietpas, R.T., Jensen, J.D., Bolon, D.N.A., 2015. A systematic survey of an intragenic
701 epistatic landscape. Mol Biol Evol 32, 229–238. https://doi.org/10.1093/molbev/msu301
702 Baquero, F., Negri, M.C., Morosini, M.I., Blázquez, J., 1998. Selection of very small differences
703 in bacterial evolution. Int Microbiol 1, 295–300.
704 Bershtein, S., Mu, W., Serohijos, A.W.R., Zhou, J., Shakhnovich, E.I., 2013. Protein Quality
705 Control Acts on Folding Intermediates to Shape the Effects of Mutations on Organismal
706 Fitness. Molecular Cell 49, 133–144. https://doi.org/10.1016/j.molcel.2012.11.004
707 Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., Tawfik, D.S., 2006. Robustness–epistasis
708 link shapes the fitness landscape of a randomly drifting protein. Nature 444, 929–932.
709 https://doi.org/10.1038/nature05385
710 Bershtein, S., Serohijos, A.W., Shakhnovich, E.I., 2017. Bridging the physical scales in
711 evolutionary biology: from protein sequence space to fitness of organisms and
712 populations. Curr. Opin. Struct. Biol. 42, 31–40.
713 https://doi.org/10.1016/j.sbi.2016.10.013
714 Blondeau, J.M., 2009. New concepts in antimicrobial susceptibility testing: the mutant
715 prevention concentration and mutant selection window approach. Vet Dermatol 20, 383–
716 396. https://doi.org/10.1111/j.1365-3164.2009.00856.x
717 Chen, J.Z., Fowler, D.M., Tokuriki, N., 2020. Comprehensive exploration of the translocation,
718 stability and substrate recognition requirements in VIM-2 lactamase. eLife 9, e56707.
719 https://doi.org/10.7554/eLife.56707
720 Dalziel, A.C., Rogers, S.M., Schulte, P.M., 2009. Linking genotypes to phenotypes and fitness:
721 how mechanistic biology can inform molecular ecology. Mol. Ecol. 18, 4997–5017.
722 https://doi.org/10.1111/j.1365-294X.2009.04427.x
723 Das, S.G., Direito, S.O., Waclaw, B., Allen, R.J., Krug, J., 2020. Predictable properties of fitness
724 landscapes induced by adaptational tradeoffs. eLife 9, e55155.
725 https://doi.org/10.7554/eLife.55155
726 Dasmeh, P., Serohijos, A.W.R., 2018. Estimating the contribution of folding stability to
727 nonspecific epistasis in protein evolution. Proteins: Structure, Function, and
728 Bioinformatics 86, 1242–1250. https://doi.org/10.1002/prot.25588
729 Dean, A.M., 1995. A molecular investigation of genotype by environment interactions. Genetics
730 139, 19–33.
731 DePristo, M.A., Weinreich, D.M., Hartl, D.L., 2005. Missense meanderings in sequence space: a
732 biophysical view of protein evolution. Nature Reviews Genetics 6, 678.
733 https://doi.org/10.1038/nrg1672
734 Diss, G., Lehner, B., 2018. The genetic landscape of a physical interaction. eLife 7, e32472.
735 https://doi.org/10.7554/eLife.32472

Domingo, J., Baeza-Centurion, P., Lehner, B., 2019. The Causes and Consequences of Genetic Interactions (Epistasis). Annu Rev Genomics Hum Genet 20, 433–460. https://doi.org/10.1146/annurev-genom-083118-014857

Drlica, K., Zhao, X., 2007. Mutant Selection Window Hypothesis Updated. Clin Infect Dis 44, 681–688. https://doi.org/10.1086/511642

Edgar, R.C., Flyvbjerg, H., 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31, 3476–3482. https://doi.org/10.1093/bioinformatics/btv401

Firnberg, E., Labonte, J.W., Gray, J.J., Ostermeier, M., 2014. A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape. Molecular Biology and Evolution 31, 1581–1592. https://doi.org/10.1093/molbev/msu081

Flynn, J.M., Rossouw, A., Cote-Hammarlof, P., Fragata, I., Mavor, D., Hollins, C., III, Bank, C., Bolon, D.N., 2020. Comprehensive fitness maps of Hsp90 show widespread environmental dependence. eLife 9, e53810. https://doi.org/10.7554/eLife.53810

Fowler, D.M., Araya, C.L., Fleishman, S.J., Kellogg, E.H., Stephany, J.J., Baker, D., Fields, S., 2010. High-resolution mapping of protein sequence-function relationships. Nature Methods 7, 741–746. https://doi.org/10.1038/nmeth.1492

Fowler, D.M., Stephany, J.J., Fields, S., 2014. Measuring the activity of protein variants on a large scale using deep mutational scanning. Nat Protoc 9, 2267–2284. https://doi.org/10.1038/nprot.2014.153

Fröhlich, C., Gama, J.A., Harms, K., Hirvonen, V.H.A., Lund, B.A., Kamp, M.W. van der, Johnsen, P.J., Samuelsen, Ø., Leiros, H.-K.S., 2020. Cryptic β-lactamase evolution is driven by low β-lactam concentrations. bioRxiv 2020.12.01.404343. https://doi.org/10.1101/2020.12.01.404343

Gray, V.E., Hause, R.J., Luebeck, J., Shendure, J., Fowler, D.M., 2018. Quantitative Missense Variant Effect Prediction Using Large-Scale Mutagenesis Data. Cell Syst 6, 116-124.e3. https://doi.org/10.1016/j.cels.2017.11.003

Gullberg, E., Cao, S., Berg, O.G., Ilbäck, C., Sandegren, L., Hughes, D., Andersson, D.I., 2011. Selection of resistant bacteria at very low antibiotic concentrations. PLoS Pathog. 7, e1002158. https://doi.org/10.1371/journal.ppat.1002158

Hartl, D.L., Dykhuizen, D.E., Dean, A.M., 1985. Limits of adaptation: the evolution of selective neutrality. Genetics 111, 655–674.

Hietpas, R.T., Jensen, J.D., Bolon, D.N.A., 2011. Experimental illumination of a fitness landscape. Proc. Natl. Acad. Sci. U.S.A. 108, 7896–7901. https://doi.org/10.1073/pnas.1016024108

Horovitz, A., Fleisher, R.C., Mondal, T., 2019. Double-mutant cycles: new directions and applications. Current Opinion in Structural Biology, Cryo electron microscopy ● Biophysical and computational methods ● Biophysical and computational methods - Part B 58, 10–17. https://doi.org/10.1016/j.sbi.2019.03.025

Jacquier, H., Birgy, A., Nagard, H.L., 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. PNAS 110, 13067–13072. https://doi.org/10.1073/pnas.1215206110

Jasinska, W., Manhart, M., Lerner, J., Gauthier, L., Serohijos, A.W.R., Bershtein, S., 2020. Chromosomal barcoding of E. coli populations reveals lineage diversity dynamics at high

resolution. Nature Ecology & Evolution 4, 437–452. https://doi.org/10.1038/s41559-020-1103-z

Kaltenbach, M., Tokuriki, N., 2014. Dynamics and constraints of enzyme evolution. J. Exp. Zool. B Mol. Dev. Evol. 322, 468–487. https://doi.org/10.1002/jez.b.22562

Kemble, H., Eisenhauer, C., Couce, A., Chapron, A., Magnan, M., Gautier, G., Nagard, H.L., Nghe, P., Tenaillon, O., 2020. Flux, toxicity, and expression costs generate complex genetic interactions in a metabolic pathway. Science Advances 6, eabb2236. https://doi.org/10.1126/sciadv.abb2236

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

Klesmith, J.R., Bacik, J.-P., Wrenbeck, E.E., Michalczyk, R., Whitehead, T.A., 2017. Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. Proceedings of the National Academy of Sciences 114, 2265–2270. https://doi.org/10.1073/pnas.1614437114

Liu, A., Fong, A., Becket, E., Yuan, J., Tamae, C., Medrano, L., Maiz, M., Wahba, C., Lee, C., Lee, K., Tran, K.P., Yang, H., Hoffman, R.M., Salih, A., Miller, J.H., 2011. Selective advantage of resistant strains at trace levels of antibiotics: a simple and ultrasensitive color test for detection of antibiotics and genotoxic agents. Antimicrob. Agents Chemother. 55, 1204–1210. https://doi.org/10.1128/AAC.01182-10

Lundin, E., Tang, P.-C., Guy, L., Näsvall, J., Andersson, D.I., 2018. Experimental Determination and Prediction of the Fitness Effects of Random Point Mutations in the Biosynthetic Enzyme HisA. Mol Biol Evol 35, 704–718. https://doi.org/10.1093/molbev/msx325

Lunzer, M., Miller, S.P., Felsheim, R., Dean, A.M., 2005. The Biochemical Architecture of an Ancient Adaptive Landscape. Science 310, 499–501. https://doi.org/10.1126/science.1115649

Mavor, D., Barlow, K., Thompson, S., Barad, B.A., Bonny, A.R., Cario, C.L., Gaskins, G., Liu, Z., Deming, L., Axen, S.D., Caceres, E., Chen, W., Cuesta, A., Gate, R.E., Green, E.M., Hulce, K.R., Ji, W., Kenner, L.R., Mensa, B., Morinishi, L.S., Moss, S.M., Mravic, M., Muir, R.K., Niekamp, S., Nnadi, C.I., Palovcak, E., Poss, E.M., Ross, T.D., Salcedo, E.C., See, S.K., Subramaniam, M., Wong, A.W., Li, J., Thorn, K.S., Conchúir, S.Ó., Roscoe, B.P., Chow, E.D., DeRisi, J.L., Kortemme, T., Bolon, D.N., Fraser, J.S., 2016. Determination of ubiquitin fitness landscapes under different chemical stresses in a classroom setting. Elife 5. https://doi.org/10.7554/eLife.15802

Mavor, D., Barlow, K.A., Asarnow, D., Birman, Y., Britain, D., Chen, W., Green, E.M., Kenner, L.R., Mensa, B., Morinishi, L.S., Nelson, C.A., Poss, E.M., Suresh, P., Tian, R., Arhar, T., Ary, B.E., Bauer, D.P., Bergman, I.D., Brunetti, R.M., Chio, C.M., Dai, S.A., Dickinson, M.S., Elledge, S.K., Helsell, C.V.M., Hendel, N.L., Kang, E., Kern, N., Khoroshkin, M.S., Kirkemo, L.L., Lewis, G.R., Lou, K., Marin, W.M., Maxwell, A.M., McTigue, P.F., Myers-Turnbull, D., Nagy, T.L., Natale, A.M., Oltion, K., Pourmal, S., Reder, G.K., Rettko, N.J., Rohweder, P.J., Schwarz, D.M.C., Tan, S.K., Thomas, P.V., Tibble, R.W., Town, J.P., Tsai, M.K., Ugur, F.S., Wassarman, D.R., Wolff, A.M., Wu, T.S., Bogdanoff, D., Li, J., Thorn, K.S., O'Conchúir, S., Swaney, D.L., Chow, E.D., Madhani, H.D., Redding, S., Bolon, D.N., Kortemme, T., DeRisi, J.L., Kampmann, M., Fraser, J.S., 2018. Extending chemical perturbations of the ubiquitin fitness landscape in

a classroom setting reveals new constraints on sequence tolerance. Biol Open 7. https://doi.org/10.1242/bio.036103

Meer, J.-Y. van der, Poddar, H., Baas, B.-J., Miao, Y., Rahimi, M., Kunzendorf, A., Merkerk, R. van, Tepper, P.G., Geertsema, E.M., Thunnissen, A.-M.W.H., Quax, W.J., Poelarends, G.J., 2016. Using mutability landscapes of a promiscuous tautomerase to guide the engineering of enantioselective Michaelases. Nat Commun 7, 1–16. https://doi.org/10.1038/ncomms10911

Mehlhoff, J.D., Stearns, F.W., Rohm, D., Wang, B., Tsou, E.-Y., Dutta, N., Hsiao, M.-H., Gonzalez, C.E., Rubin, A.F., Ostermeier, M., 2020. Collateral fitness effects of mutations. PNAS 117, 11597–11607. https://doi.org/10.1073/pnas.1918680117

Melnikov, A., Rogov, P., Wang, L., Gnirke, A., Mikkelsen, T.S., 2014. Comprehensive mutational scanning of a kinase *in vivo* reveals substrate-dependent fitness landscapes. Nucleic Acids Research 42, e112–e112. https://doi.org/10.1093/nar/gku511

Miton, C.M., Tokuriki, N., 2016. How mutational epistasis impairs predictability in protein evolution and design. Protein Sci. 25, 1260–1272. https://doi.org/10.1002/pro.2876

Noda-García, L., Davidi, D., Korenblum, E., Elazar, A., Putintseva, E., Aharoni, A., Tawfik, D.S., 2019. Chance and pleiotropy dominate genetic diversity in complex bacterial environments. Nat Microbiol 4, 1221–1230. https://doi.org/10.1038/s41564-019-0412-y

Olson, C.A., Wu, N.C., 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

Pál, C., Papp, B., Lercher, M.J., 2006. An integrated view of protein evolution. Nat. Rev. Genet. 7, 337–348. https://doi.org/10.1038/nrg1838

Palmer, A.C., Toprak, E., Baym, M., Kim, S., Veres, A., Bershtein, S., Kishony, R., 2015. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. Nat Commun 6, 7385. https://doi.org/10.1038/ncomms8385

Pokusaeva, V.O., Usmanova, D.R., Putintseva, E.V., Espinar, L., Sarkisyan, K.S., Mishin, A.S., Bogatyreva, N.S., Ivankov, D.N., Akopyan, A.V., Avvakumov, S.Y., Povolotskaya, I.S., Filion, G.J., Carey, L.B., Kondrashov, F.A., 2019. An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. PLOS Genetics 15, e1008079. https://doi.org/10.1371/journal.pgen.1008079

Rockah-Shmuel, L., Tóth-Petróczy, Á., Tawfik, D.S., 2015. Systematic Mapping of Protein Mutational Space by Prolonged Drift Reveals the Deleterious Effects of Seemingly Neutral Mutations. PLoS Comput. Biol. 11, e1004421. https://doi.org/10.1371/journal.pcbi.1004421

Salinas, V.H., Ranganathan, R., 2018. Coevolution-based inference of amino acid interactions underlying protein function. eLife 7, e34300. https://doi.org/10.7554/eLife.34300

Sarkisyan, K.S., Bolotin, D.A., Meer, M.V., Usmanova, D.R., Mishin, A.S., Sharonov, G.V., Ivankov, D.N., Bozhanova, N.G., Baranov, M.S., Soylemez, O., Bogatyreva, N.S., Vlasov, P.K., Egorov, E.S., Logacheva, M.D., Kondrashov, A.S., Chudakov, D.M., Putintseva, E.V., Mamedov, I.Z., Tawfik, D.S., Lukyanov, K.A., Kondrashov, F.A., 2016. Local fitness landscape of the green fluorescent protein. Nature 533, 397–401. https://doi.org/10.1038/nature17995

Sebaugh, J.L., McCray, P.D., 2003. Defining the linear portion of a sigmoid-shaped curve: bend points. Pharmaceutical Statistics 2, 167–174. https://doi.org/10.1002/pst.62

871 Starita, L.M., Pruneda, J.N., Lo, R.S., Fowler, D.M., Kim, H.J., Hiatt, J.B., Shendure, J.,
872 Brzovic, P.S., Fields, S., Klevit, R.E., 2013. Activity-enhancing mutations in an E3
873 ubiquitin ligase identified by high-throughput mutagenesis. Proceedings of the National
874 Academy of Sciences 110, E1263–E1272. https://doi.org/10.1073/pnas.1303309110
875 Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K.H.D., Dingens, A.S., Navarro,
876 M.J., Bowen, J.E., Tortorici, M.A., Walls, A.C., King, N.P., Veesler, D., Bloom, J.D.,
877 2020. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals
878 Constraints on Folding and ACE2 Binding. Cell 182, 1295-1310.e20.
879 https://doi.org/10.1016/j.cell.2020.08.012
880 Starr, T.N., Picton, L.K., Thornton, J.W., 2017. Alternative evolutionary histories in the
881 sequence space of an ancient protein. Nature 549, 409–413.
882 https://doi.org/10.1038/nature23902
883 Starr, T.N., Thornton, J.W., 2016. Epistasis in protein evolution. Protein Sci. 25, 1204–1218.
884 https://doi.org/10.1002/pro.2897
885 Steinberg, B., Ostermeier, M., 2016. Environmental changes bridge evolutionary valleys. Sci.
886 Adv. 2, e1500921. https://doi.org/10.1126/sciadv.1500921
887 Stiffler, M.A., Hekstra, D.R., Ranganathan, R., 2015. Evolvability as a Function of Purifying
888 Selection in TEM-1 β-Lactamase. Cell 160, 882–892.
889 https://doi.org/10.1016/j.cell.2015.01.035
890 Thompson, S., Zhang, Y., Ingle, C., Reynolds, K.A., Kortemme, T., 2020. Altered expression of
891 a quality control protease in E. coli reshapes the in vivo mutational landscape of a model
892 enzyme. eLife 9, e53476. https://doi.org/10.7554/eLife.53476
893 Tokuriki, N., Tawfik, D.S., 2009. Stability effects of mutations and protein evolvability. Current
894 Opinion in Structural Biology, Carbohydradtes and glycoconjugates / Biophysical
895 methods 19, 596–604. https://doi.org/10.1016/j.sbi.2009.08.003
896 Westhoff, S., van Leeuwe, T.M., Qachach, O., Zhang, Z., van Wezel, G.P., Rozen, D.E., 2017.
897 The evolution of no-cost resistance at sub-MIC concentrations of streptomycin in
898 Streptomyces coelicolor. ISME J 11, 1168–1178. https://doi.org/10.1038/ismej.2016.194
899 Wrenbeck, E.E., Azouz, L.R., Whitehead, T.A., 2017. Single-mutation fitness landscapes for an
900 enzyme on multiple substrates reveal specificity is globally encoded. Nature
901 Communications 8, 15695. https://doi.org/10.1038/ncomms15695
902 Yang, G., Anderson, D.W., Baier, F., Dohmen, E., Hong, N., Carr, P.D., Kamerlin, S.C.L.,
903 Jackson, C.J., Bornberg-Bauer, E., Tokuriki, N., 2019. Higher-order epistasis shapes the
904 fitness landscape of a xenobiotic-degrading enzyme. Nature Chemical Biology 15, 1120–
905 1128. https://doi.org/10.1038/s41589-019-0386-3
906 Yi, X., Dean, A.M., 2019. Adaptive Landscapes in the Age of Synthetic Biology. Mol. Biol.
907 Evol. 36, 890–907. https://doi.org/10.1093/molbev/msz004
908 Zwart, M.P., Schenk, M.F., Hwang, S., Koopmanschap, B., de Lange, N., van de Pol, L., Nga,
909 T.T.T., Szendro, I.G., Krug, J., de Visser, J.A.G.M., 2018. Unraveling the causes of
910 adaptive benefits of synonymous mutations in TEM-1 β-lactamase. Heredity 121, 406–
911 421. https://doi.org/10.1038/s41437-018-0104-z
912

Supplemental Figures

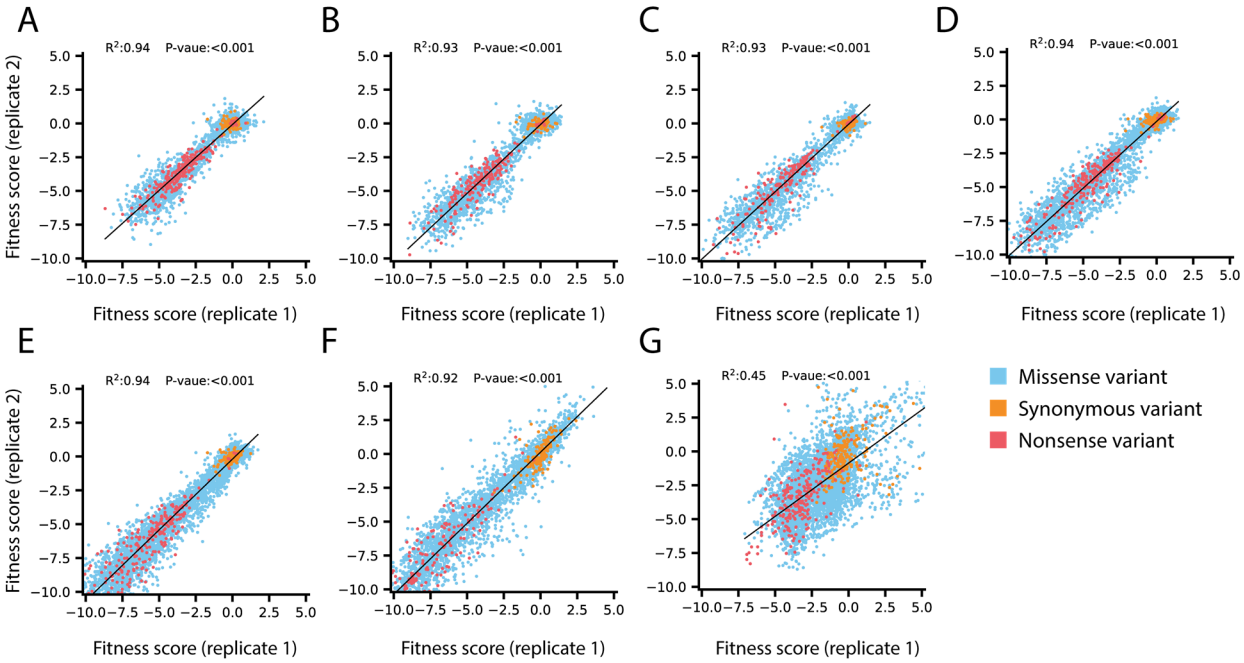


Figure S1. Replicate correlation of DMS fitness scores, Related to Figure 1. Each scatter plot shows the replicate fitness scores of all variants under selection in (**A**) 4.0, (**B**) 8.0, (**C**) 16, (**D**) 32, (**E**) 64, (**F**) 128 and (**G**) 256 µg/mL AMP. In each panel, variants are colored according to mutation type with the legend in the lower right. The solid black line indicate the line of best fit for a linear regression, with the R^2 and P-value of the regression above each plot.

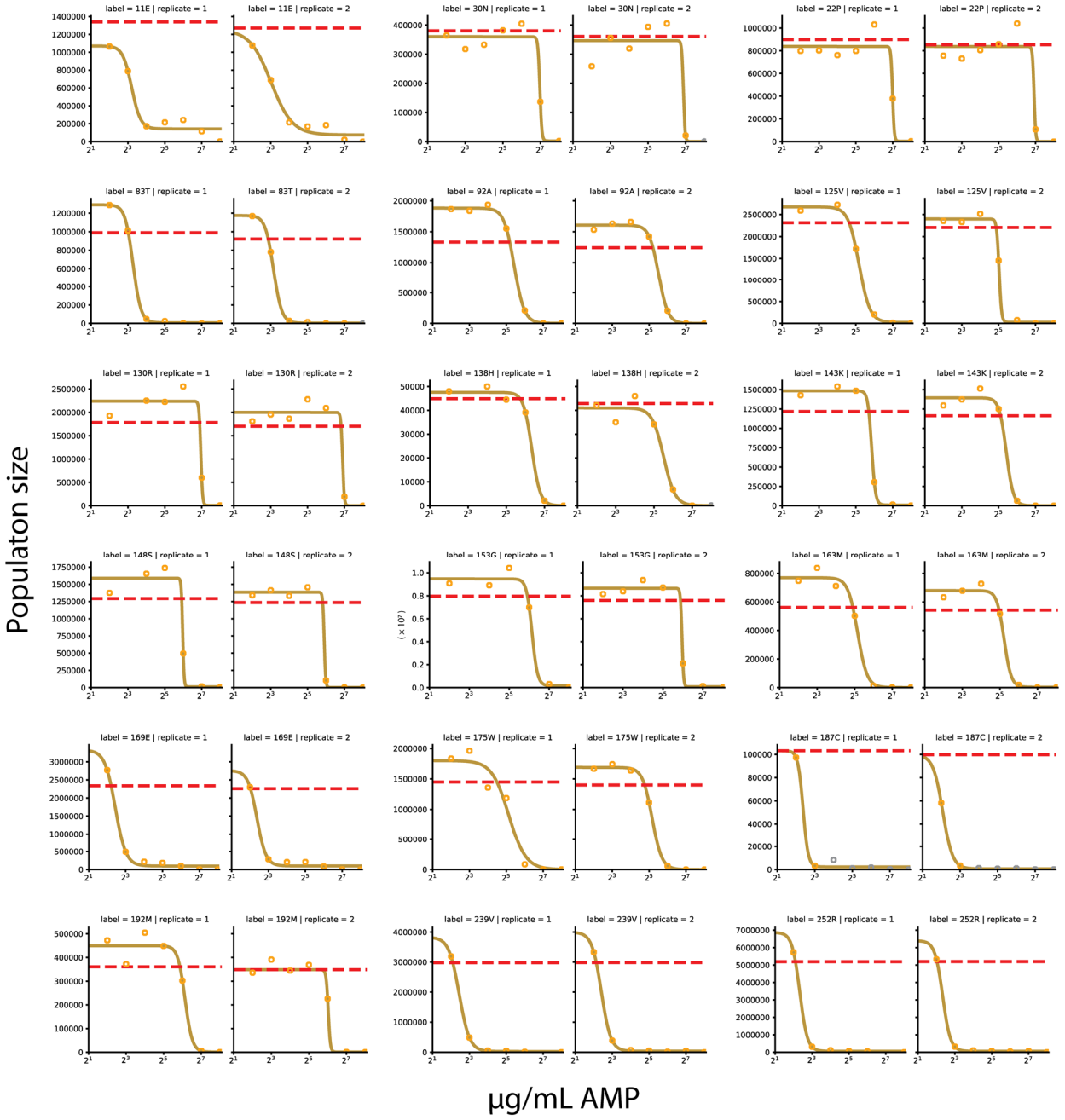


Figure S2. Example dose-response curves for *EC*$_{50}$ calculation based on deep sequencing data, Related to Figure 1. Each pair of plots represent the two replicate dose-response curves of a single variant in the library. The solid curve indicates the results of fitting to the sigmoidal **equation (2)**. The horizontal dashed line marks the average population size of the variant in the non-selected conditions. Dose-response curves shown here were randomly sampled from the entire pool of all variants with a successful fit.

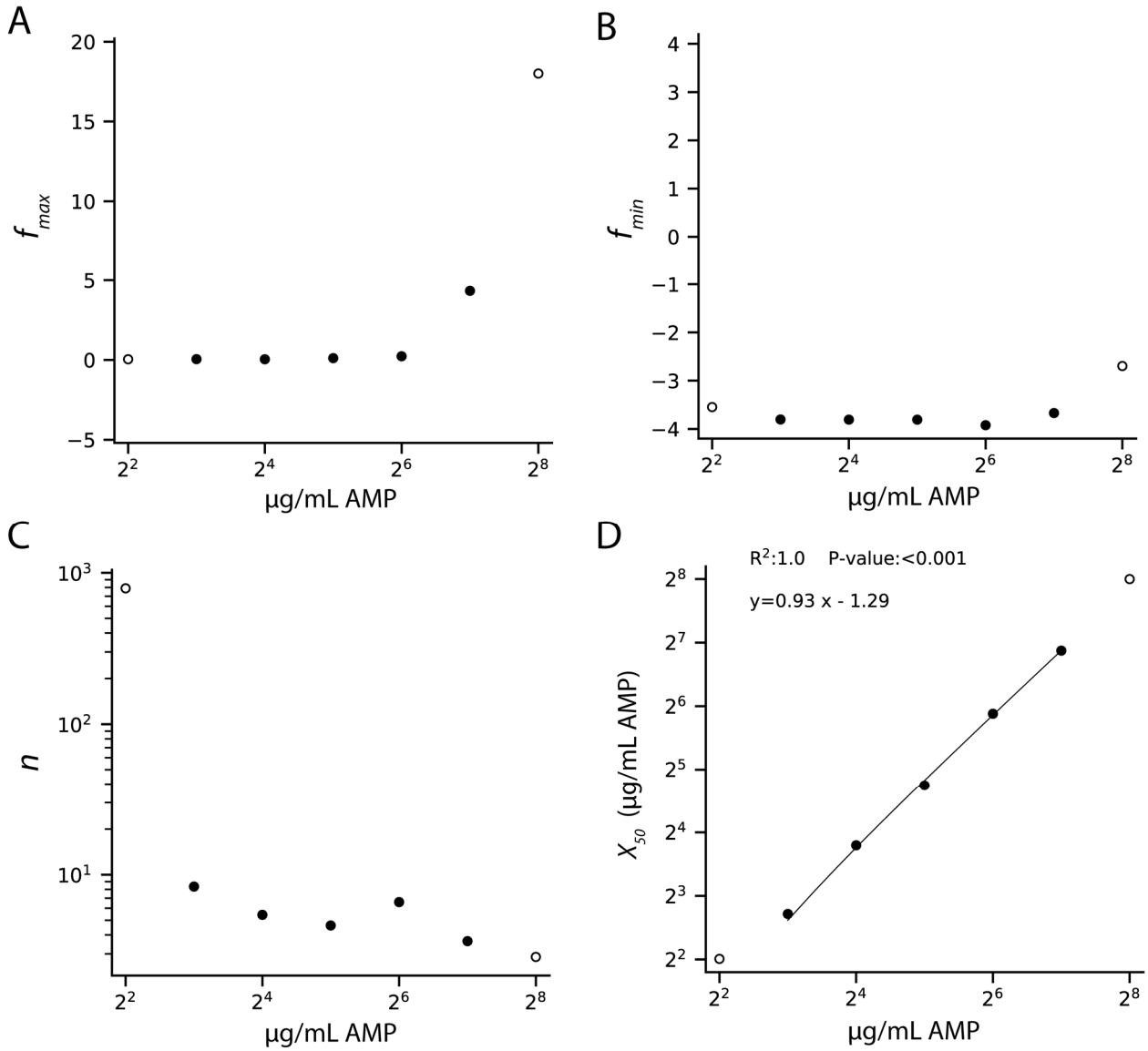


Figure S3. Fit parameters for the phenotype-fitness relationship across AMP concentrations, Related to Figure 3. Each plot shows one of the final fitted value of a parameter in the four-parameter sigmoidal curve in **equation (5)** when fitted to the fitness scores and EC_{50} values from the AMP concentration indicated in the x-axis. (**A**) Plot for maximum fitness (f_{max}) (**B**) Plot for minimum fitness (f_{min}), (**C**) Plot for Hill coefficient (n) (**D**) Plot for the inflection point (X_{50}). For all plots, solid points represent AMP concentrations where the linear portion of the fit were not limited by the AMP concentration range, while hollow points indicate the opposite. For **D**), a linear regression was conducted with the solid data points. The black line indicates the line of best fit for a linear regression, and the R^2, P-value and the equation of the fit shown at the top.