Variations and predictability of epistasis on an intragenic fitness landscape.

Sarvesh Baheti[1], Namratha Raj[2], Supreet Saini[2,*]

[1] Department of Chemical Engineering, Indian Institute of Technology Jodhpur, India

[2] Department of Chemical Engineering, Indian Institute of Technology Bombay, India

* Corresponding author. Email: saini@che.iitb.ac.in, Phone: +91 22 2576 7216

Abstract

How epistasis hinders or facilitates movement on fitness landscapes has been a longstanding question of interest. Several high throughput experiments have demonstrated that despite its idiosyncrasy, epistatic effects exhibit global statistical patterns. Recently, Papkou et. al. constructed a fitness landscape for a 9-base region in the *folA* gene, which encodes for dihydrofolate reductase (DHFR), in *E. coli*, and demonstrated that despite being highly rugged, the landscape is highly navigable. In this work, using the *folA* landscape, we ask two questions: (1) How does the nature of epistatic interactions change as a function of the genomic background? (2) How predictable is epistasis within a gene? Our results show that epistasis is “fluid” - the nature of epistasis exhibited by a pair of mutations is strongly contingent on the genetic background. Mutations exhibit one of two binary “states”: a small fraction of mutations exhibit extremely strong patterns of global epistasis, while most do not. Despite these observations, we observe that the distribution of fitness effects (DFE) of a genotype is highly predictable based on its fitness. These results offer a new perspective on how epistasis operates within a gene, and how it can be predicted.

Keywords: *epistasis, fitness landscapes, folA, diminishing returns, increasing costs*

Significance Statement.

How a mutation changes organismal fitness is dependent on the genome in which it occurs. This phenomenon is known as epistasis and makes evolution unpredictable. Recent efforts to understand epistasis have led to the identification of statistical patterns in its manifestations. To study how epistasis operates in protein evolution, we analyze a recently reported landscape which quantifies fitness of ~260000 sequences of an *E. coli* gene. We show two previously unknown properties of epistasis: “fluid” (epistasis between mutations is controlled via other epistatic interactions) and “binary” (only a few mutations exhibit statistical patterns; most do not). This work sheds new light on how epistasis manifests in gene sequences. Our results have important consequences for protein & organismal evolution.

Introduction

Mutations decide the fitness of an organism in an environment-dependent fashion. But, the effect of mutations also depends on the genetic background they are occurring in[1]. This phenomenon is referred to as epistasis. Hence, epistasis influences adaptation[2,3]. However, it is largely unpredictable, although a few statistical patterns based on macroscopic traits have been reported in the last few years[4-10]. One way to study genotype-phenotype relationship and patterns of epistasis is by using fitness landscapes[11-13].

Fitness landscape is a multidimensional surface, in which one dimension represents a fitness-related phenotype, and the others genotype. Hence, it serves as a genotype-phenotype map, whose shape, in a given environment, is a consequence of genetic interactions or epistasis[14-17]. If the landscape has many peaks, its structure is called rugged[18-20]. Populations navigating on a rugged landscape are likely to be trapped in local peaks, whose precise identity is dictated by chance and population's starting point on the landscape[21-23]. Alternatively, on smooth landscapes, populations starting from different points on the sequence space all converge to the same global peak[24-27]. Thus, the structure of the landscape also dictates our ability to predict evolution, and this has wide-ranging implications.

Although fitness landscape was conceptualized to explain the relationship between a population's genotype and fitness, it has evolved to explain the relationship between fitness and functional protein-coding sequences[13,28]. This effort to characterize fitness landscapes started by considering a handful of "important" biallelic sites in a protein[27,29-37]. Although such landscapes could explain the pervasiveness of epistasis in protein sequences, they do not allow us to make wide-ranging statistical predictions of the characteristics of the fitness landscape[38]. However, with advancing high-throughput technologies, it has become possible to construct high-dimensional landscapes[39-45].

In this study, we use one such high-dimensional fitness landscape constructed by Papkou and coworkers to understand how epistasis operates in a 9-base pair region of the *folA* gene in *E. coli*[40]. *folA* encodes for dihydrofolate reductase (DHFR) and mutations in the gene are known to confer resistance to the antibiotic trimethoprim[46-50].

Via analysis of the *folA* landscape, we ask three specific questions (**Supplement Figure S1**): (1) How does the nature of epistasis between two given sites change as a function of the genetic background? (2) Are these changes dependent on the fitness or genotype? (3) Does a given mutation follow already known patterns of global epistasis? If yes, what does it depend on?

Our results show that epistasis is “fluid” – i.e., the nature of epistasis two mutations exhibit is a function of the genetic background. We also show that only a small fraction of mutations follow global epistasis. In fact, mutations can be classified into two groups consisting of ones that show global epistasis and others (comprising of a majority) that do not. We also propose a novel way to estimate and predict distribution of fitness effects (DFE) of a given genotype.

Results

folA fitness landscape in *E. coli*.

A fitness landscape of a 9-base pair region of *folA* was generated recently by Papkou and coworkers[40]. The landscape explored a 9-bp region of the gene that has been shown to be important for resistance to the antibiotic trimethoprim[47,51]. All 4^9 variants of *folA* were generated, and grown in media containing the antibiotic. Deep sequencing was used to quantify fitness, and relative fitness of ~99.7% of all variants was obtained. A striking feature of the landscape reported by Papkou and coworkers was that despite being highly rugged with 514 peaks, a majority of the landscape had adaptive access to high fitness peaks. This was because, compared to lower peaks, high fitness peaks had a large basin of attraction.

Our analysis using this dataset shows that with increasing size of the landscape, the number of peaks increase; however, the density of the peaks decreases (**Supplement Figure S2** and **S3**). With this increasing size of the landscape, however, the accessibility of the global peak decreases (**Supplement Figure S4**). Globally, only small regions of the landscape could be represented as a maximally rugged NK landscape (**Supplement Figure S5**).

By scanning all mutations on a 9-dimensional landscape, Papkou *et. al.* have created a dataset that allows us to ask specific questions about epistasis and its manifestations on a fitness landscape. The genotypes on the *folA* landscape have been divided into two categories by Papkou *et. al.* into functional (~7% of all points) and non-functional (~93% of all points). This distinction is based on a statistical segregation of the 4^9 points. Since this segregation does not have any functional basis, we call the two groups as “high fitness” and “low fitness”.

Nature of epistasis between two mutations is context dependent or “fluid” nature of epistasis.

Epistasis between two mutations, *A* and *B*, can manifest as no, positive, negative, or sign epistasis[52]. While we know several examples of pairs of mutations exhibiting epistasis of each kind[9,52], we do not know if or how often the nature of epistasis exhibited by a pair of mutations changes with the genomic background.

To answer this question, we pick a pair of mutations and compute the fraction of genomes in which these mutations exhibit (a) Positive epistasis (PE), (b) Negative epistasis (NE), (c) Sign epistasis, and (d) No epistasis (No). Sign epistasis was further classified into (i) Reciprocal Sign epistasis (RSE), (ii) Single Sign epistasis (SSE) and (iii) Other Sign epistasis (OSE) based on the number of paths restricted by Darwinian evolution (**Figure 1A**).

For example, the mutation pair (G → A at position 3 and T → C at position 7) exhibits PE in 26.08%; NE in 34.36%; RSE in 0.67%; SSE in 2.31%; OSE in 5.00%; and No Epistasis in 31.57% of all high fitness backgrounds. The corresponding figures for low fitness backgrounds are: PE: 19.41%; NE: 22.61%; RSE: 7.65%; SSE: 5.71%; OSE: 13.16%; and No Epistasis: 30.92%. The exact numbers for all mutations pairs are provided in **Supplement Data-File1**.

We repeat this process for all 5184 ($\binom{9}{2}$ *loci* x $(12)^2$) possible pairs of mutations in the 9-base pair region of *folA*. The frequency distribution of the fraction of all genotypes where a pair exhibits each type of epistasis is shown in **Figure 1B** (for high fitness backgrounds) and **Figure 1C** (for low fitness backgrounds).

Barring a few pair of mutations which exhibit positive epistasis in all/nearly all high fitness backgrounds, nature of epistasis between a pair of mutations is strongly dependent on the genetic background (**Figure 1** and **Supplement Table 1** and **Supplement Data-File2**). In high fitness backgrounds, mutation pairs exhibit positive epistasis most frequently (median 41% of the genotypes), followed by negative epistasis (median 23%) and no epistasis (median 16%), with sign epistasis being relatively rare (**Figure 1B**). In low fitness backgrounds, mutation pairs exhibit no epistasis most frequently (median 30%), followed by negative epistasis (median 22%), positive epistasis (median 21%) and other sign epistasis (median 13%) (**Figure 1C**).

Because of this contingency of the nature of epistasis between two mutations on the genetic background, we propose that epistasis is “fluid”.

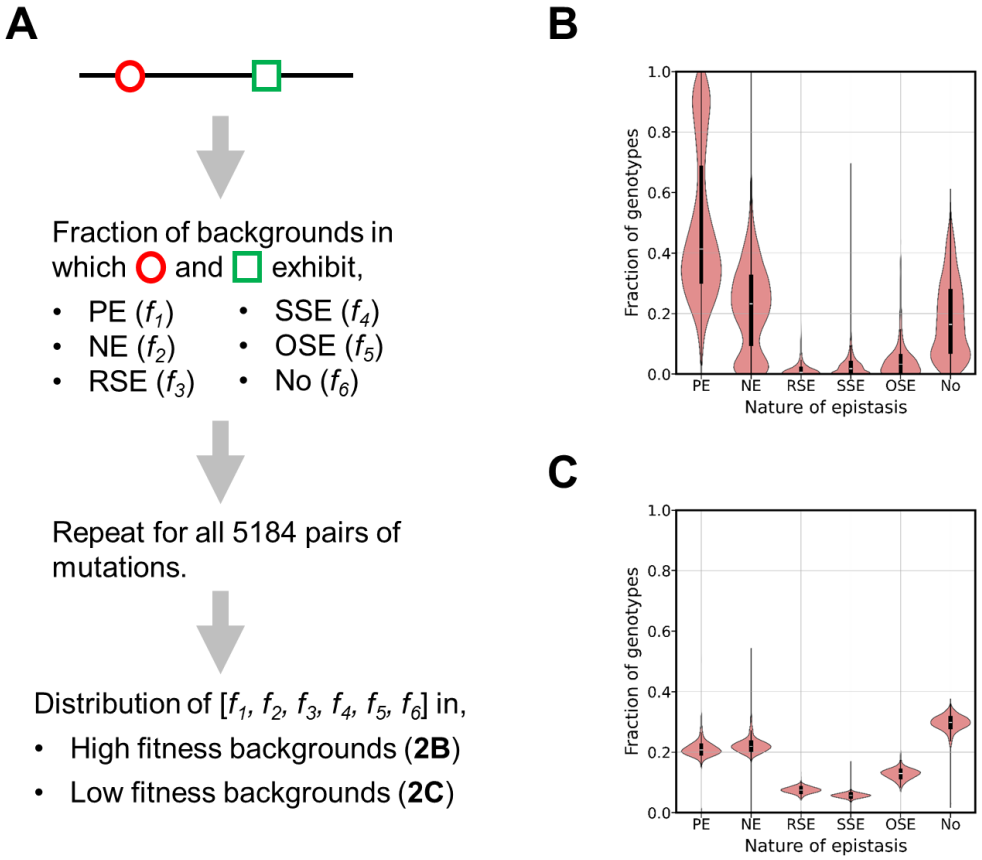

Figure 1. Nature of epistasis exhibited by two mutations is contingent on the genetic background. (A) The nature of epistasis exhibited by two mutations (indicated by red circle and green square) in all 4^7 genetic backgrounds was quantified. Numbers f_1 to f_6 indicate the fraction of genomes in which the two mutations exhibit Positive epistasis (PE), Negative epistasis (NE), Reciprocal Sign epistasis (RSE), Single Sign epistasis (SSE), Other Sign epistasis (OSE), and No epistasis (No), and were calculated for this pair of mutations. The process was repeated for every possible pair of mutations. Distribution of the six fractions f_1 to f_6 in **(B)** high and **(C)** low fitness backgrounds is plotted.

Functionally important sites cause switch in epistasis more frequently.

It has previously been shown that both the secondary structure and location of a residue in a protein dictate the nature of epistasis exhibited by residues[53]. In **Figure 1**, we saw that epistasis changes with genomic background. But, are certain positions more robust to changes in the nature of epistasis than others? In other words, does the location of a mutation in the genetic background dictate the likelihood of epistasis change?

The effect of a locus *X* on changing the nature of epistasis between two mutations was quantified as shown in **Figure 2A**. As shown in **Figure 2B** and **Figure 2C**, upon the introduction of a single mutation at locus *X*, the nature of epistasis switches in more than 50% of cases. The contributions of different loci is different. Positions 4 and 5 on the landscape, which are critical for protein function, are more responsible for changing the nature of epistasis between two mutations, than other sites on the landscape (**Supplement Table 2** and **Supplement Table 3**). This control of

nature of epistasis between other sites by functionally important sites is likely an important factor controlling protein evolution.

The switch of the nature of epistasis is most frequent to positive, negative, or no epistasis (**Supplement Figure S6** and **S7**) (also see **Supplement Data-File3**). Switch to sign epistasis is relatively infrequent. Interestingly, in high fitness backgrounds, mutations at functionally important positions (4 and 5, on the landscape) cause a switch to sign-epistasis more frequently, as compared to a mutation at any of the other seven positions. This pattern is not seen in low fitness backgrounds.

The pervasiveness of the change in epistatic interactions is surprising, and makes prediction of evolutionary trajectories harder. However, we also note that change of nature of epistasis due to a mutation is not unique to *folA*. An analysis of a previously published five-point landscape of beta-lactamase gene[29] shows that change in nature of epistasis in common in intramolecular landscapes (**Supplement Figure S8**). The above results further emphasize that epistasis is "fluid", and that different sites control the switch of epistasis differently.

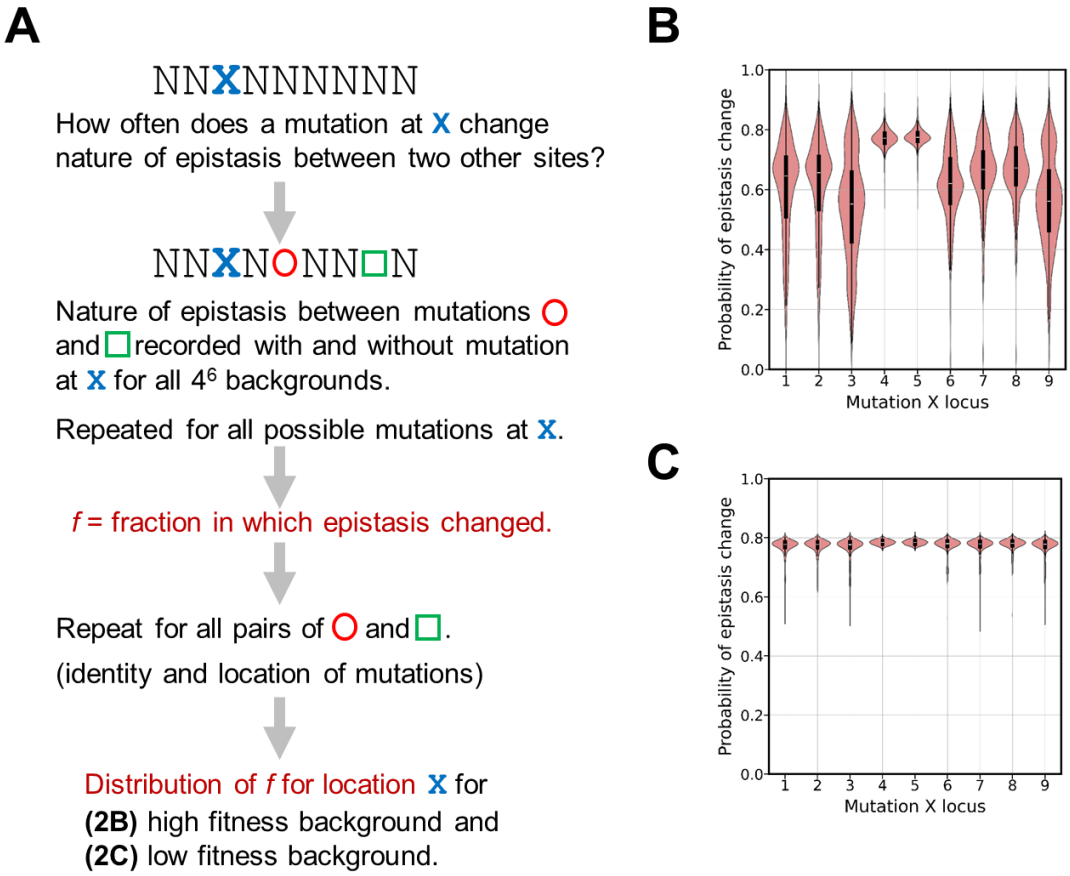

Figure 2. Different sites on the landscape dictate change in epistasis differently. (A) To determine the impact of mutation at locus **X** in changing the nature of epistasis between two mutations (indicated by a red circle and a green square), we took the following approach. For a particular pair of mutations, the six sites (indicated by *N*) were fixed, and the nature of epistasis recorded before and after a mutation at **X**. This process was repeated for all 4^6 backgrounds, and the fraction of genomes in which, upon introduction of a

mutation at **X**, epistasis between the mutations (red circle and green square) changed, was noted. This gives us ***f***, as indicated in the Figure. This process was repeated for all possible pairs of mutations (red circle and green square); giving a distribution of ***f*** for locus **X**. This distribution of ***f*** is shown in **(B)** for high fitness backgrounds and **(C)** for low fitness backgrounds. Functionally important sites (4 and 5) cause change in epistasis more frequently than others.

Synonymous mutations can cause change of nature of epistasis.

While historically synonymous mutations were thought to be neutral[54,55], several studies have demonstrated that they can have a wide range of effects on cellular fitness[56-59]. However, their role in changing epistasis between two mutations has not been studied. To study this in the *folA* landscape, we note that any two mutations lie in either one or two codons in the *folA* sequence. Depending on the location of these two mutations, there is/are one/two codon(s) that remain unaffected by these mutations. We introduce all possible synonymous mutations in the unaffected codon(s) and ask how frequently does the nature of mutation change? Through this, we seek the likelihood that a synonymous mutation will change the nature of epistasis between two mutations (**Figure 3A**).

In our analysis, we consider T**A**A ↔ T**G**A and TA**A** ↔ TA**G** as synonymous mutations (see discussion). All three codons encode for a stop signal for translation.

Our results shows that synonymous changes can cause change in nature of epistasis frequently (**Figures 3B** and **3C**). In high fitness backgrounds, the likelihood of change of nature of epistasis upon introduction of a synonymous mutation is ~0.45 (**Figure 3B**). In low fitness backgrounds, this number is ~0.75 (**Figure 3C**).

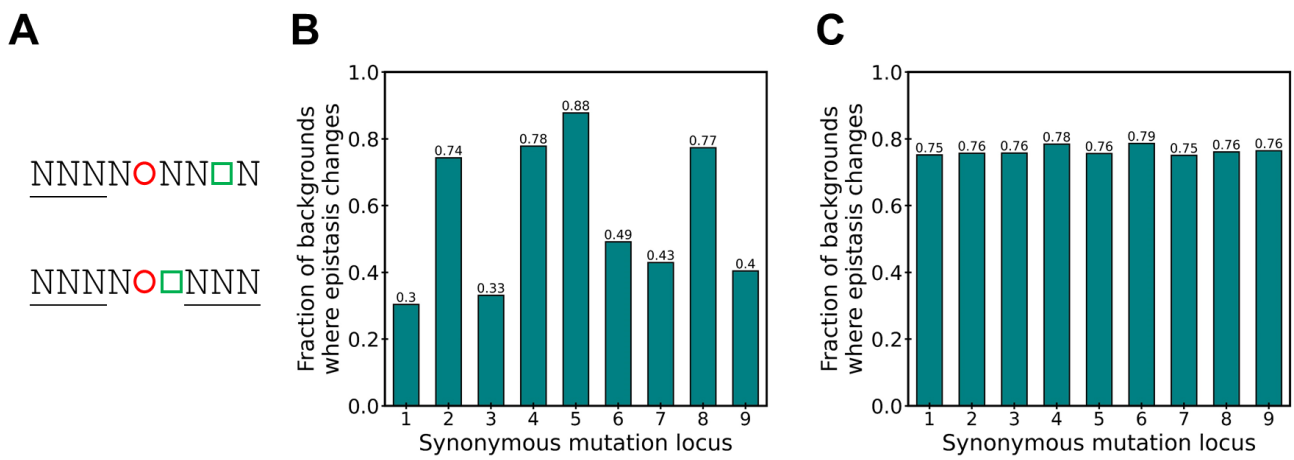

Figure 3. Introduction of a synonymous mutation changes the nature of epistasis between two interacting mutations. (A) For a given sequence, two mutations (red circle and green square) could occur in either one or two codons, resulting in two or one unaltered codons (underlined in figure). All possible synonymous mutations were introduced in the unaltered codon(s) and change of nature of epistasis (if any) between the two mutations was noted. This was done for all backgrounds and all pairs of two mutations. Probability that a synonymous mutation at locus *X* leads to a change in nature of epistasis between two

mutations in **(B)** high and **(C)** low fitness backgrounds is shown. Mutations resulting in T**A**A ↔ T**G**A and TA**A** ↔ TA**G** are considered as synonymous mutations.

These results clearly demonstrate that change of nature of epistasis can take place via the acquisition of even a synonymous mutation, reaffirming "fluidity" in the nature of epistasis. This property of epistasis makes predicting evolution difficult. However, in recent years, epistasis has been shown to exhibit several statistical patterns, collectively termed as global epistasis[1]. We next check how these patterns hold on the *folA* landscape.

Mutations on the landscape exhibit diminishing returns and increasing costs.

Global epistasis suggests that the fitness effect of a mutation is a decreasing function of the background fitness[7,60 61] (**Figure 4A**). In **Figure 4B**, a point represents fitness effect of a single mutation (y-axis) against the fitness of the genotype in which the mutation is introduced (x-axis) on the *folA* landscape. This is done for all mutations in all backgrounds. The line in red indicates the linear fit between the two variables. In both high and low fitness backgrounds, the negative slope of the line indicates presence of global epistasis, although the correlation is not very strong (R^2 = 0.1284 for high fitness backgrounds, & R^2 = 0.1236 for low fitness backgrounds).

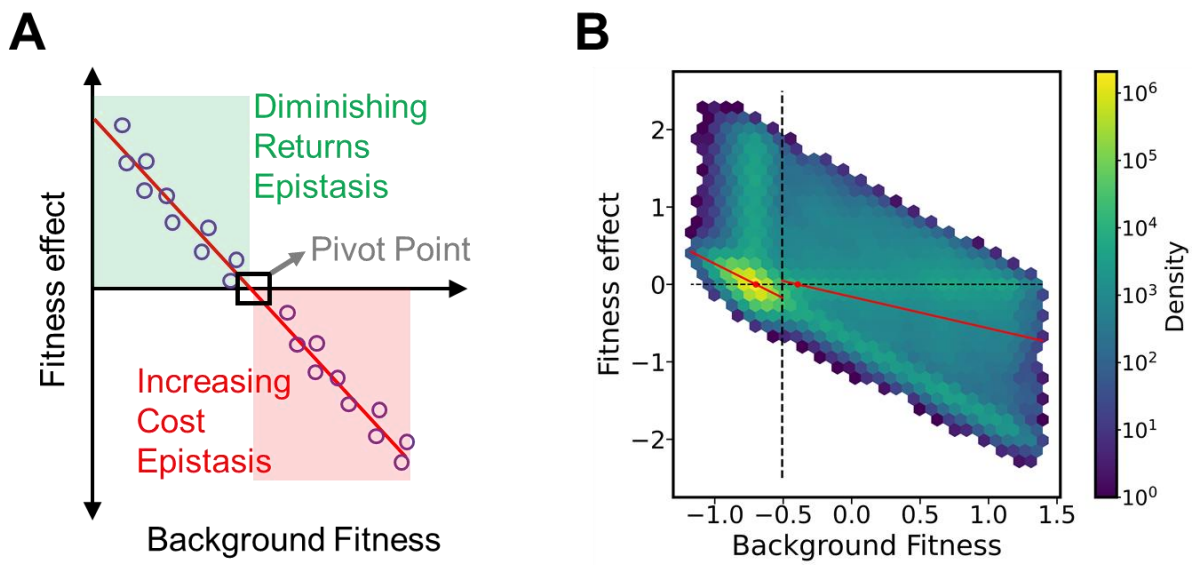

Figure 4. The *folA* fitness landscape exhibits weak patterns of global epistasis. (A) Cartoon to illustrate statistical patterns of global epistasis. The beneficial effect of a mutation decreases with an increase in background fitness (Diminishing Returns Epistasis). Beyond a certain background fitness (Pivot Point), the mutation becomes deleterious and its deleterious effects increase as the fitness of the genetic background increases (Increasing Costs Epistasis). **(B)** Mutational effects vs. the background fitness show weak correlation for low fitness (R^2 = 0.1236) (left of dotted line) and high fitness (R^2 = 0.1284) (right of dotted line) backgrounds. The red lines show the best linear fits for the two groups.

Only a small fraction of mutations exhibit global epistasis/ "binary" nature of global epistasis.

We next investigate the effect on fitness of each of the 108 (12 possible mutations at each of the 9 sites) mutations on all possible genotypes. **Figure 5** shows that only a few mutations exhibit strong patterns of global epistasis. These mutations are primarily (14 out of 16) at positions (nucleotide 4 and 5) which are functionally most important for *folA*[46]. An overwhelming majority of mutations (77/108) exhibit no correlation ($R^2 < 0.2$) between their fitness effect and the fitness of the background on which they occur (**Supplement Figure S9** and **Supplement Table S4**).

Therefore, mutations exhibit one of two states depending on whether they follow global epistasis, or not. This indicates the "binary" nature of mutations. In the case of *folA* landscape, only mutations at nucleotide positions critical for function exhibit global epistasis.

A recent work[61] demonstrated that the growth rate at which any mutation switches from being beneficial to deleterious is conserved for all mutations. This growth rate, around which the nature of mutation changes from being beneficial to deleterious was referred to as the pivot growth rate. While the mechanistic origins of pivot growth rate are yet unknown, this phenomenon likely represents a deep fundamental "rule" of how a cell works. We test the existence of the pivot growth rate for each of the 108 mutations in the *folA* landscape. Most (~80%) mutations pivot from being beneficial to deleterious at a growth rate -0.657 ± 0.0657 (**Figure 6** and **Supplement Table 4**). The presence of a pivot growth rate despite poor statistical correlations between background fitness and the fitness effect of a mutation is a surprising observation.

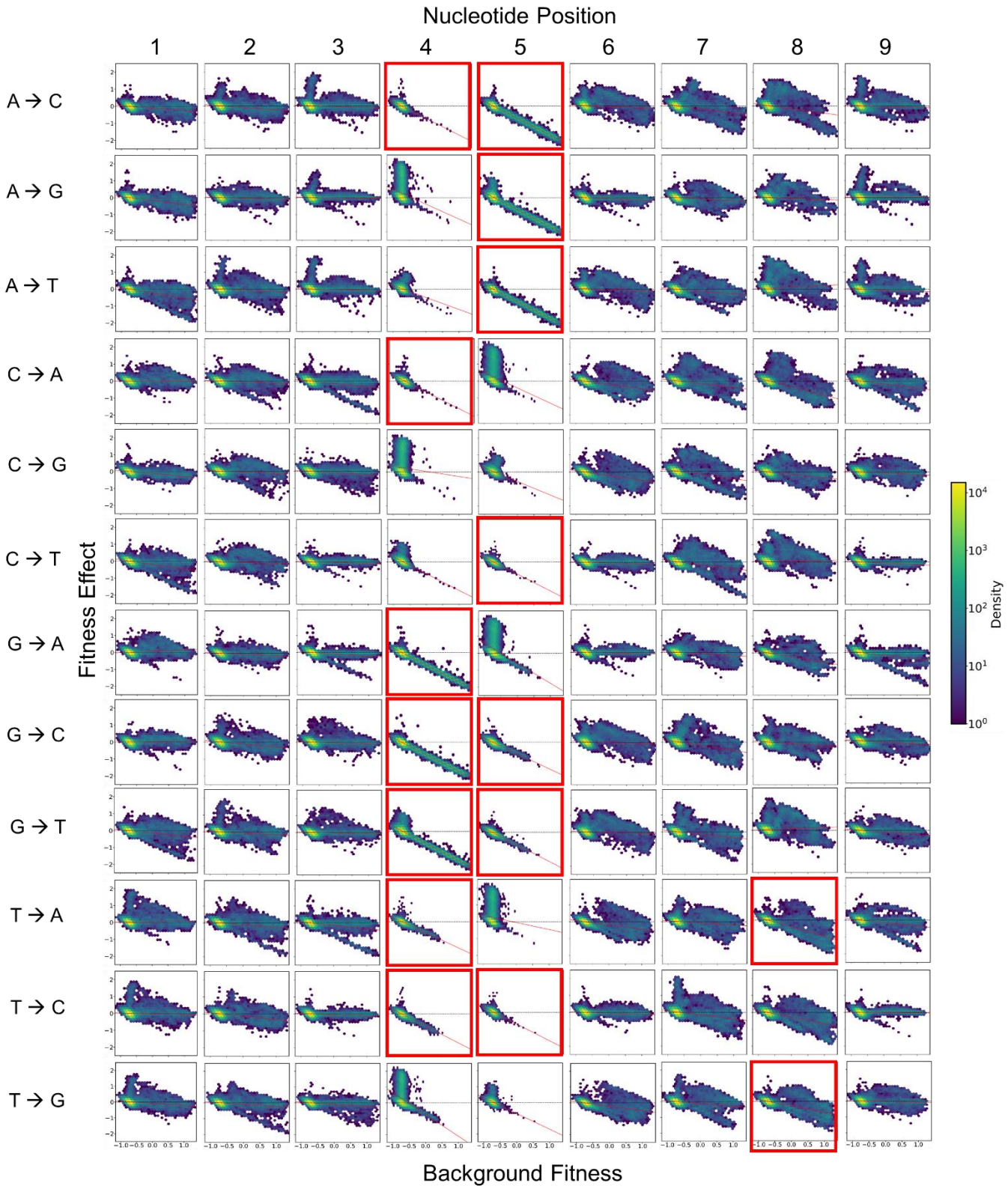

Figure 5. A small fraction of mutations follow global epistasis. Only 16 (14 of which are at position 4 and 5) (highlighted in red boxes) out of 108 possible mutations exhibit statistically significant patterns (R^2 > 0.4) of global epistasis. See **Supplement Table 1** for statistics for each mutation.

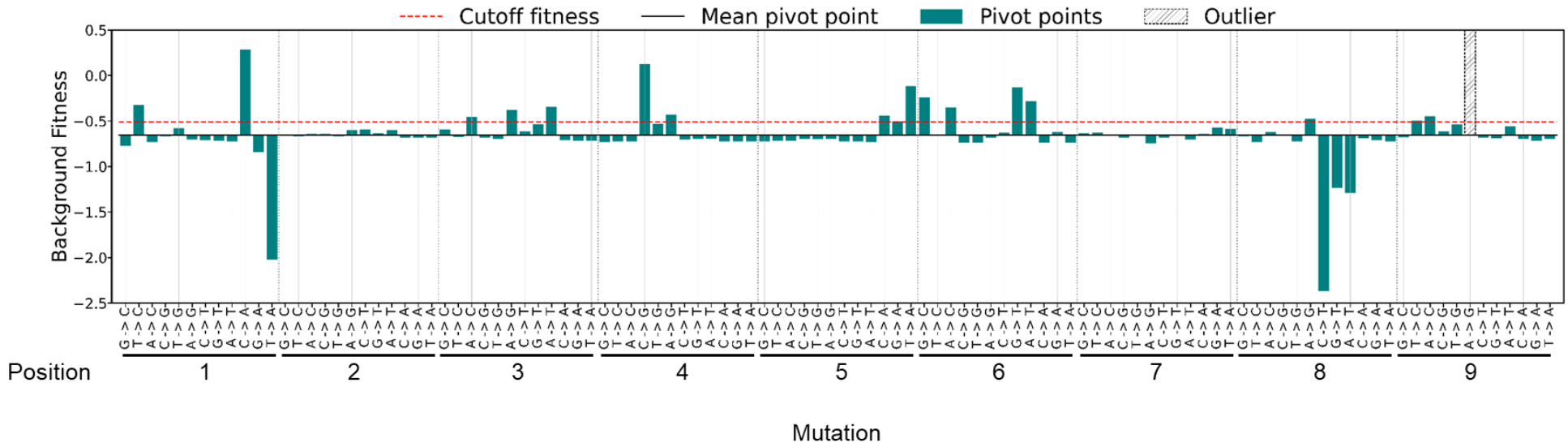

Figure 6. Most mutations switch from being beneficial to deleterious at the pivot point. Background fitness (y-axis) at which each of the 108 mutations (x-axis) switches from being beneficial to deleterious. The black solid line shows the average fitness (-0.657) (or pivot) at which a mutation changes from being beneficial to deleterious. Individual bars show the deviation from the mean pivot point for each mutation. More than 80% (86 out of 108) of the mutations exhibit the pivot fitness in the range -0.657 +/- 0.0657. The dotted line in red indicates the growth rate used in Papkou et al[40] to differentiate between high and low fitness variants.

Predicting DFE from phenotype.

The “binary” and “fluid” nature of epistasis discussed in this work make prediction of evolutionary trajectories difficult. As shown in **Figure 5**, only functionally important sites exhibit global epistasis. For most mutations, fitness effects are idiosyncratic. Hence, in the absence of knowledge of the sites which exhibit global epistasis, predictions are likely going to be accurate. From the context of the “fluidity” of epistasis, in the absence of complete knowledge of the genetic background, it is barely possible to comment on nature of epistasis between two mutations. Thus, these two features of epistasis make evolution unpredictable. We now ask if there exists any statistical pattern at all, which would enable the prediction of evolution.

In this context, we define a quantity called “phenotypic DFE”, which represents the collective DFE of all genotypes exhibiting near identical fitness (**Figure 7A**). To compute the phenotypic DFE, we binned all genotypes in non-overlapping narrow fitness intervals (see methods). DFE of each genotype was computed by introducing all possible 27 (by introducing all three mutations at each of the 9 positions for a given genotype) mutations. DFE of all genotypes in one interval was averaged to obtain the phenotypic DFE. We next ask how robustly the DFE of a particular genotype of near identical fitness can be predicted solely from the phenotypic DFE.

To test this, we compute the phenotypic DFE from randomly sampled 90% of the genotypes with fitness between f_o and $f_o + \Delta$. The DFE of the remaining 10% of the genotypes in a fitness window was computed separately, and each DFE was compared with the phenotypic DFE. The likelihood of equivalence of each genotype’s DFE with the corresponding phenotypic DFE was estimated as the *p*-value of the Mann-Whitney U test. This comparison gives us a distribution of *p*-values, for each background fitness.

The phenotypic DFE of high fitness backgrounds comprised of two peaks. The first peak corresponds to mutations with large deleterious effects, whose magnitude increases with increasing background fitness (**Figures 7B**). The second peak is roughly centered at fitness effect ~ 0. For low fitness backgrounds, the DFE comprised of only one peak, whose mean decreased as the background fitness increased (**Figure 7C**).

Phenotypic DFE is better predicted, than fitness effects of individual mutations, by the background fitness of the genotype. The fraction of genomes for which the phenotypic DFE is statistically significantly different from the actual DFE reduces as the background fitness increases (**Figure 7D**). This effect can also be seen from the distribution of *p*-values of comparisons between phenotypic DFE and actual DFEs, for different background fitness (**Supplement Figure 10**).

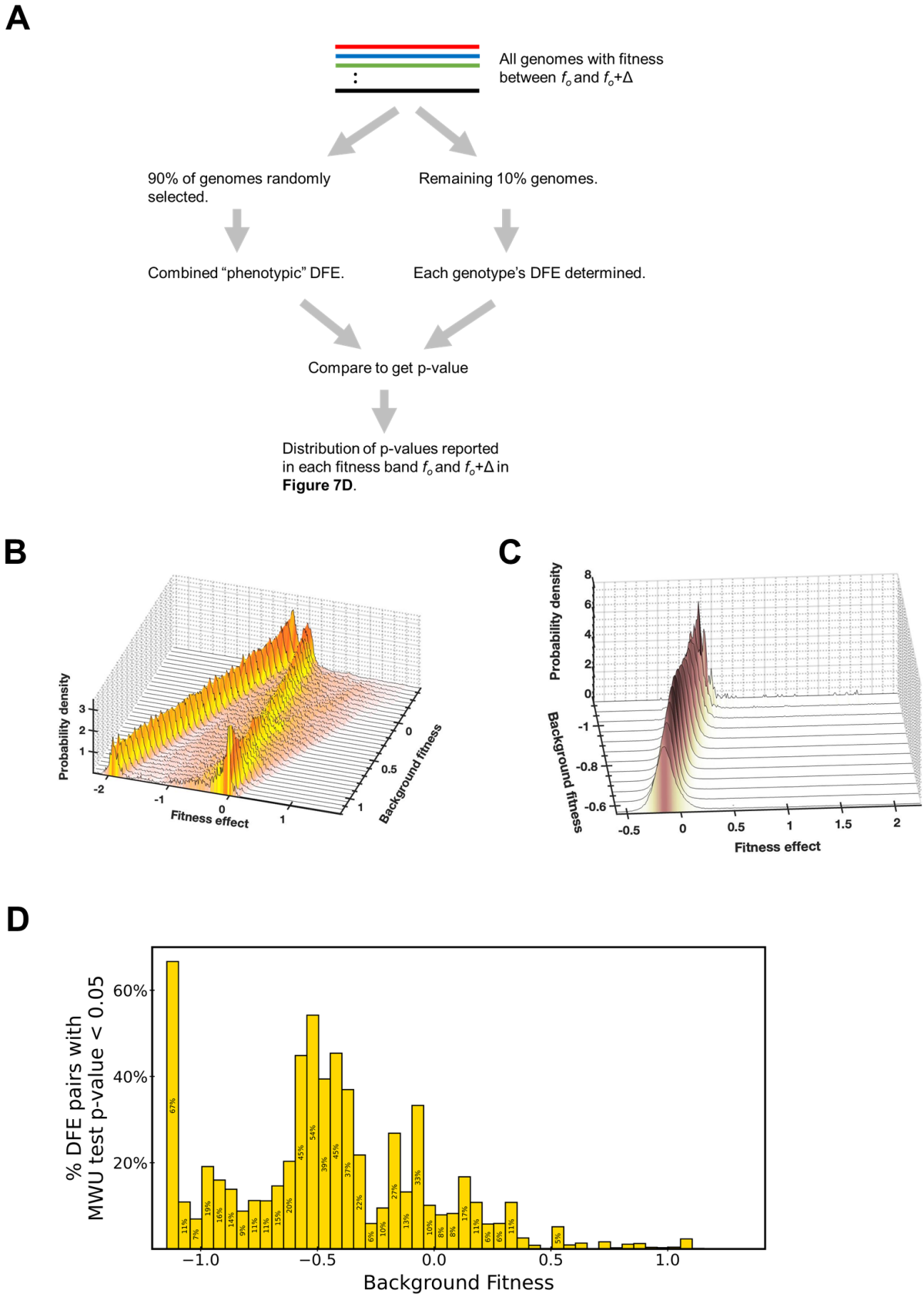

Figure 7. "Phenotypic DFE" is a predictor of DFE. (A) Backgrounds with near identical fitness were randomly distributed in two groups comprising 90% and 10% of all backgrounds. The first group was used to define a "phenotypic DFE" (a mean DFE of all genomes in the group). The phenotypic DFE was compared with individual DFEs in the 10% group, and the *p*-value distribution of this comparison was obtained. The

process was repeated for genotypes with other fitness. **(B)** Phenotypic DFE of high fitness backgrounds exhibited two peaks. The first peak was at fitness effect ~ 0. The second peak comprised of deleterious mutations, whose magnitude increased with increasing background fitness. **(C)** Phenotypic DFE of low fitness backgrounds exhibited a single peak, whose mean decreased with increasing background fitness. **(D)** Percent of DFEs in a fitness window with p-value < 0.05 when compared with the Phenotypic DFE**.** As fitness increases, Phenotypic DFE becomes a better predictor of DFE of a genotype.

Discussion

Genotype-phenotype mapping, especially of proteins, is of great interest in evolutionary biology, cell biology, and genetics[11,46,62-66]. The underlying rules that dictate this mapping are a combination of individual mutation effects, epistatic interactions between the mutations, and between the mutations and the genetic background they occur in[7,29,35,60,61,67-70]. The extent and ubiquity of epistatic interactions is of particular interest, because they are mostly unpredictable and have a direct effect on the shape of the fitness landscape, and consequently, adaptation[34,38,71-76]. Therefore, in order to understand the statistical rules which govern epistasis on a landscape, we analyzed a recently published *folA* landscape which was constructed by quantifying fitness of more than 260,000 sequences in a 9-base pair region[40]. Mutations in these nine base pairs have been reported to be adaptive[47,49].

Navigability of fitness landscapes decides the likelihood of a population to reach the global fitness maximum. Most fitness landscapes indicate that protein landscapes are rugged, and hence, protein evolution is constrained[47,70,77-79]. But, fitness landscapes are constructed by considering a handful of sites, and there is theoretical evidence to suggest that mutations in other "dimensions" could enable populations to navigate valleys in the landscape[11]. Fisher suggested the same in his correspondence with Wright[80].

A particularly confounding mode of epistasis is sign epistasis, which alters the qualitative nature of a mutational effect, and creates valleys in fitness landscapes. In the *folA* landscape, sign epistasis frequently changes to positive or negative epistasis, indicating “fluidity” in its effects. We also report that synonymous mutations can change the nature of epistasis between two existing mutations. Hence, even if valleys existed, they are not that difficult to navigate. Similar findings have been reported in the past - deleterious mutations making peaks accessible, and neutral mutations being adaptive[81,82]. Additionally, synonymous mutations are known to have fitness effects by changing protein amount (by altering mRNA stability or translation rates)[56,83-87] or structure[88-92]. Our results provide a novel mechanism via which synonymous mutations are relevant, for driving evolutionary change and controlling disease states[93-96].

Interestingly, our analysis shows that the nature of epistasis between two mutations can change by simply changing the stop codon (TAA ↔ TGA or TAA ↔ TAG). This observation holds even when the stop codon is encoded by the first three bases of the 9-bp landscape, and the interacting mutations whose nature of epistasis changes are in the subsequent two codons (which are presumably not even translated). This indicates that the change in the nature of epistasis has likely not do with the protein synthesis but with the mRNA and its effect on cellular fitness.

Premature termination of translation has been known to destabilize the entire transcript[97]. In eukaryotes, elaborate mechanisms are present to deal with potentially toxic effects of truncated proteins[98-101]. However, prokaryotes lack these mechanisms. In fact, a recent study shows that in *E. coli* under stress, despite a premature stop codon in the gene sequence, stop codon read-through rates may be as high as 80%, due to a high probability of a mismatch at a premature stop codon[102]. This is a likely explanation for change of epistasis (or change of fitness) due to alternate premature stop codons in *folA*, when *E. coli* is grown in antibiotic stress.

Neutral mutations can improve evolvability of a sequence, and hence aid the movement on a fitness landscape[103-108]. However, we do not see any evolvability enhancing mutations in the 9-bp *folA* landscape (**Supplement Figure S11**).

Predicting evolution is a longstanding goal in the field, and global epistasis patterns offered a tool to predict epistatic effects. However, our analyses show that global epistasis often does not hold. Instead, we observe that sequences with similar fitness have similar DFEs, offering some predictability, based on a macroscopic trait, of the adaptive potential of a population[109-111]. The means of these DFEs decreased linearly with an increase in background fitness, despite the mutations on this landscape not following global epistasis patterns (**Supplement Figure S12**).

We show that the navigability of a landscape can change via mutations in the same protein. Can it also change via mutations elsewhere on the genome? High-dimensional landscapes are necessary to answer this question. Additionally, increasing the dimensionality of the landscape is likely going to provide qualitatively new perspectives of adaptation.

Methods

Calculating Hamming Distance.

To calculate the hamming distance between two sequences of equal length, we compare the sequences and count the number of dissimilar loci.

Construction of sequence spaces and landscapes.

In order to construct an n-base pair sequence space from a parent p-base pair sequence space ($p > n$), we choose any $p - n$ loci in the parent sequence space and find all variants in the parent space in which the selected loci are fixed (these loci contain a same sequence). The set of these selected variants are assigned their one hamming distance neighbours to construct a new $n\text{-}base$ sequence space.

By repeating this process over all $\binom{p}{p-n}$ combinations of selecting the loci to be fixed and the 4^{p-n} permutations of choosing the fixed sequence in each case, we are able to break the parent sequence space into $\left[\binom{p}{p-n} \times 4^{p-n}\right]$ n-base pair sequence spaces (**Supplement Table 2**).

For our study, we use a nine-base pair parent sequence space to generate n-base pair sequence spaces $(1 \leq n \leq 9)$. Landscapes are generated by mapping fitness values of each sequence in a sequence space corresponding to the ones assigned in the empirical fitness landscape generated by Papkou et al. using a nine base pair gene.

In rare cases where fitness value was not known, we disregarded those variants in all studies.

Finding peaks in a fitness landscape.

<u>Number of Peaks in Landscape</u>: To count the number of peaks in fitness landscapes, we find the number of variants that have a higher fitness value than all its one hamming distance neighbours in that landscape.

<u>Peak Probability</u>: To quantify the probability of encountering a fitness peak in any landscape, we find the ratio of number of peaks found to the total number of sequences in that fitness landscape.

Expected number of peaks: In our case (as we are dealing with 4 letter genome), the expected number of peaks in maximally rugged (uncorrelated) NK landscapes is found by $\left(\frac{number\ of\ functional\ variants}{3n+1}\right)$ for an n-base pair landscape[20]. The number of peaks predicted by NK landscapes is rounded off to nearest integer.

Fitness effect of mutations.

To find the fitness effect of a mutation acting on a genotype, we start with a set of all genotypes in the empirical Papkou et al. fitness landscape which would result in a new sequence formed following the mutation. In all such genotypes, we find the background fitness f_b and the fitness of resulting mutant f_m. The fitness effect of this mutation in this background is determined as the fitness difference between the mutant and the background ($s = f_m - f_b$).

Finding phenotypic DFE: We found the Distribution of Fitness Effects for variants lying in a small slice of fitness value (for our study, we kept the range of this slice to be 0.05). The DFE was constructed by analysing the frequency of fitness effects of all mutations acting on the backgrounds lying in the selected range.

Finding genotypic DFE: For any 9 length genotypic sequence, we found the $9 loci \times 3 bases = 27$ mutations that may result in a new sequence. We used the frequency of fitness effect of all these mutations to constitute the distribution of fitness effects for any genotype.

Non parametric tests: We found the p-value of the two parameter KS test and the Mann–Whitney *U* test using the *python scipy.stats* library.

Linear regression: We found the pivot point fitness of individual mutations via linear regression of background fitness and the selection coefficient of the mutation using the “LinearRegression” model from python sklearn.linear_model library.

Epistasis as function of genetic background.

Classifying Epistasis: To quantify the epistasis present in a mutation pair acting on two differing genetic loci, we compute the fitness effect of the individual mutations on a genetic background and the cumulative fitness effect of the two mutations on the same genetic background. If the fitness effect of the individual mutations were s_1 and s_2, while the cumulative effect of the two mutations was s_{12}, then we classify the epistasis into following categories,

1. No Epistasis: If $|s_{12} - (s_1 + s_2)| < 0.05$ i.e. we assume no epistasis was present if the cumulative fitness effect of the two mutations was about the same as the two mutations acting independently on the genetic background.

2. Sign Epistasis: If $s_{12} \times (s_1 + s_2) < 0$ i.e. we classify sign epistasis if the cumulative effect of the two mutations lead to a different sign of fitness effect than the sum of individual fitness effects of the two mutations on the same background. For example, if the combined effect of the mutation pair was beneficial even though the sum of fitness effects of the two mutations on the same background was deleterious, and vice versa.

3. We further classified this epistasis into three categories,
   A. Reciprocal Sign Epistasis: If $s_1 < 0$, $s_2 < 0$ and $s_{12} > 0$ i.e. if the cumulative effect of the two mutations lead the background to a higher fitness, but both shortest paths to this higher fitness point are blocked to darwinian evolution.
   B. Single Sign Epistasis: If exclusively $s_1 < 0$ or $s_2 < 0$ and $s_{12} > 0$ i.e. if exactly one of the two shortest paths leading background to higher fitness are blocked to darwinian evolution.
   C. Other Sign Epistasis: All other cases cases classified to sign epistasis.

4. Positive Epistasis: If $s_{12} > s_1 + s_2$ i.e. if the combined fitness effect of the two mutations was more beneficial / less deleterious than the sum of individual effects of the mutation pair on the genetic background.

5. Negative Epistasis: If $s_{12} < s_1 + s_2$ i.e. if the combined fitness effect of the two mutations was less beneficial / more deleterious than the sum of individual effects of the mutation pair on the genetic background.

Epistasis Change with Genetic Background: Having the epistasis dossier generated for all mutation pairs, we compiled all cases of *Positive, Negative, Sign* and *No Epistasis*. For each of these cases, we select the genetic backgrounds and their one mutant neighbours such that their differing mutation locus is unrelated to the loci involved in Epistasis (In our case, we can find $(9 - 2)\,loci \times 3\,bases = 21$ such neighbours for each background).

We then check the nature of epistasis in each of these 21 neighbours on the same mutation pair, and quantify the number of these cases in which nature of epistasis changes.

Finding paths in a sequence space.

Set of all variants: We start by listing all the mutations required to convert the starting sequence $A$to the target sequence T. If h denotes the minimum number of mutations required to change sequence A to T, then for each step $s \in 1, \cdots, h-1$, we list all variants in the sequence space which are at s hamming distance from the starting variant and $h-s$ hamming distance from the target sequence. The resulting set includes all variants involved at each step for shortest traversal from A to T.

Having the set of all variants at each step in traversal of sequence A to T, we recursively find all $h!$ permutations of paths such that each step only allows one base change while leading the sequence to the target in minimum number of steps.

Effect of neutral mutations on evolvability.

To perform this study, we compiled all the $(9 loci \times 4 bases) = 36 mutations$ that are possible among all sequences of *folA* landscape. We then listed all the backgrounds on which these mutations change the genotype, but showcase neutral fitness effect (magnitude of fitness effect < 0.05).

We then allow a second mutation which changes both the background and the mutant genotype. If the selection coefficient of the second mutation on the background is s_1 and on the neutral mutant is s_2, then the relative increase in evolvability is quantified as: $\left(\frac{s_2 - s_1}{|s_1|}\right)$.

Consider a variant A and a neutral mutation X which results in a variant B $(A \neq B | s_{AB}| < 0.05)$. For any mutation Y taking place on genotypes A and B forming A' and B' such that $(A \neq A' B \neq B')$, we quantify the relative change in evolvability of A from mutation Y due to a neutral mutation X as $\left(\frac{s_{BB'} - s_{AA'}}{|s_{AA'}|}\right)$.

Finding synonymous mutations

We identified the all synonymous codons, i.e. the codons that encode the same amino acid / termination function. For each codon c_i , we then found the list of all synonymous codons which are at 1 Hamming distance from the codon c_i.

Using this data, we were able to identify the set of synonymous mutations for each codon.

Change in mode of epistasis due to synonymous mutation in an extrinsic codon

For any background in DHFR gene, we tested all mutation pairs (*A* and *B*) and identified the type of epistasis exhibited. Since these two mutations can mutate a minimum of one and a maximum of two codons in the 9 base pair gene, at least one codon remains un-mutated. For the mutation pair, we found the un-mutated codon(s), and derived the set of all possible synonymous mutations on the codon(s) of the given background.

We noted the number of instances where introduction of a synonymous mutation (X) at a particular locus (on an un-mutated codon) does or does not change the nature of epistasis for the background and given mutation pair.

We did this analysis separately (all / high fitness / low fitness) backgrounds, their possible mutation pairs (*A* and *B*) and their respective synonymous mutations (X) to identify the overall probability of epistasis change due to synonymous mutation on each locus.

Codes.

All codes used in this work and Supplement Data Files are available at: https://github.com/SainiSupreet/Ecoli-folA-DHFR. The “readme” file at the repository gives details of how to run codes.

Funding.

This work was funded by a Wellcome Trust/DBT (India Alliance) grant (Award Number: IA/S/19/2/504632) to SS. NMR was funded by Prime Minister's Research Fellowship (PMRF ID 1301163).

Acknowledgements.

We thank Christian Landry and Krishna Swamy for feedback on the manuscript.

References

1 Johnson, M. S., Reddy, G. & Desai, M. M. Epistasis and evolution: recent advances and an outlook for prediction. *BMC Biol* **21**, 120, doi:10.1186/s12915-023-01585-3 (2023).

2 Bank, C. Epistasis and Adaptation on Fitness Landscapes. *Annual Review of Ecology, Evolution, and Systematics* **53** (2022).

3 Ostman, B., Hintze, A. & Adami, C. Impact of epistasis and pleiotropy on evolutionary adaptation. *Proc Biol Sci* **279**, 247-256, doi:10.1098/rspb.2011.0870 (2012).

4 Good, B. H. & Desai, M. M. The impact of macroscopic epistasis on long-term evolutionary dynamics. *Genetics* **199**, 177-190, doi:10.1534/genetics.114.172460 (2015).

5 Johnson, M. S. & Desai, M. M. Mutational robustness changes during long-term adaptation in laboratory budding yeast populations. *Elife* **11**, doi:10.7554/eLife.76491 (2022).

6 Wunsche, A. *et al.* Diminishing-returns epistasis decreases adaptability along an evolutionary trajectory. *Nat Ecol Evol* **1**, 61, doi:10.1038/s41559-016-0061 (2017).

7 Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519-1522, doi:10.1126/science.1250939 (2014).

8 Park, Y., Metzger, B. P. H. & Thornton, J. W. The simplicity of protein sequence-function relationships. *Nat Commun* **15**, 7953, doi:10.1038/s41467-024-51895-5 (2024).

9 Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. *Protein Sci* **25**, 1204-1218, doi:10.1002/pro.2897 (2016).

10 de Visser, J. A., Cooper, T. F. & Elena, S. F. The causes of epistasis. *Proc Biol Sci* **278**, 3617-3624, doi:10.1098/rspb.2011.1537 (2011).

11 Greenbury, S. F., Louis, A. A. & Ahnert, S. E. The structure of genotype-phenotype maps makes fitness landscapes navigable. *Nat Ecol Evol* **6**, 1742-1752, doi:10.1038/s41559-022-01867-z (2022).

12 Srivastava, M. & Payne, J. L. On the incongruence of genotype-phenotype and fitness landscapes. *PLoS Comput Biol* **18**, e1010524, doi:10.1371/journal.pcbi.1010524 (2022).

13 de Visser, J. A. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet* **15**, 480-490, doi:10.1038/nrg3744 (2014).

14 Blanquart, F. & Bataillon, T. Epistasis and the Structure of Fitness Landscapes: Are Experimental Fitness Landscapes Compatible with Fisher's Geometric Model? *Genetics* **203**, 847-862, doi:10.1534/genetics.115.182691 (2016).

15 Fraisse, C. & Welch, J. J. The distribution of epistasis on simple fitness landscapes. *Biol Lett* **15**, 20180881, doi:10.1098/rsbl.2018.0881 (2019).

16 Diaz-Colunga, J. *et al.* Global epistasis on fitness landscapes. *Philos Trans R Soc Lond B Biol Sci* **378**, 20220053, doi:10.1098/rstb.2022.0053 (2023).

17 Park, Y., Metzger, B. P. H. & Thornton, J. W. The simplicity of protein sequence-function relationships. *bioRxiv*, doi:10.1101/2023.09.02.556057 (2024).

18 Van Cleve, J. & Weissman, D. B. Measuring ruggedness in fitness landscapes. *Proc Natl Acad Sci U S A* **112**, 7345-7346, doi:10.1073/pnas.1507916112 (2015).

19 Meger, A. T. *et al.* Rugged fitness landscapes minimize promiscuity in the evolution of transcriptional repressors. *Cell Syst* **15**, 374-387 e376, doi:10.1016/j.cels.2024.03.002 (2024).

20 Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol* **128**, 11-45, doi:10.1016/s0022-5193(87)80029-2 (1987).

21 Hayashi, Y. *et al.* Experimental rugged fitness landscape in protein sequence space. *PLoS One* **1**, e96, doi:10.1371/journal.pone.0000096 (2006).

22 Neidhart, J., Szendro, I. G. & Krug, J. Adaptation in tunably rugged fitness landscapes: the rough Mount Fuji model. *Genetics* **198**, 699-721, doi:10.1534/genetics.114.167668 (2014).

23 Fontanari, D. B. S. a. J. F. Evolutionary dynamics on rugged fitness landscapes: Exact dynamics and information theoretical aspects. *Physical Review E* **80** (2009).
24 Carneiro, M. & Hartl, D. L. Colloquium papers: Adaptive landscapes and protein evolution. *Proc Natl Acad Sci U S A* **107 Suppl 1**, 1747-1751, doi:10.1073/pnas.0906192106 (2010).
25 Franke, J., Klozer, A., de Visser, J. A. & Krug, J. Evolutionary accessibility of mutational pathways. *PLoS Comput Biol* **7**, e1002134, doi:10.1371/journal.pcbi.1002134 (2011).
26 de Visser, J. A., Park, S. C. & Krug, J. Exploring the effect of sex on empirical fitness landscapes. *Am Nat* **174 Suppl 1**, S15-30, doi:10.1086/599081 (2009).
27 Hall, D. W., Agan, M. & Pope, S. C. Fitness epistasis among 6 biosynthetic loci in the budding yeast Saccharomyces cerevisiae. *J Hered* **101 Suppl 1**, S75-84, doi:10.1093/jhered/esq007 (2010).
28 Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress on Genetics.* **1**, 355-366 (1932).
29 Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111-114, doi:10.1126/science.1123539 (2006).
30 Smith, J. M. Natural selection and the concept of a protein space. *Nature* **225**, 563-564, doi:10.1038/225563a0 (1970).
31 Malcolm, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F. & Wilson, A. C. Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature* **345**, 86-89, doi:10.1038/345086a0 (1990).
32 de Visser, J., Hoekstra, R. F. & van den Ende, H. Test of Interaction between Genetic Markers That Affect Fitness in Aspergillus Niger. *Evolution* **51**, 1499-1505, doi:10.1111/j.1558-5646.1997.tb01473.x (1997).
33 Kouyos, R. D., Silander, O. K. & Bonhoeffer, S. Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol Evol* **22**, 308-315, doi:10.1016/j.tree.2007.02.014 (2007).
34 Hietpas, R. T., Jensen, J. D. & Bolon, D. N. Experimental illumination of a fitness landscape. *Proc Natl Acad Sci U S A* **108**, 7896-7901, doi:10.1073/pnas.1016024108 (2011).
35 Starr, T. N., Flynn, J. M., Mishra, P., Bolon, D. N. A. & Thornton, J. W. Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *Proc Natl Acad Sci U S A* **115**, 4453-4458, doi:10.1073/pnas.1718133115 (2018).
36 Gorter, F. A., Aarts, M. G. M., Zwaan, B. J. & de Visser, J. Local Fitness Landscapes Predict Yeast Evolutionary Dynamics in Directionally Changing Environments. *Genetics* **208**, 307-322, doi:10.1534/genetics.117.300519 (2018).
37 Schenk, M. F., Szendro, I. G., Salverda, M. L., Krug, J. & de Visser, J. A. Patterns of Epistasis between beneficial mutations in an antibiotic resistance gene. *Mol Biol Evol* **30**, 1779-1787, doi:10.1093/molbev/mst096 (2013).
38 Buda, K., Miton, C. M. & Tokuriki, N. Pervasive epistasis exposes intramolecular networks in adaptive enzyme evolution. *Nat Commun* **14**, 8508, doi:10.1038/s41467-023-44333-5 (2023).
39 Li, C., Qian, W., Maclean, C. J. & Zhang, J. The fitness landscape of a tRNA gene. *Science* **352**, 837-840, doi:10.1126/science.aae0568 (2016).
40 Papkou, A., Garcia-Pastor, L., Escudero, J. A. & Wagner, A. A rugged yet easily navigable fitness landscape. *Science* **382**, eadh3860, doi:10.1126/science.adh3860 (2023).
41 Flynn, J. M. *et al.* Comprehensive fitness landscape of SARS-CoV-2 M(pro) reveals insights into viral resistance mechanisms. *Elife* **11**, doi:10.7554/eLife.77433 (2022).
42 Wu, N. C., Dai, L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *Elife* **5**, doi:10.7554/eLife.16965 (2016).

43 Pokusaeva, V. O. *et al.* An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. *PLoS Genet* **15**, e1008079, doi:10.1371/journal.pgen.1008079 (2019).
44 Karapanagioti, F., Atlason, U. A., Slotboom, D. J., Poolman, B. & Obermaier, S. Fitness landscape of substrate-adaptive mutations in evolved amino acid-polyamine-organocation transporters. *Elife* **13**, doi:10.7554/eLife.93971 (2024).
45 Flynn, J. M. *et al.* Comprehensive fitness maps of Hsp90 show widespread environmental dependence. *Elife* **9**, doi:10.7554/eLife.53810 (2020).
46 Bershtein, S., Choi, J. M., Bhattacharyya, S., Budnik, B. & Shakhnovich, E. Systems-level response to point mutations in a core metabolic enzyme modulates genotype-phenotype relationship. *Cell Rep* **11**, 645-656, doi:10.1016/j.celrep.2015.03.051 (2015).
47 Tamer, Y. T. *et al.* High-Order Epistasis in Catalytic Power of Dihydrofolate Reductase Gives Rise to a Rugged Fitness Landscape in the Presence of Trimethoprim Selection. *Mol Biol Evol* **36**, 1533-1550, doi:10.1093/molbev/msz086 (2019).
48 Matthews, D. A. *et al.* Dihydrofolate reductase: x-ray structure of the binary complex with methotrexate. *Science* **197**, 452-455, doi:10.1126/science.17920 (1977).
49 Benkovic, S. J., Fierke, C. A. & Naylor, A. M. Insights into enzyme function from studies on mutants of dihydrofolate reductase. *Science* **239**, 1105-1110, doi:10.1126/science.3125607 (1988).
50 Schnell, J. R., Dyson, H. J. & Wright, P. E. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biophys Biomol Struct* **33**, 119-140, doi:10.1146/annurev.biophys.33.110502.133613 (2004).
51 Toprak, E. *et al.* Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat Genet* **44**, 101-105, doi:10.1038/ng.1034 (2011).
52 Phillips, P. C. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**, 855-867, doi:10.1038/nrg2452 (2008).
53 Gonzalez, C. E. & Ostermeier, M. Pervasive Pairwise Intragenic Epistasis among Sequential Mutations in TEM-1 beta-Lactamase. *J Mol Biol* **431**, 1981-1992, doi:10.1016/j.jmb.2019.03.020 (2019).
54 Kimura, M. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet Res* **11**, 247-269, doi:10.1017/s0016672300011459 (1968).
55 King, J. L. & Jukes, T. H. Non-Darwinian evolution. *Science* **164**, 788-798, doi:10.1126/science.164.3881.788 (1969).
56 Agashe, D., Martinez-Gomez, N. C., Drummond, D. A. & Marx, C. J. Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *Mol Biol Evol* **30**, 549-560, doi:10.1093/molbev/mss273 (2013).
57 Kristofich, J. *et al.* Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. *PLoS Genet* **14**, e1007615, doi:10.1371/journal.pgen.1007615 (2018).
58 Lebeuf-Taylor, E., McCloskey, N., Bailey, S. F., Hinz, A. & Kassen, R. The distribution of fitness effects among synonymous mutations in a gene under directional selection. *Elife* **8**, doi:10.7554/eLife.45952 (2019).
59 Walsh, I. M., Bowman, M. A., Soto Santarriaga, I. F., Rodriguez, A. & Clark, P. L. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc Natl Acad Sci U S A* **117**, 3528-3534, doi:10.1073/pnas.1907126117 (2020).
60 Johnson, M. S., Martsul, A., Kryazhimskiy, S. & Desai, M. M. Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science* **366**, 490-493, doi:10.1126/science.aay4199 (2019).

61 Ardell, S., Martsul, A., Johnson, M. S. & Kryazhimskiy, S. Environment-independent distribution of mutational effects emerges from microscopic epistasis. *bioRxiv*, doi:10.1101/2023.11.18.567655 (2024).
62 Ahnert, S. E. Structural properties of genotype-phenotype maps. *J R Soc Interface* **14**, doi:10.1098/rsif.2017.0275 (2017).
63 Mak, H. C. & Justman, Q. Genotype-Phenotype Mapping Meets Single Cell Biology. *Cell Syst* **4**, 1-2, doi:10.1016/j.cels.2017.01.008 (2017).
64 Pigliucci, M. Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. *Philos Trans R Soc Lond B Biol Sci* **365**, 557-566, doi:10.1098/rstb.2009.0241 (2010).
65 Alberch, P. From genes to phenotype: dynamical systems and evolvability. *Genetica* **84**, 5-11, doi:10.1007/BF00123979 (1991).
66 Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425-431, doi:10.1126/science.1180823 (2010).
67 Hinz, A., Amado, A., Kassen, R., Bank, C. & Wong, A. Unpredictability of the Fitness Effects of Antimicrobial Resistance Mutations Across Environments in Escherichia coli. *Mol Biol Evol* **41**, doi:10.1093/molbev/msae086 (2024).
68 Rauscher, R. *et al.* Positive epistasis between disease-causing missense mutations and silent polymorphism with effect on mRNA translation velocity. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2010612118 (2021).
69 Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E. & Cooper, T. F. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193-1196, doi:10.1126/science.1203801 (2011).
70 Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165-1174 (2005).
71 Sailer, Z. R. & Harms, M. J. Molecular ensembles make evolution unpredictable. *Proc Natl Acad Sci U S A* **114**, 11938-11943, doi:10.1073/pnas.1711927114 (2017).
72 Morris, S. C. Evolution: like any other science it is predictable. *Philos Trans R Soc Lond B Biol Sci* **365**, 133-145, doi:10.1098/rstb.2009.0154 (2010).
73 Harms, M. J. & Thornton, J. W. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* **512**, 203-207, doi:10.1038/nature13410 (2014).
74 Miton, C. M. & Tokuriki, N. How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci* **25**, 1260-1272, doi:10.1002/pro.2876 (2016).
75 Lassig, M., Mustonen, V. & Walczak, A. M. Predicting evolution. *Nat Ecol Evol* **1**, 77, doi:10.1038/s41559-017-0077 (2017).
76 Sarkisyan, K. S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397-401, doi:10.1038/nature17995 (2016).
77 Starr, T. N. *et al.* ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature* **603**, 913-918, doi:10.1038/s41586-022-04464-z (2022).
78 Park, Y., Metzger, B. P. H. & Thornton, J. W. Epistatic drift causes gradual decay of predictability in protein evolution. *Science* **376**, 823-830, doi:10.1126/science.abn6895 (2022).
79 Macken, C. A. & Perelson, A. S. Protein evolution on rugged landscapes. *Proc Natl Acad Sci U S A* **86**, 6191-6195, doi:10.1073/pnas.86.16.6191 (1989).
80 Provine, W. B. *Sewall Wright and Evolutionary Biology*. (Chicago University Press, 1986).
81 Despres, P. C. *et al.* Compensatory mutations potentiate constructive neutral evolution by gene duplication. *bioRxiv*, doi:10.1101/2024.02.12.579783 (2024).
82 Douglas, S. M., Chubiz, L. M., Harcombe, W. R. & Marx, C. J. Identification of the potentiating mutations and synergistic epistasis that enabled the evolution of inter-species cooperation. *PLoS One* **12**, e0174345, doi:10.1371/journal.pone.0174345 (2017).
83 Ando, H., Miyoshi-Akiyama, T., Watanabe, S. & Kirikae, T. A silent mutation in mabA confers isoniazid resistance on Mycobacterium tuberculosis. *Mol Microbiol* **91**, 538-547, doi:10.1111/mmi.12476 (2014).

84 Kershner, J. P. *et al.* A Synonymous Mutation Upstream of the Gene Encoding a Weak-Link Enzyme Causes an Ultrasensitive Response in Growth Rate. *J Bacteriol* **198**, 2853-2863, doi:10.1128/JB.00262-16 (2016).
85 Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in Escherichia coli. *Science* **324**, 255-258, doi:10.1126/science.1170160 (2009).
86 Goodman, D. B., Church, G. M. & Kosuri, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342**, 475-479, doi:10.1126/science.1241934 (2013).
87 Bailey, S. F., Alonso Morales, L. A. & Kassen, R. Effects of Synonymous Mutations beyond Codon Bias: The Evidence for Adaptive Synonymous Substitutions from Microbial Evolution Experiments. *Genome Biol Evol* **13**, doi:10.1093/gbe/evab141 (2021).
88 Jiang, Y. *et al.* How synonymous mutations alter enzyme structure and function over long timescales. *Nat Chem* **15**, 308-318, doi:10.1038/s41557-022-01091-z (2023).
89 Lan, P. D. *et al.* Synonymous Mutations Can Alter Protein Dimerization Through Localized Interface Misfolding Involving Self-entanglements. *J Mol Biol* **436**, 168487, doi:10.1016/j.jmb.2024.168487 (2024).
90 Deane, C. M. & Saunders, R. The imprint of codons on protein structure. *Biotechnol J* **6**, 641-649, doi:10.1002/biot.201000329 (2011).
91 Yu, C. H. *et al.* Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Mol Cell* **59**, 744-754, doi:10.1016/j.molcel.2015.07.018 (2015).
92 Buhr, F. *et al.* Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Mol Cell* **61**, 341-351, doi:10.1016/j.molcel.2016.01.008 (2016).
93 Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* **12**, 683-691, doi:10.1038/nrg3051 (2011).
94 Hunt, R. C., Simhadri, V. L., Iandoli, M., Sauna, Z. E. & Kimchi-Sarfaty, C. Exposing synonymous mutations. *Trends Genet* **30**, 308-321, doi:10.1016/j.tig.2014.04.006 (2014).
95 Supek, F., Minana, B., Valcarcel, J., Gabaldon, T. & Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**, 1324-1335, doi:10.1016/j.cell.2014.01.051 (2014).
96 Sharma, Y. *et al.* A pan-cancer analysis of synonymous mutations. *Nat Commun* **10**, 2569, doi:10.1038/s41467-019-10489-2 (2019).
97 Nilsson, G., Belasco, J. G., Cohen, S. N. & von Gabain, A. Effect of premature termination of translation on mRNA stability depends on the site of ribosome release. *Proc Natl Acad Sci U S A* **84**, 4890-4894, doi:10.1073/pnas.84.14.4890 (1987).
98 Shi, M. *et al.* Premature Termination Codons Are Recognized in the Nucleus in A Reading-Frame Dependent Manner. *Cell Discov* **1**, 15001-, doi:10.1038/celldisc.2015.1 (2015).
99 Kim, J. H. *et al.* SMG-6 mRNA cleavage stalls ribosomes near premature stop codons in vivo. *Nucleic Acids Res* **50**, 8852-8866, doi:10.1093/nar/gkac681 (2022).
100 Arribere, J. A. & Fire, A. Z. Nonsense mRNA suppression via nonstop decay. *Elife* **7**, doi:10.7554/eLife.33292 (2018).
101 Garcia-Rodriguez, R. *et al.* Premature termination codons in the DMD gene cause reduced local mRNA synthesis. *Proc Natl Acad Sci U S A* **117**, 16456-16464, doi:10.1073/pnas.1910456117 (2020).
102 Romero Romero, M. L. *et al.* Environment modulates protein heterogeneity through transcriptional and translational stop codon readthrough. *Nat Commun* **15**, 4446, doi:10.1038/s41467-024-48387-x (2024).
103 Payne, J. L. & Wagner, A. The causes of evolvability and their evolution. *Nat Rev Genet* **20**, 24-38, doi:10.1038/s41576-018-0069-z (2019).
104 Pigliucci, M. Is evolvability evolvable? *Nat Rev Genet* **9**, 75-82, doi:10.1038/nrg2278 (2008).

105 Raynes, Y., Gazzara, M. R. & Sniegowski, P. D. Mutator dynamics in sexual and asexual experimental populations of yeast. *BMC Evol Biol* **11**, 158, doi:10.1186/1471-2148-11-158 (2011).
106 van Nimwegen, E., Crutchfield, J. P. & Huynen, M. Neutral evolution of mutational robustness. *Proc Natl Acad Sci U S A* **96**, 9716-9720, doi:10.1073/pnas.96.17.9716 (1999).
107 Tenaillon, O., Toupance, B., Le Nagard, H., Taddei, F. & Godelle, B. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* **152**, 485-493, doi:10.1093/genetics/152.2.485 (1999).
108 Wagner, A. Evolvability-enhancing mutations in the fitness landscapes of an RNA and a protein. *Nat Commun* **14**, 3624, doi:10.1038/s41467-023-39321-8 (2023).
109 Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O. & Desai, M. M. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc Natl Acad Sci U S A* **109**, 4950-4955, doi:10.1073/pnas.1119910109 (2012).
110 Good, B. H. & Desai, M. M. Deleterious passengers in adapting populations. *Genetics* **198**, 1183-1208, doi:10.1534/genetics.114.170233 (2014).
111 Hallatschek, O. The noisy edge of traveling waves. *Proc Natl Acad Sci U S A* **108**, 1783-1787, doi:10.1073/pnas.1013529108 (2011).