bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

On the (un-)predictability of a large intragenic fitness landscape

Claudia Bank[a,b,c,1], Sebastian Matuszewski[b,c,1], Ryan T. Hietpas[d,e], Jeffrey D. Jensen[b,c,1,*]

[a]*Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal*

[b]*School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

[c]*Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland*

[d]*Eli Lilly and Company, Indianapolis, IN 46225*

[e]*Department of Biochemistry & Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605*

Abstract

The study of fitness landscapes, which aims at mapping genotypes to fitness, is receiving ever-increasing attention. Novel experimental approaches combined with NGS methods enable accurate and extensive studies of the fitness effects of mutations – allowing us to test theoretical predictions and improve our understanding of the shape of the true underlying fitness landscape, and its implications for the predictability and repeatability of evolution.

Here, we present a uniquely large multi-allelic fitness landscape comprised of 640 engineered mutants that represent all possible combinations of 13 amino-acid changing mutations at six sites in the heat-shock protein Hsp90 in *Saccharomyces cerevisiae* under elevated salinity. Despite a prevalent pattern of negative epistasis in the landscape, we find that the global fitness peak is reached via four positively epistatic mutations. Combining traditional and extending recently proposed theoretical and statistical approaches, we quantify features of the global multi-allelic fitness landscape. Using subsets of this data, we demonstrate that extrapolation beyond a known part of the landscape is difficult owing to both local ruggedness and amino-acid specific epistatic hotspots, and that inference is additionally confounded by the non-random choice of mutations for experimental fitness landscapes.

***Key words*:** evolution, adaptation, epistasis, fitness landscapes

*To whom correspondence should be addressed. E-mail: jeffrey.jensen@epfl.ch

[1]contributed equally

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author Summary

The study of fitness landscapes is fundamentally concerned with understanding the relative roles of stochastic and deterministic processes in adaptive evolution. Here, the authors present a uniquely large and complete multi-allelic intragenic fitness landscape of 640 systematically engineered mutations in yeast Hsp90. Using a combination of traditional and recently proposed theoretical approaches, they study the accessibility of the global fitness peak, and the potential for predictability of the fitness landscape topology. They report local ruggedness of the landscape and the existence of epistatic hotspot mutations, which together make extrapolation and hence predictability inherently difficult, if mutation-specific information is not considered.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Since first proposed by Sewall Wright in 1932 [1], the idea of a fitness landscape relating genotype (or phenotype) to the reproductive success of an individual has inspired evolutionary biologists and mathematicians alike [2, 3, 4]. With the advancement of molecular and systems biology towards large and accurate data sets, it is a concept that receives increasing attention across other subfields of biology [5, 6, 7, 8, 9]. The shape of a fitness landscape carries information on the repeatability and predictability of evolution, the potential for adaptation, the importance of genetic drift, the likelihood of convergent and parallel evolution, and the degree of optimization that is (theoretically) achievable [4]. Unfortunately, the dimensionality of a complete fitness landscape of an organism – that is, a mapping of all possible combinations of mutations to their respective fitness effects – is much too high to be assessed experimentally. With the development of experimental approaches that allow for the assessment of full fitness landscapes of tens to hundreds of mutations, there is growing interest in statistics that capture the features of the landscape, and that relate an experimental landscape to theoretical landscapes of similar architecture, which have been studied extensively [10]. It is, however, unclear whether this categorization allows for an extrapolation to unknown parts of the landscape, which would be the first step towards quantifying predictability – an advance that would yield impacts far beyond the field of evolutionary biology, in particular for the clinical study of drug resistance evolution in pathogens and the development of effective vaccine and treatment strategies [8].

Existing research in this rapidly growing field comes from two sides. Firstly, different empirical landscapes have been assessed (reviewed in [4]), generally based on the combination of previously observed beneficial mutations or on the dissection of an observed adaptive walk (i.e., a combination of mutations that have been observed to be beneficial in concert). Secondly, theoretical research has proposed different landscape architectures (such as the House-of-cards, the Kaufmann NK, and the Rough-Mount-Fuji model), studied their respective properties, and developed a number of statistics that characterize the landscape and quantify the expected degree of epistasis (i.e., interaction effects between mutations) [11, 12, 13, 10, 14].

The picture that emerges from these studies is mixed, reporting both smooth [15] and rugged [16, 17] landscapes with both positive epistasis (i.e., two mutations in concert are more advantageous than expected; [18]) and negative epistasis (i.e., two mutations in concert are more deleterious than expected; [19, 20]; but see [21, 22]). Current statistical approaches have been used to rank the existing landscapes by certain features [10, 14] and to assess whether they are compatible with Fisher's Geometric Model [23]. A crucial remaining question is the extent to which the non-random choice of mutations for the experiment affects the topology of the landscape, and whether the local topology is indeed informative as to the rest of the landscape.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

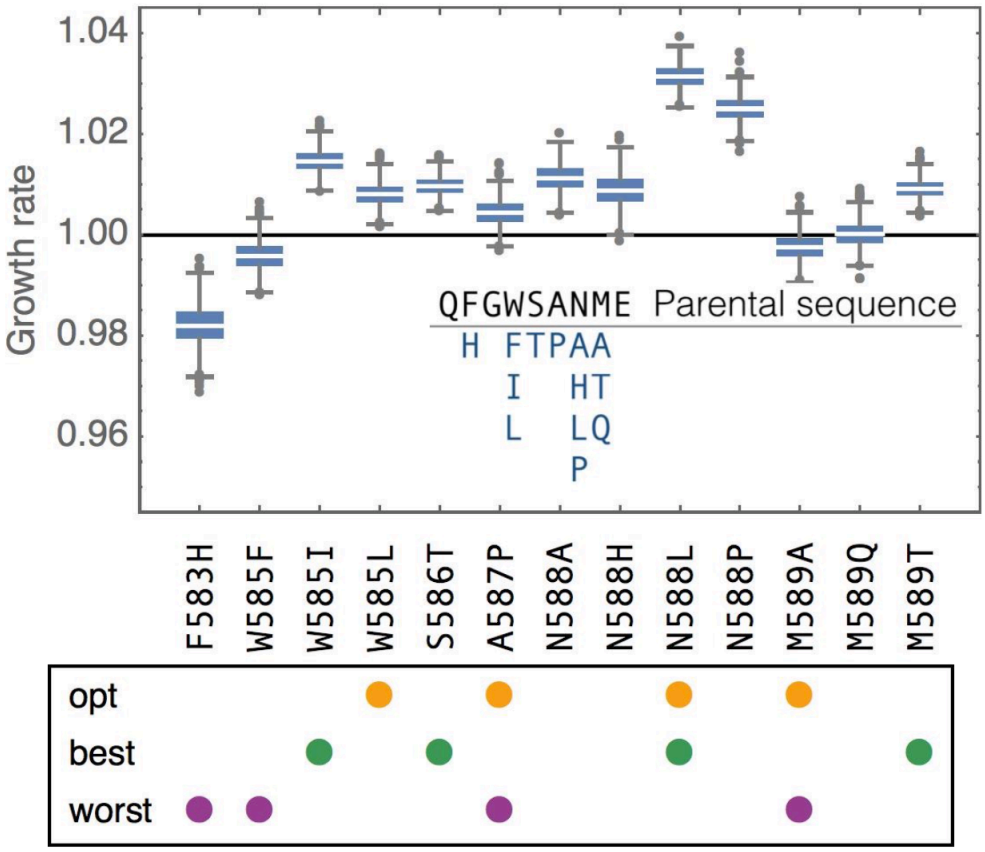


Figure 1: Individual amino-acid substitutions and their effect on the parental background in elevated salinity, obtained from 1000 samples from the posterior distribution of growth rates. The box indicates with colored dots which mutations are involved in the focal landscapes discussed throughout the main text: the four mutations leading to the global optimum ("opt"), the four individually most beneficial mutations ("best") and the four mutations with the individually lowest growth rates ("worst"). Inset: parental sequence at positions 582-590 and assessed amino acids by position.

Here, we present an intragenic fitness landscape of 640 amino-acid changing mutations in the heat shock protein Hsp90 in *Saccharomyces cerevisiae* in a challenging environment imposed by high salinity. With all possible combinations of 13 mutations of various fitness effects at 6 positions, the presented landscape is not only uniquely large but also distinguishes itself from previously published work regarding several other experimental features – namely, by its systematic and controlled experimental setup using engineered mutations of various selective effects, and by considering multiple alleles simultaneously. We begin by describing the landscape and identifying the global peak, which is reached through a highly positively epistatic combination of four mutations. Based on a variety of implemented statistical measures and models, we describe the accessibility of the peak, the pattern of epistasis, and the topology of the landscape. In order to accommodate our data, we extend several previously used models and statistics to the multi-allelic case. Using subsets of the landscape, we discuss the predictive potential of such modeling and the problem of selecting non-random mutations

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

when attempting to quantify local landscapes in order to extrapolate global features.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Material and Methods

Here, we briefly outline the materials and methods used. A more detailed treatment of the theoretical work is presented in the Supporting Information.

Data generation

Codon substitution libraries consisting of 640 combinations (single, double, triple and quadruple mutants) of 13 previously isolated individual mutants within the 582-590 region of yeast Hsp90 were generated from optimized cassette ligation strategies as previously described in [24] and cloned into the p417GPD plasmid that constitutively expresses Hsp90.

Constitutively expressed libraries of Hsp90 mutation combinations were introduced into the *S. cerevisiae* shutoff strain DBY288 (can1-100 ade2-1 his3-11,15 leu2-3,12 trp1-1 ura3-1 hsp82: :leu2 hsc82: :leu2 ho: : :pgals-hsp82-his3) using the lithium acetate method [25]. Following transformation the library was amplified for 12 hours at 30deg C under nonselective conditions using galactose (Gal) medium with 100 μg/mL ampicillin (1.7 g yeast nitrogen base without amino acids, 5 g ammonium sulfate, 0.1 g aspartic acid, 0.02 g arginine, 0.03 g valine, 0.1 g glutamic acid, 0.4 g serine, 0.2 g threonine, 0.03 g isoleucine, 0.05 g phenylalanine, 0.03 g tyrosine, 0.04 g adenine hemisulfate, 0.02 g methionine, 0.1 g leucine, 0.03 g lysine, 0.01 g uracil per liter with 1% raffinose and 1% galactose). After amplification the library culture was transferred to selective medium similar to Gal medium but raffinose and galactose are replaced with 2% dextrose. The culture was grown for 8 hours at 30deg C to allow shutoff of the wild-type copy of Hsp90 and then shifted to selective medium containing 0.5M NaCl for 12 generations. Samples were taken at specific time points and stored at -80degC.

Yeast lysis, DNA isolation and preparation for Illumina sequencing were performed as previously described [26]. Sequencing was performed by Elim Biopharmaceuticals, Inc and produced ≈30 million reads of 99% confidence at each read position based on PHRED scoring [27, 28]. Analysis of sequencing data was performed as previously described [29].

Estimation of growth rates

Individual growth rates were estimated according to the approach described by [20] using a Bayesian Monte Carlo Markov Chain (MCMC) approach proposed in [30]. Nucleotide sequences coding for the same amino acid sequence were interpreted as replicates with equal growth rates. The resulting MCMC output consisted of 10,000 posterior estimates for each amino acid mutation corresponding to an average effective samples size of 7,419 (minimum 725). Convergence was assessed using the Hellinger distance approach [31] combined with visual inspection of the resulting trace files.

Adaptive walks

In the strong selection weak mutation (SSWM) limit [32], adaptation can be modeled as a Markov process only consisting of subsequent fitness-increasing one-step substitutions that continue until an optimum is reached (so-called adaptive walks). This process is characterized

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by an absorbing Markov chain with a total of n different states (i.e., mutants), consisting of k absorbing (i.e., optima) and $n-k$ transient states (i.e., non-optima). Defining $w(g)$ as the fitness of genotype g, and $g_{[i]}$ as the genotype g carrying a mutant allele at locus i, the selection coefficient is denoted by $s_j(g) = w(g_{[j]}) - w(g)$, such that the transition probabilities $p_{g,g_{[i]}}$ for going from any mutant g to any mutant $g_{[i]}$ are given by the selection coefficient normalized by the sum over all adaptive, one-mutant neighbours of the current genotype g. If g is a (local) optimum, $p_{g,g} = 1$. Putting the transition matrix $\mathbf{P}$ in its canonical form and computing the fundamental matrix, then allows to determine the expectation and the variance in the number of steps before reaching *any* optimum, and to calculate the probability to reach optimum g when starting from genotype g' [33]. Robustness of the results and the influence of specific mutations were assessed by deleting the corresponding columns and rows in $\mathbf{P}$ (i.e., by essentially treating the corresponding mutation as unobserved), and re-calculating and comparing all statistics to those obtained from the full data set.

Correlation of fitness effects of mutations

Strength and type of epistasis was assessed by calculating the correlation of fitness effects of mutations γ [14], which quantifies how the selective effect of a focal mutation is altered when put onto a different genetic background, averaged over all genotypes of the fitness landscape. Extending recent theory [14], we calculated the matrix of epistatic effects between different pairs of alleles (A_i,B_i) and (A_j,B_j) termed $\gamma_{(A_i,B_i)\rightarrow(A_j,B_j)}$ (eq. S1_8), the vector of epistatic effects between a specific pair of alleles (A_i,B_i) on all other pairs of alleles $\gamma_{(A_i,B_i)\rightarrow}$ (eq. S1_9), the vector of epistatic effects between all pairs of alleles on a specific allele pair (A_j,B_j) termed $\gamma_{\rightarrow(A_j,B_j)}$ (eq. S1_12), and the decay of correlation of fitness effects γ_d (eq. S1_15) with Hamming distance d averaged over all genotypes g of the fitness landscape.

Fraction of epistasis

Following [34] and [35], we quantified whether specific pairs of alleles between two loci interact epistatically, and if so whether these display magnitude epistasis (i.e., fitness effects are non-additive, but fitness increases with the number of mutations), sign epistasis (i.e., one of the two mutations considered has an opposite effect in both backgrounds) or reciprocal sign epistasis (i.e., if both mutations show sign epistasis). In particular, we calculated the type of epistatic interaction between mutations $g_{[i]}$ and $g_{[j]}$ (with $i \neq j$) with respect to a given reference genotype g over the entire fitness landscape. There was no epistatic interaction if $|s_i(g_{[j]}) - s_i(g)| < \varepsilon = 10^{-6}$, magnitude epistasis if $s_j(g)s_j(g_{[i]}) \geq 0$ and $s_i(g)s_i(g_{[j]}) \geq 0$, reciprocal sign epistasis if $s_j(g)s_j(g_{[i]}) < 0$ and $s_i(g)s_i(g_{[j]}) < 0$, and sign epistasis in all other cases [36].

Roughness-to-slope ratio

Following [11], we calculated the roughness-to-slope ratio ρ by fitting the fitness landscape to a multidimensional linear model using the least-squares method. The slope of the linear model corresponds to the average additive fitness effect [10, 23], whereas the roughness is

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

given by the variance of the residuals. Generally, the better the linear model fit, the smaller the variance in residuals such that the roughness-to-slope ratio approaches 0 in a perfectly additive model. Conversely, a very rugged fitness landscape would have a large residual variance and, thus, a very large roughness-to-slope ratio (as in the House-of-Cards model). In addition, we calculated a test statistic Z_ρ by randomly shuffling fitness values in the sequence space to evaluate the statistical significance of the obtained roughness-to-slope ratio from the data set ρ_{data} given by $Z_\rho = \frac{\rho_{\mathrm{data}} - \mathrm{E}[\rho_{\mathrm{RSL}}]}{\sqrt{\mathrm{Var}[\rho_{\mathrm{RSL}}]}}$.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results and Discussion

We used the EMPIRIC approach [24, 26] to assess the growth rate of 640 mutants in yeast Hsp90 (see Materials and Methods). Based on previous screenings of fitness effects in different environments [29] and on different genetic backgrounds [20], and on expectations of their biophysical role, 13 amino-acid changing point mutations at 6 sites were chosen for the fitness landscape presented here (Fig. 1). The fitness landscape was created by assessing the growth rate associated with each individual mutation on the parental background, and all possible mutational combinations. A previously described MCMC approach was used to assess fitness and credibility intervals ([30]; see Materials and Methods).

The fitness landscape and its global peak

Figure 2 presents the resulting fitness landscape, with each mutant represented based on its Hamming distance from the parental genotype and its median estimated growth rate. Lines connect single-step substitutions, with vertical lines occurring when there are multiple mutations at the same position (Fig. 1). With increasing Hamming distance from the parental type, many mutational combinations become strongly deleterious. This indicates strong negative epistasis between the substitutions that, as single steps on the background of the parental type, have small effects. This pattern is consistent with Fisher's Geometric Model [37] when combinations of individually beneficial or small-effect mutations overshoot the optimum. It is also intuitively comprehensible on the protein level, where the accumulation of too many mutations is likely to destabilize the protein and render it dysfunctional [38].

Curiously, the global peak of the fitness landscape is located 4 mutational steps away from the parental type (Fig. 2B), with 98% of posterior samples identifying the peak. The fitness advantage of the global peak reaches nearly 10% over the parental type, and is consistent between replicates (see Materials and Methods; Supplementary Fig. 1). Though perhaps surprising given the degree of conservation of the studied genomic region ([24], Fig. S5), it is important to note that these fitness effects are measured under highly artificial experimental conditions including high salinity, which are unlikely to represent a natural environment of yeast. The effects of the individual mutations comprising the peak in a previous experiment without added NaCl were -0.04135, -0.01876, -0.03816, -0.02115 for mutations W585L, A587P, N588L and M589A, respectively, emphasizing the potential cost of adaptation associated with the increased salinity environment (data from [20]; see also [29]).

The global peak is not reached by combining the most beneficial single-step mutations, but via a highly synergistic combination of one beneficial and three 'neutral' mutations (i.e., mutations that are individually indistinguishable from the parental type in terms of growth rate). Figure 2C demonstrates that a multiplicative combination of the four mutations involved in the peak (termed "opt") predicts only a 4% fitness advantage. Furthermore, even a combination of the four individually most beneficial single-step mutations in the data set ("best"; considering at most one mutation per position) only predicts a benefit of 6%. Notably, the actual combination of these four mutations is highly deleterious and thus exhibits strong negative epistasis. Although negative epistasis between beneficial mutations during

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

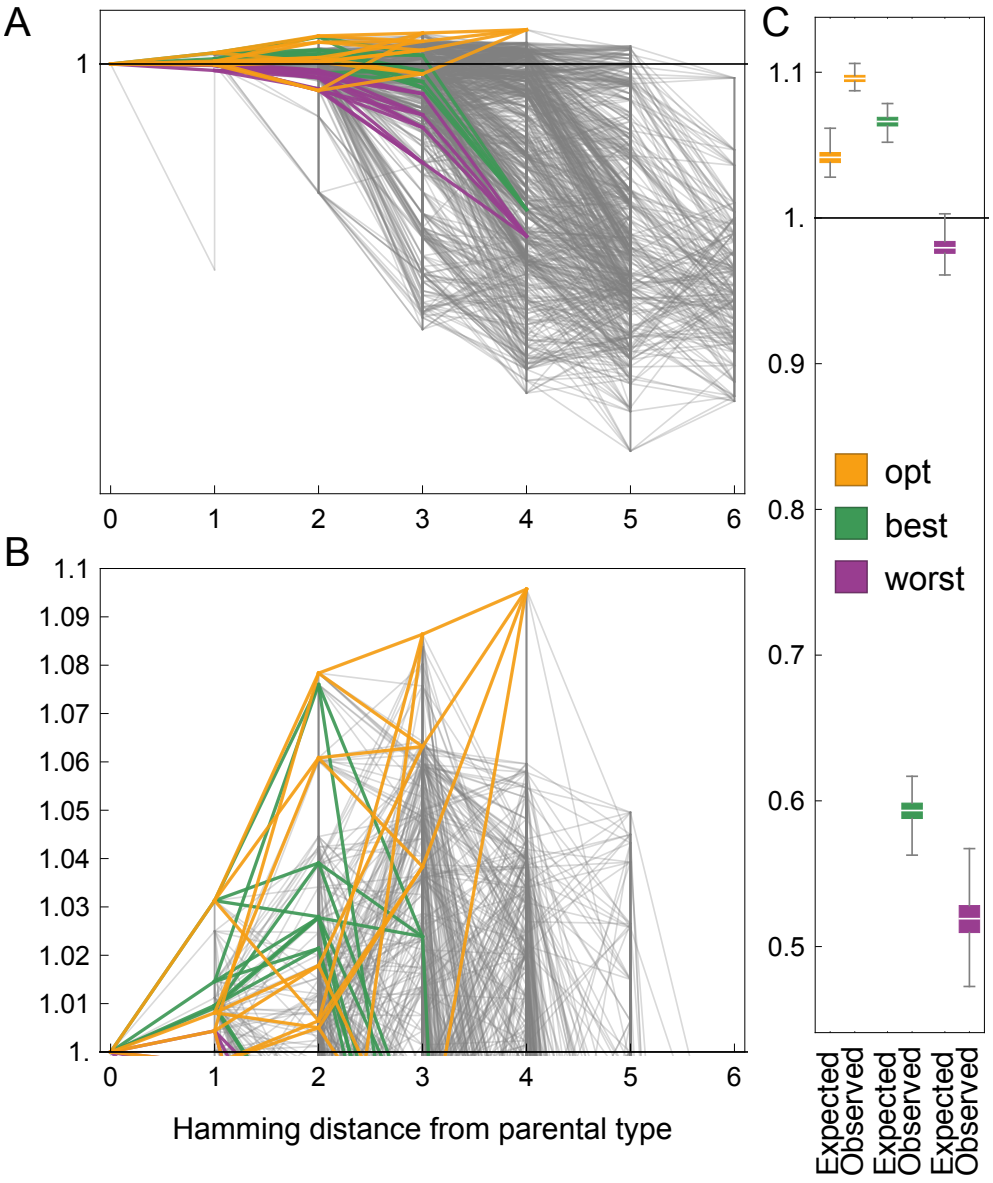


Figure 2: (A) Empirical fitness landscape by mutational distance to parental type. Each line represents a single substitution. Vertical lines appear when multiple alleles have been screened at the same position. There is a global pattern of negative epistasis. The three focal landscapes are highlighted. (B) Close-up on the beneficial portion of the landscape. (C) Expected (multiplicative) fitness versus observed fitness for the focal landscapes, obtained from 1000 posterior samples. We observe strong positive epistasis in the landscape that contains the global optimum, whereas the other two are dominated by negative epistasis. In all panels, the y axis depicts growth rate as a proxy for fitness.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

adaptation has been reported more frequently, positive epistasis has also been observed occasionally [18, 39], particularly in the context of compensatory evolution. In fact, negative epistasis between beneficial mutations and positive epistasis between neutral mutations has been predicted by de Visser *et al.* [40]. Furthermore, our results support the pattern recently found in the gene underlying the antibiotic resistance enzyme TEM-1 b-lactamase in *E. coli*, showing that large-effect mutations interact more strongly than small-effect mutations such that the fitness landscape of large-effect mutations tends to be more rugged than the landscape of small-effect mutations [13]. However, this conclusion is highly dependent on the measures of epistasis used and the selection of mutants for the landscape [10].

Adaptive walks on the fitness landscape

Next, we studied the empirical fitness landscape within a framework recently proposed by Draghi and Plotkin [41]. Given the empirical landscape, we simulated adaptive walks and studied the accessibility of the six observed local optima. In addition, we evaluated the length of adaptive walks starting from any mutant in the landscape, until an optimum is reached. In the strong selection weak mutation limit [42], we can express the resulting dynamics as an absorbing Markov chain, where local optima correspond to the absorbing states, and in which the transition probabilities correspond to the relative fitness increases attainable by the neighboring mutations (see Materials and Methods). This allowed us to derive analytical solutions for the mean and variance of the number of steps to reach a fitness optimum (see extended Materials and Methods), and the probability to reach a particular optimum starting from any given mutant in the landscape (Figs. 3, S2_2, S2_3).

Using this framework, we find that the global optimum can be reached with non-zero probability from almost 95% of starting points in the landscape, and is reached with high probability from a majority of starting points - indicating high accessibility of the global optimum (Fig. 3). The picture changes when restricting the analysis to adaptive walks initiating from the parental type (Figs. 3, S2_2, S2_3). Here, although 73% of all edges and 78% of all vertices are included in an adaptive walk to the global optimum, it is reached with only 26% probability. A local optimum two substitutions away from the parental type (Fig. 3C) is reached with a much higher probability of 47%. Hence, adaptation on the studied landscape is likely to stall at a sub-optimal fitness peak. This indicates that the local and global landscape pattern may be quite different, an observation that is confirmed and discussed in more detail below. In line with the existence of multiple local fitness peaks, we find that pairs of alleles at different loci show pervasive sign (30%) and reciprocal sign (8%) epistasis [34], whereas the remaining 62% are attributed to magnitude epistasis (i.e., there is no purely additive interaction between alleles; for a discussion of the contribution of experimental error see Fig. S2_4).

Epistasis measures and the topology of the fitness landscape

Next, we considered the global topology of the fitness landscape. Various measures of epistasis and ruggedness have been proposed, most of them correlated and hence capturing similar features of the landscape [10]. However, drawing conclusions has proven difficult because

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

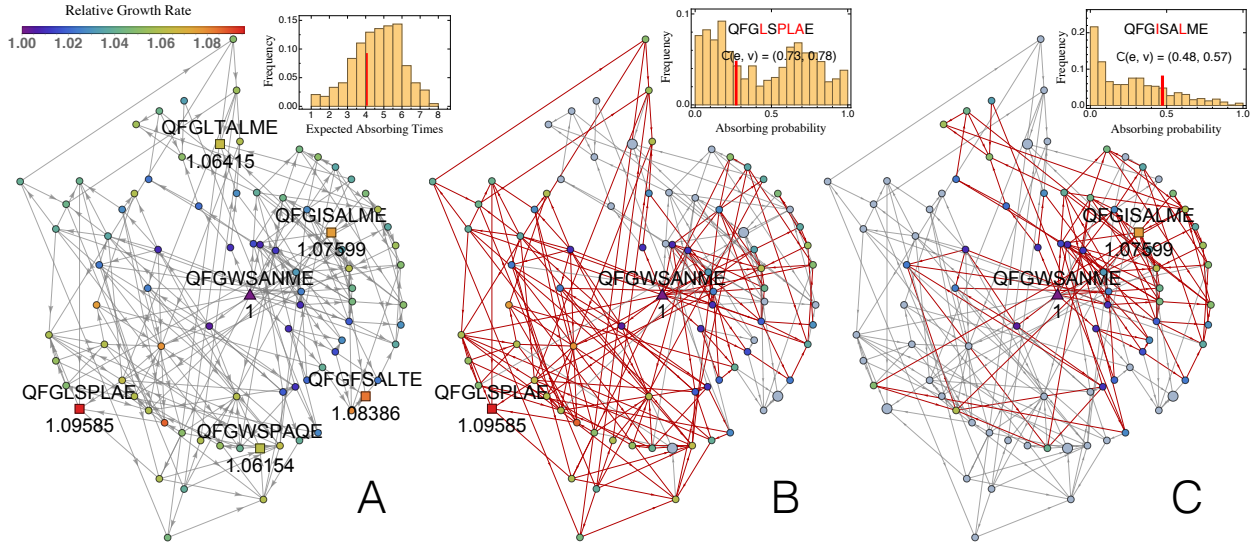


Figure 3: Graphical illustration of the sub-landscape of all beneficial mutants with respect to the parental type. (A) Each vertex (circle) corresponds to a mutant that is beneficial with respect to the parental type (triangle), colored according to their fitness. Arrows connect mutants that differ by a single substitution, with their direction indicating an increase in fitness. The five local optima in this sub-landscape are indicated as squares. Inset: distribution of average lengths of adaptive walks starting from any type in the full landscape (i.e., absorbing times of the Markov chain). The red line indicates mean path lengths for adaptive walks from the parental genotype (B) Paths leading towards the global optimum when beginning an adaptive walk at the parental genotype. Note that while 73% of all edges and 78% of all vertices are included in such a walk, it is reached in only 26% of all possible walks. Conversely, 47% of walks will reach the local optimum in panel C, which is closer to the parental type in terms of its Hamming distance. This demonstrates that evolution on this landscape is not well predicted by fitness only, and that a population may become stuck in a sub-optimal fitness peak. Inset B & C: Distribution of absorbing probabilities, that is, the probability to reach a specific optimum starting from a given mutant computed for all mutants in the data set. The red line corresponds to the respective probability when starting from the parental sequence. The global optimum is in general reached with a very high probability, but there are starting points from which it is poorly accessible.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the studied landscapes were created according to different criteria. Furthermore, published complete landscapes are too small to be divided into subsets, preventing tests for the consistency and hence the predictive potential of landscape statistics. The landscape studied here provides us with this opportunity. Moreover, because multiple alleles at the same site are contained within the landscape, we may study whether changes in the shape of the landscape are site- or amino-acid specific.

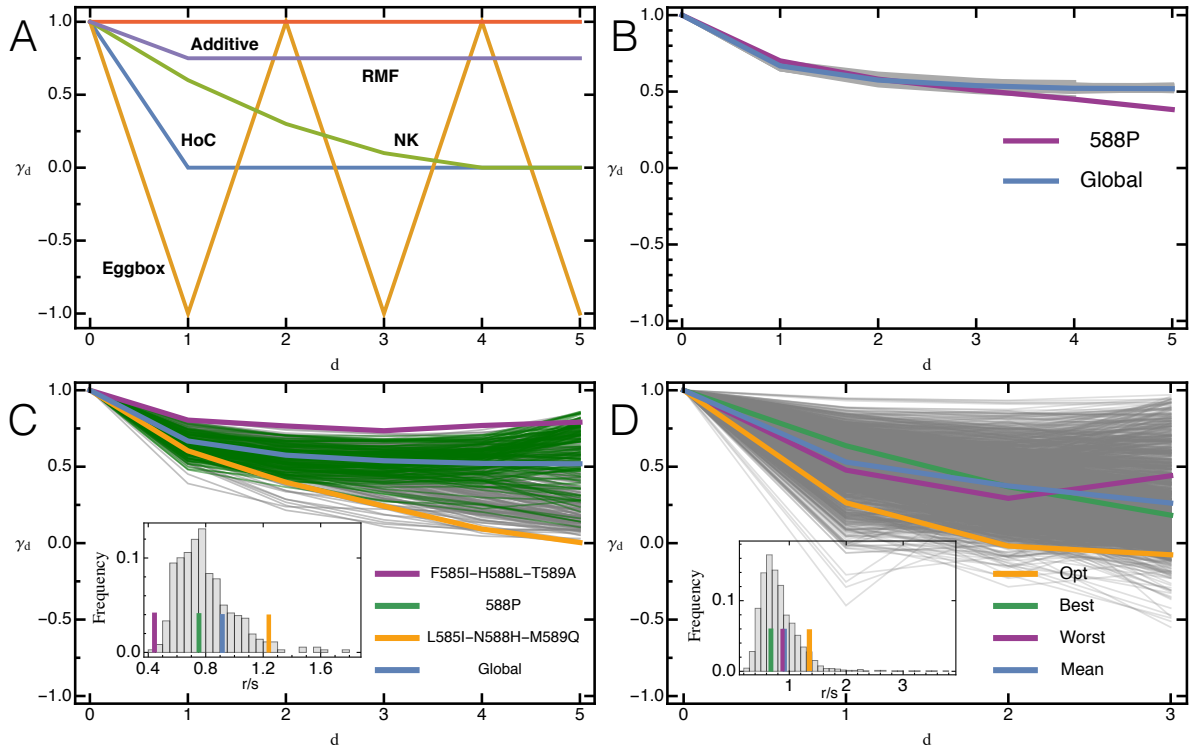


Figure 4: (A) Expected pattern of landscape-wide epistasis measure γ_d (eq. S1_15) with mutational distance for theoretical fitness landscapes with 6 loci from [14]. (B) Observed decay of γ_d with mutational distance under the drop-one approach is quite homogenous, except when hotspot mutation 588P is removed. (C) Observed decay of γ_d for all di-allelic 6-locus sub-landscapes. Depending on the underlying mutations, γ_d is vastly different suggesting qualitative differences in the topology of the underlying fitness landscape and in the extent of additivity in the landscape. Three focal landscapes representative of different types, and γ_d for the full landscape have been highlighted. (D) Observed decay of γ_d for all di-allelic 4-locus sub-landscapes containing the parental type, indicative of locally different landscape topologies. Highlighted are the three focal landscapes. Insets: Histograms of the roughness-to-slope ratio r/s for the respective subset, with horizontal lines indicating values for highlighted landscapes (r/s for global landscape (blue) is significantly different from HoC expectation, $p = 0.01$). Similar to γ_d, there is a huge variation in r/s for subparts of the fitness landscape, especially when considering the 4-locus subsets.

We computed various landscape statistics (roughness-to-slope ratio, fraction of epistasis, and the recently proposed gamma statistics; see Supplementary Material) [11, 10, 14], and

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

compared them to expectations from theoretical landscape models (NK, Rough-Mount-Fuji (RMF), House-of-Cards (HoC), Egg-Box landscapes). Whenever necessary, we provide an analytical extension of the used statistic to the case of multi-allelic landscapes (see Materials and Methods). To assess consistency and predictive potential, we computed the whole set of statistics for (1) all landscapes in which one amino acid was completely removed from the landscape (a cross-validation approach [43], subsequently referred to as the 'drop-one' approach), (2) all possible 360 di-allelic sub-landscapes, and (3) for all 1,570 di-allelic 4-step landscapes containing the parental genotype, highlighting as special examples the three focal landscapes discussed.

We find that the general topology of the fitness landscape resembles that of a RMF landscape with intermediate ruggedness, which is characterized by a mixture of a random HoC component and an additive component (Fig. 4A,B). Whereas the whole set of landscape statistics supports this topology and our conclusions, the gamma statistics measuring landscape-averaged correlations in fitness effects, recently proposed by [14], proved to be particularly illustrative. We will therefore focus on these in the main text; we refer to the Supplementary Material for additional results.

Predictive potential of landscape statistics

When computed based on the whole landscape and on a drop-one approach, the landscape appears quite homogeneous, and the gamma statistics show relatively little epistasis (Fig. 4B, 5B). On first sight, this contradicts our earlier statement of strong negative and positive epistasis but can be understood given the different definitions of the epistasis measures used: Above, we have measured epistasis based on the deviation from the multiplicative combination of the single-step fitness effects of mutations on the parental background. As these effects were small, epistasis was strong in comparison. Conversely, the gamma measure is independent of a reference genotype and captures the fitness decay with a growing number of substitutions as a dominant and quite additive component of the landscape.

Only mutation 588P has a pronounced effect on the global landscape statistics, and seems to act as an epistatic hotspot by making a majority of subsequent mutations (of indivually small effect) on its background strongly deleterious (clearly visible in Fig. 4B, 5C). This can be explained by looking at the biophysical properties of this mutation. In wild-type Hsp90, amino acid 588N is oriented away from solvent and forms hydrogen bond interactions with neighboring amino acids [24]. Proline lacks an amide proton, which inhibits hydrogen bond interactions. As a result, substituting 588N with a proline could disrupt hydrogen bond interactions with residues that may be involved in main chain hydrogen bonding and destabilize the protein. In addition, the pyrrolidine ring of proline is extremely rigid and can constrain the main chain, which may restrict the conformation of the residue preceding it in the protein sequence [44].

The variation between inferred landscape topologies increases dramatically for the 360 di-allelic 6-locus sub-landscapes (Fig.4C). Whereas they are still largely compatible with an RMF landscape, the decay of landscape-wide epistasis with mutational distance (as measured by γ_d) shows a large variance, suggesting large differences in the degree of additivity.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Interestingly, sub-sections of the landscape, typically carrying mutation *588P, show a relaxation of epistatic constraint with increasing mutational distance that cannot be captured by any of the proposed theoretical fitness landscape models, suggestive of non-random compensatory interactions. The variation in the shape of the fitness (sub-)landscapes is also reflected in the corresponding roughness-to-slope ratio (inset of Fig. 4C-D), further emphasizing inhomogeneity of the fitness landscape with local epistatic hotspots.

Finally, the 1,570 di-allelic 4-locus landscapes containing the parental genotype, though highly correlated genetically, reflect a variety of possible landscape topologies (Fig.4D), ranging from almost additive to egg-box shapes, accompanied by an extensive range of roughness-to-slope ratios. The three focal landscapes discussed above are not strongly different compared with the overall variation; yet show diverse patterns of epistasis between substitutions (Fig. 5).

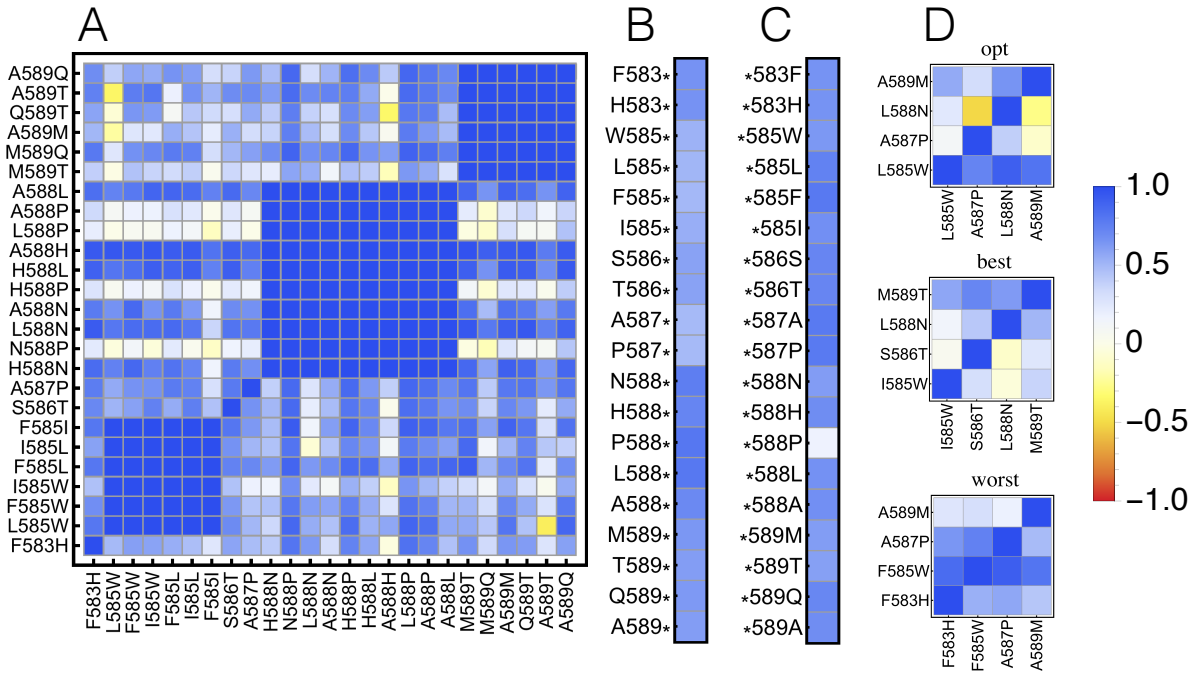


Figure 5: (A) Epistasis measure $\gamma_{(A_i,B_i)\to(A_j,B_j)}$ (eq. S1_8) between any two substitutions, averaged across the entire landscape. The majority of interactions are small to moderate (blue). Parts of the fitness landscape show highly localized and mutation-specific epistasis, ranging from strong magnitude epistasis (white) to sign and reciprocal sign epistasis (yellow). (B) The average epistatic effect $\gamma_{A_i\to}$ (eq. S1_9) of a mutation occurring on any background is always small. (C) The average epistatic effect $\gamma_{\to A_j}$ (eq. S1_13) of a background on any new mutation is usually small, except for background P588*, which shows a strong magnitude effect. (D) Locus-specific gamma for the 4 mutations leading to the global optimum (top), the 4 largest single effect mutations (middle), and the single effect mutations with the lowest fitness (bottom). The "opt" landscape exhibits strong sign epistasis between locus 587 and 588, and 588 and 589. Also the "best" landscape exhibits pervasive epistasis with sign epistasis between locus 588 and locus 585 and 586, respectively. We observe almost no epistasis in the "worst" landscape.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Thus, predicting fitness landscapes is difficult indeed. Extrapolation of the landscape, even across only a single mutation, may fail due to the existence of local epistatic hotspot mutations. While the integration of biophysical properties into landscape models is an important step forward [e.g. 45], we demonstrate that such models need to be mutation-specific. Considering a site-specific model (e.g., BLOSUM matrix; [46]) is not sufficient. Newer models such as DeepAlign may provide the opportunity to allow integration of mutation-specific effects via aligning two protein structures based on spatial proximity of equivalent residues, evolutionary relationship and hydrogen bonding similarity [47].

Conclusion

Originally introduced as a metaphor to describe adaptive evolution, fitness landscapes promise to become a powerful tool in biology to address complex questions regarding the predictability of evolution and the prevalence of epistasis within and between genomic regions. Due to their high-dimensional nature, however, the ability to extrapolate will be paramount to progress in this area, and the optimal quantitative and qualitative approaches to achieve this goal are yet to be determined.

Here, we have taken an important step towards addressing this question via the creation and analysis of a landscape comprising 640 engineered mutants of the Hsp90 protein in yeast. The unprecedented size of the fitness landscape along with the multi-allelic nature allows us to test whether global features could be extrapolated from subsets of the data. Although the global pattern indicates a rather homogeneous landscape, smaller sub-landscapes are a poor predictor of the overall global pattern because of 'epistatic hotspots'.

In combination, our results highlight the inherent difficulty imposed by the duality of epistasis for predicting evolution. In the absence of epistasis (i.e., in a purely additive landscape) evolution is globally highly predictable as the population will eventually reach the single fitness optimum, but the path taken is locally entirely unpredictable. Conversely, in the presence of (sign and reciprocal sign) epistasis evolution is globally unpredictable, as there are multiple optima and the probability to reach any one of them depends strongly on the starting genotype. At the same time, evolution may become locally predictable with the population following obligatory adaptive paths that are a direct result of the creation of fitness valleys owing to epistatic interactions.

The empirical fitness landscape studied here appears to be intermediate between these extremes. Although the global peak is within reach from almost any starting point, there is a local optimum that will be reached with appreciable probability, particular when starting from the parental genotype. From a practical standpoint, these results thus highlight the danger inherent to the common practice of constructing fitness landscapes from ascertained mutational combinations. However, this work also suggests that one promising way forward for increasing predictive power will be the utilization of multiple small landscapes used to gather information about the properties of individual mutations, combined with the integration of site-specific biophysical properties.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

This project was funded by grants from the Swiss National Science Foundation (FNS) and a European Research Council (ERC) Starting Grant to JDJ. Computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the Swiss Institute of Bioinformatics (SIB). We thank Inês Fragata and Ana-Hermina Ghenu for helpful comments and discussion.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

[1] Wright S (1932) The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proc. 6th Int. Congress Genet.* pp. 356–366.

[2] Coyne JA, Barton NH, Turelli M (1997) Perspective: A Critique of Sewall Wright's Shifting Balance Theory of Evolution. *Evolution* 51(3):643–671.

[3] Gavrilets S (2004) *Fitness landscapes and the origin of species.* (Princeton University Press, Princeton, NJ).

[4] de Visser JAGM, Krug J (2014) Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet* 15(7):480–490.

[5] Costanzo M et al. (2010) The Genetic Landscape of a Cell. *Science* 327(5964):425–431.

[6] Jiménez JI, Xulvi-Brunet R, Campbell GW, Turk-Macleod R, Chen IA (2013) Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc Natl Acad Sci USA* 110(37):14984–14989.

[7] Huang S (2013) Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer Metastasis Rev.* 32(3-4):423–448.

[8] Mann JK et al. (2014) The fitness landscape of HIV-1 gag: advanced modeling approaches and validation of model predictions by in vitro testing. *PLoS Comput. Biol.* 10(8):e1003776.

[9] Liu G, Rancati G (2016) Adaptive Evolution: Don't Fix What's Broken. *Curr Biol* 26(4):R169–71.

[10] Szendro IG, Schenk MF, Franke J, Krug J (2013) Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech.*

[11] Aita T, Iwakura M, Husimi Y (2001) A cross-section of the fitness landscape of dihydrofolate reductase. *Protein Eng.* 14(9):633–638.

[12] Franke J, Klözer A, de Visser JAGM, Krug J (2011) Evolutionary accessibility of mutational pathways. *PLoS Comput. Biol.* 7(8):e1002134.

[13] Schenk MF, Szendro IG, Salverda MLM, Krug J, de Visser JAGM (2013) Patterns of Epistasis between Beneficial Mutations in an Antibiotic Resistance Gene. *Molecular Biology and Evolution* 30(8):1779–1787.

[14] Ferretti L et al. (2016) Measuring epistasis in fitness landscapes: the correlation of fitness effects of mutations. *Journal of Theoretical Biology.*

[15] Kryazhimskiy S, Rice DP, Jerison ER, Desai MM (2014) Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* 344(6191):1519–1522.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[16] Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332(6034):1193–1196.

[17] Schoustra SE, Bataillon T, Gifford DR, Kassen R (2009) The properties of adaptive walks in evolving populations of fungus. *PLoS Biol* 7(11):e1000250.

[18] Weinreich D, Delaney N, DePristo M (2006) Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science* 312(5770):111–114.

[19] Dickinson WJ (2008) Synergistic fitness interactions and a high frequency of beneficial changes among mutations accumulated under relaxed selection in Saccharomyces cerevisiae. *Genetics* 178(3):1571–1578.

[20] Bank C, Hietpas RT, Jensen JD, Bolon DNA (2015) A systematic survey of an intragenic epistatic landscape. *Molecular Biology and Evolution* 32(1):229–238.

[21] MacLean RC, Hall AR, Perron GG, Buckling A (2010) The population genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. *Nat Rev Genet* 11(6):405–414.

[22] Jasmin JN, Lenormand T (2016) Accelerating Mutational Load Is Not Due to Synergistic Epistasis or Mutator Alleles in Mutation Accumulation Lines of Yeast. *Genetics* 202(2):751–763.

[23] Blanquart F, Achaz G, Bataillon T, Tenaillon O (2014) Properties of selected mutations and genotypic landscapes under Fisher's geometric model. *Evolution* 68(12):3537–3554.

[24] Hietpas RT, Jensen JD, Bolon DNA (2011) Experimental illumination of a fitness landscape. *Proc Natl Acad Sci USA* 108(19):7896–7901.

[25] Gietz DR, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods in Enzymology* 350:87–96.

[26] Hietpas R, Roscoe B, Jiang L, Bolon DNA (2012) Fitness analyses of all possible point mutations for regions of genes in yeast. *Nat Protoc* 7(7):1382–1396.

[27] Ewing B, Green P (1998) Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Research* 8:186–194.

[28] Ewing B, Hillier L, Wendl MC, Green P (1998) Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Research* 8:175–185.

[29] Hietpas RT, Bank C, Jensen JD, Bolon DNA (2013) Shifting fitness landscapes in response to altered environments. *Evolution* 67(12):3512–3522.

[30] 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(3):841–852.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[31] Boone EL, Merrick JRW, Krachey MJ (2012) A Hellinger distance approach to MCMC diagnostics. *Journal of Statistical Computation and Simulation* 84(4):833–849.

[32] Gillespie JH (1983) A simple stochastic gene substitution model. *Theoretical Population Biology* 23(2):202–215.

[33] Kemeny JG, Snell JL (1960) Finite markov chains (van Nostrand Princeton, NJ).

[34] Weinreich DM, Watson RA, Chao L (2005) Perspective: Sign Epistasis and Genetic Constraint on Evolutionary Trajectories. *Evolution* 59(6):1165–1174.

[35] Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445(7126):383–386.

[36] Brouillet S, Annoni H, Ferretti L, Achaz G (2015) MAGELLAN: a tool to explore small fitness landscapes. *bioRxiv* p. 031583.

[37] Weinreich DM, Knies JL (2013) Fisher's geometric model of adaptation meets the functional synthesis: data on pairwise epistasis for fitness yields insights into the shape and size of phenotype space. *Evolution* 67(10):2957–2972.

[38] Soskine M, Tawfik DS (2010) Mutational effects and the evolution of new protein functions. *Nat Rev Genet* 11(8):572–582.

[39] Trindade S et al. (2009) Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genet.* 5(7):e1000578.

[40] de Visser JA, Hoekstra RF, van den Ende H (1997) An experimental test for synergistic epistasis and its application in Chlamydomonas. *Genetics* 145(3):815–819.

[41] Draghi JA, Plotkin JB (2013) Selection biases the prevalence and type of epistasis along adaptive trajectories. *Evolution* 67(11):3120–3131.

[42] Gillespie JH (1984) Molecular Evolution Over the Mutational Landscape. *Evolution* 38(5):1116–1129.

[43] Geisser S (1993) *Predictive Inference.* (CRC Press) Vol. 55.

[44] Bajaj K et al. (2007) Stereochemical criteria for prediction of the effects of proline mutations on protein stability. *PLoS Comput. Biol.* 3(12):e241.

[45] Wylie CS, Shakhnovich EI (2011) A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc Natl Acad Sci USA* 108(24):9916–9921.

[46] Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences* 89(22):10915–10919.

[47] Wang S, Ma J, Peng J, Xu J (2013) Protein structure alignment beyond spatial proximity. *Sci Rep* 3:1448.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[48] Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Proceedings of the Royal Society of Edinborough* 52:399–433.

[49] Kingman JFC (1978) A simple model for the balance between selection and mutation. *Journal of Applied Probability* pp. 1–12.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information 1: Extended Materials and Methods

Adaptive walks

Under the strong selection weak mutation (SSWM) limit [42], adaptation follows an absorbing Markovian process characterized by a series of fitness-increasing substitutions along one-mutant neighbours until reaching a fitness optimum (forming a so-called adaptive walk), with a total of n different states (i.e., mutants), consisting of k absorbing (i.e., optima) and $n-k$ transient states (i.e., non-optima). Defining $w(g)$ as the fitness of genotype g, and $g_{[i]}$ as the genotype g carrying a mutant allele at locus i, the selection coefficient is denoted by

$$s_j(g) = w(g_{[j]}) - w(g), \tag{S1_1}$$

such that the transition probabilities for going from any mutant j to any mutant i are given by the transition matrix

$$\mathbf{P} = \begin{cases} \dfrac{s_i}{\sum\limits_{i \in M_j} s_i} & \text{if } M_j \neq \emptyset \\ 1 & \text{else,} \end{cases} \tag{S1_2}$$

where $M(g) := \{j : s_j(g) > 0, \mathcal{D}_{\mathrm{HD}}(g, g_{[j]}) = 1\}$ denotes the set of all adaptive, one-mutant neighbours of the current genotype g.

The canonical form of $\mathbf{P}$ can then be obtained by permutation, such that

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix}, \tag{S1_3}$$

where $\mathbf{Q}$ is a $(n-k) \times (n-k)$ matrix which contains the transition probabilities between transient states; $\mathbf{R}$ is a $(n-k) \times k$ matrix which gives the transition probabilities from any transient to any absorbing state; $\mathbf{0}$ is the $k \times (n-k)$ zero matrix; and $\mathbf{I}_k$ is the $k \times k$ identity matrix [33].

Using the above representation, all basic properties of the absorbing Markov chain can be calculated from the fundamental matrix

$$\mathbf{N} = \sum_{i=0}^{\infty} = \mathbf{Q}^i = \left(\mathbf{I}_{(n-k)} - \mathbf{Q}\right)^{-1}. \tag{S1_4}$$

In particular, the expected number of steps before absorption (i.e., the expected number of steps on the fitness landscape before reaching any optimum) is given by

$$\mathrm{E}[\mathbf{t}] = \mathbf{N}\mathbf{1}, \tag{S1_5}$$

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where $\mathbf{1}$ is a column vector of length $(n-k)$ with all entries being 1, and the ith entry of $\mathrm{E}[\mathbf{t}]$ gives the expected number of steps when starting from state (mutant) i.

Similarly, the variance in the number of steps before being absorbed can be computed as

$$\mathrm{Var}[\mathbf{t}] = (2\mathbf{N} - \mathbf{I}_t)\,\mathrm{E}[\mathbf{t}] - \mathrm{E}[\mathbf{t}] \odot \mathrm{E}[\mathbf{t}], \tag{S1_6}$$

where $\odot$ denotes the Hadamard product.

Finally, the probability of being absorbed in state j when starting from transient state i (i.e., reaching optimum j when the initial genotype is i), is given by the $(n-k) \times k$-matrix

$$\mathbf{B} = \mathbf{NR}. \tag{S1_7}$$

Thus, these methods give an easy and computationally fast way of quantifying and predicting adaptive walks on fitness landscapes. Furthermore, the robustness of these results and the influence of particular mutants can be assessed by deleting the ith column and row of $\mathbf{P}$ – i.e., by essentially treating mutant i as unobserved –, recalculating the above statistics and comparing these to the statistics obtained from the full data set. Similarly, entire mutations (i.e., amino acids) can be left out to assess their relative effect on the fitness landscape and the generality of our analysis (and the statistics used).

Measuring epistasis

We applied different metrics for quantifying epistasis over the entire fitness landscape as well as for particular mutations and assessed their consistency in capturing the strength of gene×gene interactions. In particular, we follow the definition of epistasis by [48] (originally termed epistacy), as the deviation from additivity when combining two genetic effects which is measured by the difference in log-fitness between the effects of the double mutant and the single mutant relative to the wild-type fitness.

Correlation of fitness effects of mutations: γ

The first measure has recently been introduced by [14] and is defined as the single-step correlation of fitness effects for mutations between neighbouring genotypes. It quantifies how the selective effect of a focal mutation is altered when it occurs in a different genetic background averaged over all genotypes of the fitness landscape. Geometrically, γ measures the correlation between slopes (with respect to genotype-fitness hypercube) of the same mutation put into different genetic backgrounds. Thus, if the fitness effect of a mutation is independent of its genetic background (i.e., if there is no epistasis), the correlation in slopes will be perfect ($\gamma = 1$), whereas it will be zero if the fitness slopes of each genotype are independent of the fitnesses of other genotypes (as in the House-of-Cards model; 49). Depending on the scale γ can either be used to quantify the strength of gene×gene interactions between specific mutations or as an overall measure for the entire fitness landscape. However, in its original form γ was defined for bi-allelic data only and thus needs to be extended by considering pairs of specific alleles at different loci [14].

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Let $\mathcal{A}_i = \{1_i, 2_i, \ldots, m_i\}$ denote the set of different alleles present at locus i for all polymorphic loci $i \in \{1, 2, \ldots, n\}$ such that $|\mathcal{A}_i| \geq 2$ for all loci i. Further, let $\mathcal{G}$ be the set of all phenotypes that can be formed by combining all alleles such that the total number of genotype is $|\mathcal{G}| = \prod_{i=1}^{n} m_i$.

Then the matrix of epistatic effects between loci i and j carrying alleles $A_i, B_i \in \mathcal{A}_i$ and $A_j, B_j \in \mathcal{A}_j$ is given by

$$\gamma_{(A_i,B_i)\to(A_j,B_j)} = \mathrm{Cor}[s_{(A_j,B_j)}(g), s_{(A_j,B_j)}(g_{[(A_i,B_i)]})] = \frac{\sum_g s_j(g) s_j(g_{[i]})}{\sum_g (s_j(g))^2}. \tag{S1_8}$$

where $g := \{x \in \mathcal{G} | x_i = A_i \text{ or } x_i = B_i \text{ and } x_j = A_j \text{ or } x_j = B_j\} \subseteq \mathcal{G}$ such that the sum is only calculated over the subset of genotypes carrying one of the two focal alleles at each focal locus. Thus, $\gamma_{(A_i,B_i)\to(A_j,B_j)}$ is a quadratic matrix of dimension $\left(\sum_{i=1}^{n} \frac{|\mathcal{A}_i|(|\mathcal{A}_i|-1)}{2}\right)$. Note that in the case where $|\mathcal{A}_i| = 2$ for all loci i, we obtain equation 9 in [14].

Likewise, the epistatic effect of a mutation in locus i with alleles (A_i, B_i) on other loci (and pairs of alleles) can be calculated as

$$\gamma_{(A_i,B_i)\to} = \mathrm{Cor}[s(g), s(g_{[(A_i,B_i)]})] = \frac{\sum_{j\neq i} \sum_{\mathfrak{a}_j} \sum_g s_j(g) s_j(g_{[i]})}{\sum_{j\neq i} \sum_{\mathfrak{a}_j} \sum_g (s_j(g))^2}, \tag{S1_9}$$

where the summation index $\mathfrak{a}_j = \{(A_j, B_j) \mid A_j, B_j \in \mathcal{A}_j \text{ and } A_j \neq B_j\}$ is over the set of subsets of size two that can be constructed from all alleles found at locus j. Note that the third summation index g changes depending on $\mathfrak{a}_j$.

An additional summation allows calculation of the epistatic effect of a mutation in locus i carrying allele (A_i) on other loci (and pairs of alleles) can be calculated as

$$\gamma_{A_i\to} = \mathrm{Cor}[s(g), s(g_{[A_i]})] = \frac{\sum_{j\neq i} \sum_{\mathfrak{f}_i} \sum_{\mathfrak{a}_j} \sum_g s_j(g) s_j(g_{[i]})}{\sum_{j\neq i} \sum_{\mathfrak{f}_i} \sum_{\mathfrak{a}_j} \sum_g (s_j(g))^2}, \tag{S1_10}$$

where $\mathfrak{f}_i = \{(A_i, B_i) \mid B_i \in \mathcal{A}_i \text{ and } A_i \neq B_i\}$ such that the sum is only calculated over the elements of the set of subsets of size two that can be constructed from all alleles found at locus i that contain allele A_i.

Then, summing over $\mathfrak{l}_i = \{(A_i, B_i) \mid A_i, B_i \in \mathcal{A}_i \text{ and } A_i \neq B_i\}$, i.e., the elements of the set of subsets of size two that can be constructed from all alleles found at locus i, gives the epistatic effect of a mutation in locus i

$$\gamma_{i\to} = \mathrm{Cor}[s(g), s(g_{[i]})] = \frac{\sum_{j\neq i} \sum_{\mathfrak{l}_i} \sum_{\mathfrak{a}_j} \sum_g s_j(g) s_j(g_{[i]})}{\sum_{j\neq i} \sum_{\mathfrak{l}_i} \sum_{\mathfrak{a}_j} \sum_g (s_j(g))^2}. \tag{S1_11}$$

Similarly, the epistatic effect of other mutations (again considering pairs of alleles first) on locus j with alleles (A_i, B_i) can be calculated as

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\gamma_{\rightarrow(A_j,B_j)} = \mathrm{Cor}[s_{(A_j,B_j)}(g), s_{(A_j,B_j)}(g_1)] = \frac{\sum_{i\neq j}\sum_{\mathfrak{a}_i}\sum_g s_j(g)s_j(g_{[i]})}{\sum_{i\neq j}\sum_{\mathfrak{a}_i}\sum_g (s_j(g))^2}, \tag{S1_12}$$

the epistatic effect of other mutations on locus j carrying allele (A_j) is given by

$$\gamma_{\rightarrow(A_j)} = \mathrm{Cor}[s_{(A_j)}(g), s_{(A_j)}(g_1)] = \frac{\sum_{i\neq j}\sum_{\mathfrak{f}_j}\sum_{\mathfrak{a}_i}\sum_g s_j(g)s_j(g_{[i]})}{\sum_{i\neq j}\sum_{\mathfrak{f}_j}\sum_{\mathfrak{a}_i}\sum_g (s_j(g))^2}, \tag{S1_13}$$

and the epistatic effect of other mutations on locus j becomes

$$\gamma_{\rightarrow j} = \mathrm{Cor}[s_j(g), s_j(g_1)] = \frac{\sum_{i\neq j}\sum_{\mathfrak{l}_j}\sum_{\mathfrak{a}_i}\sum_g s_j(g)s_j(g_{[i]})}{\sum_{i\neq j}\sum_{\mathfrak{l}_j}\sum_{\mathfrak{a}_i}\sum_g (s_j(g))^2}. \tag{S1_14}$$

Finally, γ_d, that is the decay of correlation of fitness effects with Hamming distance d, (i.e., the cumulative epistatic effect of d mutations averaged over the entire fitness landscape) is calculated as

$$\gamma_d = \mathrm{Cor}[s(g), s_j(g_d)] = \frac{\sum_g\sum_{g_d}\sum_{j\neq i_1,i_2\ldots i_d}\sum_{\mathcal{A}_j\setminus\{A_j\}} s_j(g)s_j(g_{[i_1 i_2\ldots i_d]})}{\sum_g\sum_{\mathcal{A}_j\setminus\{A_j\}}(s_j(g))^2}, \tag{S1_15}$$

where the last summation is over all different alleles present at locus j except the one carried by genotype g at locus j. Note that there is unfortunately no multi-allelic analog to equation (14) of [14]. Furthermore, as desired when there are only two alleles at a given locus, equations (S1_9 – S1_11) and equations (S1_12 – S1_14) collapse and give identical values, and reduce to their bi-allelic counterparts (i.e., eq. 7-8 in 14).

Fraction of epistasis

The second statistic quantifies whether specific pairs of alleles between two loci interact epistatically, and if so whether they display magnitude epistasis (i.e., fitness effects are non-additive, but fitness increases with the number of mutations), sign epistasis (i.e., one of the two mutations considered has an opposite effect in both backgrounds) or reciprocal sign epistasis (i.e., if both mutations show sign epistasis; 34, 35; for an implementation see [36]).

Using equation (S1_1) the type of epistatic interaction between a mutations at locus i and j (with $i \neq j$) when introduced on some reference genotype g, $\mathfrak{e}(g,i,j)$, can formally be given as

$$\mathfrak{e}(g,i,j) = \begin{cases} \text{none} & \text{if } |s_i(g_{[j]}) - s_i(g)| = 0 \\ \text{magnitude} & \text{if } s_j(g)s_j(g_{[i]}) \geq 0 \text{ and } s_i(g)s_i(g_{[j]}) \geq 0 \\ \text{reciprocal sign} & \text{if } s_j(g)s_j(g_{[i]}) < 0 \text{ and } s_i(g)s_i(g_{[j]}) < 0 \\ \text{sign} & \text{else} \end{cases}. \tag{S1_16}$$

Note that for numerical purposes we allowed for a very small deviation $\varepsilon = 10^{-6}$ in the case of no epistatic interaction. For ease of notation, we treat the elements in $\mathcal{A}_i$ as fixed and

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ordered, and define $\mathrm{I}(g_{[i]})$ as the index of the allele present at locus i in genotype g in $\mathcal{A}_i$ such that the fraction of epistasis over all $\mathcal{G}$ can be calculated as

$$\mathfrak{E}(x) = c^{-1} \sum_{g} \sum_{k=\mathrm{I}(g_{[i]})+1}^{m_i} \sum_{l=k+1}^{m_j} \mathbb{1}_x \left(\mathfrak{e}(g,k,l) \right), \tag{S1_17}$$

where $x \in \{\text{none}, \text{magnitude}, \text{sign}, \text{reciprocal sign}\}$, $\mathbb{1}_x(y)$ is the indicator function that is 1 if $x = y$ and 0 otherwise, and $c = \left(\sum_{g} \sum_{k=\mathrm{I}(g_{[i]})+1}^{m_i} \sum_{l=k+1}^{m_j} 1 \right)$ is a normalization constant.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information 2: Supporting Figures

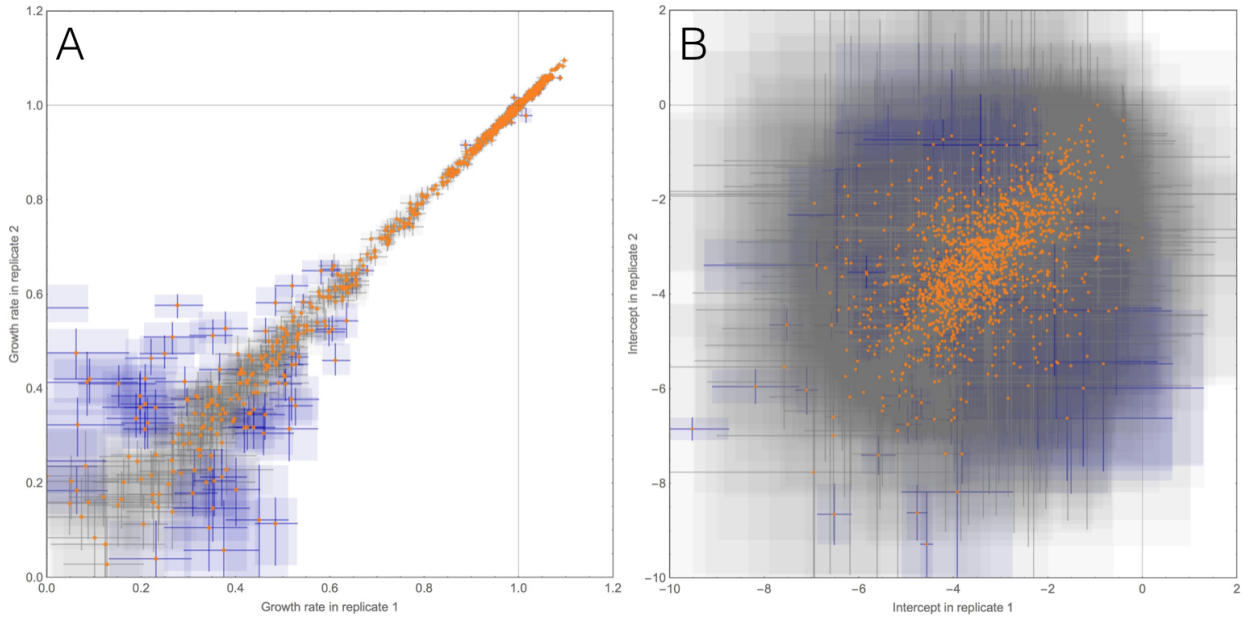


Figure S2_1: (A) Scatterplot demonstrating the correlation of estimated growth rates ($R^2 = 0.944$). Orange dots indicate median growth rates, rectangles indicate the limits of the 95% credibility intervals (in blue if they do not overlap between replicates). (B) In contrast, estimated initial population sizes are much more variable between replicates, demonstrating that initial representation in the mutant library had no effect on the estimated growth rate.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

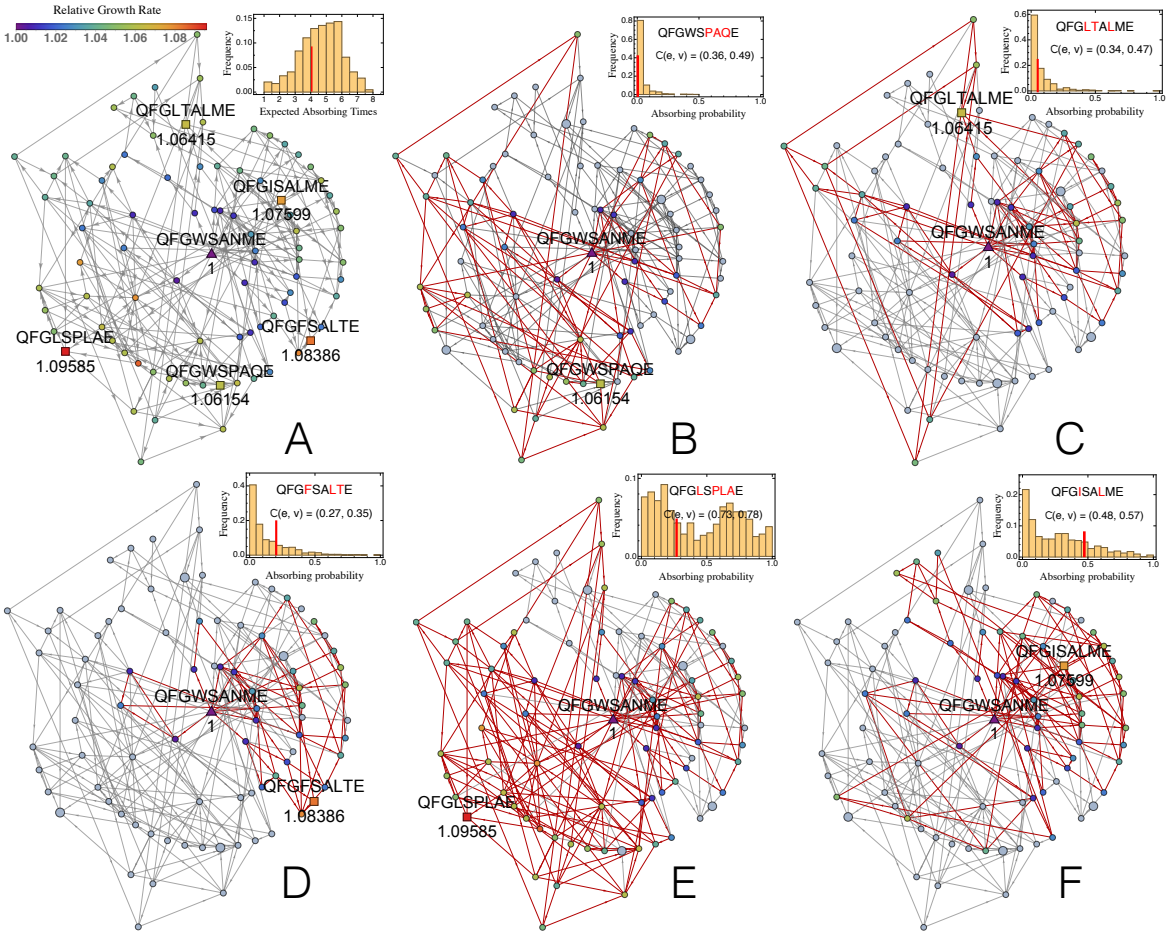


Figure S2_2: Graphical illustration of the sub-landscape of all beneficial mutants with respect to the parental type. (A) Each vertex (circle) corresponds to a mutant that is beneficial with respect to the parental type (triangle), colored according to their fitness and arranged circularly by their Hamming distance to the parental type. Arrows connect mutants that differ by a single substitution, with their direction indicating an increase in fitness. The five local optima in this sub-landscape are indicated as squares. Inset: distribution of average lengths of adaptive walks starting from any type in the full landscape (i.e., absorbing times of the Markov chain). The red line indicates mean path lengths for adaptive walks from the parental genotype. (B)-(F) Paths leading towards the different optima – highlighted by their sequence name – when beginning an adaptive walk at the parental genotype. Inset (B) - (F): Distribution of absorbing probabilities, that is, the probability to reach a specific optimum starting from a given mutant computed for all mutants in the data set. The red line corresponds to the respective probability when starting from the parental sequence. The global optimum is in general reached with a very high probability, but there are starting points from which it is poorly accessible.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

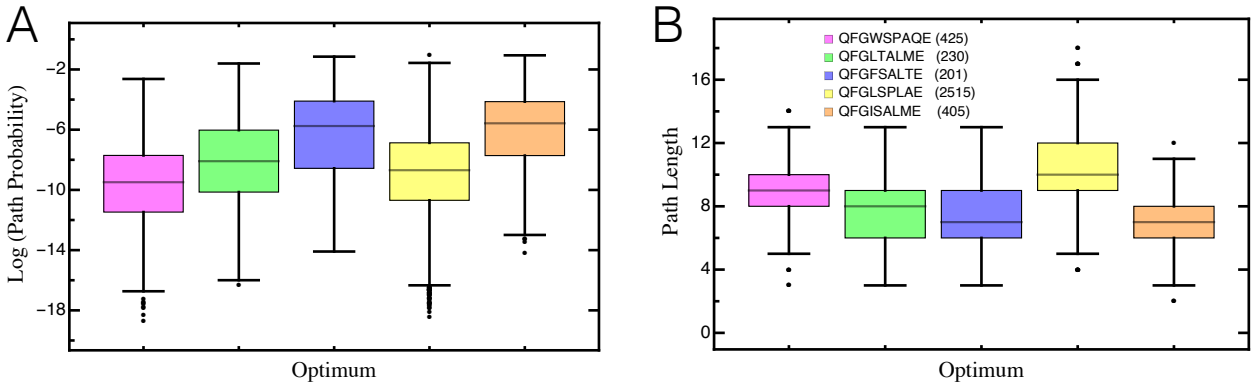
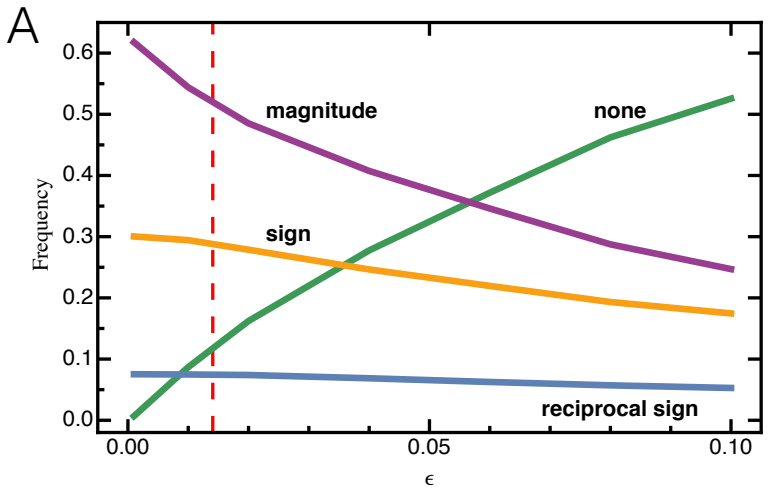


Figure S2_3: (A) Path probability of reaching a specific local optimum when beginning an adaptive walk at the parental genotype. Results are presented on log scale. Boxes represent the interquartile range (i.e., the 50% C.I.), whiskers extend to the highest/lowest data point within the box ± 1.5 times the interquartile range, and black circles represent close outliers. (B) Path length (i.e., the number of substitutions) to reach a specific local optimum. Inset: Color coding for the different optima with the total number of different paths leading to the respective optimum in brackets. There is a total of 3776 different paths that can be taken from the parental genotype, 66% of which lead to the global optimum (yellow). Most of these, however, require multiple substitutions, such that the probability to reach the global optimum is generally low. In contrast, the local optimum close to the parental genotype (orange) can be reached with only a few substitutions, and thus, with relatively high probability.

bioRxiv preprint doi: https://doi.org/10.1101/048769; this version posted April 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Type	Mean	95% IQR
none	8.24E-06	[0, 9.47E-05]
magnitude	0.62	[0.612, 0.628]
sign	0.298	[0.291, 0.305]
reciprocal sign	0.082	[0.078, 0.086]

Figure S2_4: (A) The fraction of epistasis (eq. S1_17) calculated for different values of ε, quantifying the error threshold beyond which interactions between different alleles are considered to be epistatic. Expectedly, as ε increases the number of non-epistatic interactions between alleles increases, which in the majority of cases were previously classified as magnitude epistatic interactions. Likewise, the fraction of sign and reciprocal sign epistasis decreases. The red dashed line shows the mean absolute difference between median growth rates obtained from the two independent replicates calculated over all mutants with median growth rate larger than the mean median growth rate estimated for the substitution F583$*$ (i.e., when substituting a stop codon) between the two replicates. Using this threshold for ε, we find that approximately 10% of all interactions are additive, approximately 50% show magnitude epistasis, whereas the sign and reciprocal sign epistasis remains roughly constant at 30% and 8%, respectively. (B) The mean and the 95% interquantile range (IQR) of the fraction of epistasis calculated over 10,000 posterior samples obtained from the MCMC simulations using the standard threshold $\varepsilon = 10^{-6}$. There is only little to no variation in the fraction of epistasis and the mean value is essentially identical to the proportions reported in the main text, indicating robustness of the inferred fraction of epistasis, and confirming that variation in posterior samples is typically small (i.e., growth rates are estimated with high precision).