# Experimental Determination and Prediction of the Fitness Effects of Random Point Mutations in the Biosynthetic Enzyme HisA

Erik Lundin,[1] Po-Cheng Tang,[1] Lionel Guy,[1] Joakim Näsvall,[1] and Dan I. Andersson*,[1]

[1]Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

*Corresponding author: E-mail: dan.andersson@imbim.uu.se.

**Associate editor:** Jianzhi Zhang

## Abstract

**The distribution of fitness effects of mutations is a factor of fundamental importance in evolutionary biology. We determined the distribution of fitness effects of 510 mutants that each carried between 1 and 10 mutations (synonymous and nonsynonymous) in the *hisA* gene, encoding an essential enzyme in the L-histidine biosynthesis pathway of *Salmonella enterica*. For the full set of mutants, the distribution was bimodal with many apparently neutral mutations and many lethal mutations. For a subset of 81 single, nonsynonymous mutants most mutations appeared neutral at high expression levels, whereas at low expression levels only a few mutations were neutral. Furthermore, we examined how the magnitude of the observed fitness effects was correlated to several measures of biophysical properties and phylogenetic conservation. We conclude that for HisA: (i) The effect of mutations can be masked by high expression levels, such that mutations that are deleterious to the function of the protein can still be neutral with regard to organism fitness if the protein is expressed at a sufficiently high level; (ii) the shape of the fitness distribution is dependent on the extent to which the protein is rate-limiting for growth; (iii) negative epistatic interactions, on an average, amplified the combined effect of nonsynonymous mutations; and (iv) no single sequence-based predictor could confidently predict the fitness effects of mutations in HisA, but a combination of multiple predictors could predict the effect with a SD of 0.04 resulting in 80% of the mutations predicted within 12% of their observed selection coefficients.**

*Key words:* mutation, fitness, *Salmonella enterica*, HisA protein.

## Introduction

The distribution of fitness effects (DFE) of mutations is of fundamental importance in understanding major evolutionary questions regarding, for example, disease development (Yue et al. 2005; Eyre-Walker et al. 2006), the maintenance of healthy population sizes of endangered species (Silander et al. 2007), and adaptive evolution (Jacquier et al. 2013; Firnberg et al. 2014; Knight et al. 2015). Fitness landscapes are vast and have as many dimensions as there are possible mutations in a system (de Visser and Krug 2014). The effect of mutations can be classified into three major categories: beneficial, neutral, or deleterious. Beneficial mutations are rare and tend to be exponentially distributed, whereas deleterious mutations tend to show a bimodal U-shaped distribution with most mutations being lethal or close to neutral (Eyre-Walker and Keightley 2007).

Previous studies have used many different estimates of fitness. For example, fitness in the presence of antibiotics (minimum inhibitory concentrations, MIC) have been assessed, but with limited resolution in the assay as the reported MIC values have fixed discrete levels (Walkiewicz et al. 2012; Jacquier et al. 2013). Fitness effects of mutations in green fluorescent protein (GFP) were measured by over-expression in a nonnative organism (*Escherichia coli*), with a measure of fitness (fluorescence of GFP) not selected for by evolution in *E. coli* (Sarkisyan et al. 2016). DNA sequence data have been used either on only a small region of a single gene (Hietpas et al. 2011) or a whole genome (Keightley and Eyre-Walker 2010). The DFE of mutations in a whole genome has been studied in yeast (Wloch et al. 2001), bacteria (Elena et al. 1998), and viruses with great accuracy (Sanjuan et al. 2004; Sanjuan 2010), but without explaining the underlying cause(s) of the observed pattern.

When measuring the DFE of mutations in a protein, it is important to distinguish between the fitness of the individual protein and organism fitness (Soskine and Tawfik 2010; Jiang et al. 2013; Boucher et al. 2016). Thus, organism fitness could either be correlated or uncorrelated to protein fitness depending on the experimental set-up and the extent to which the studied protein is rate-limiting for growth under the particular study condition. Consequently, it is expected that the shapes of the DFEs may vary extensively depending on the sensitivity of the experimental set-up and how rate-limiting the studied protein is for growth.

**Open Access**

We determined the DFE of a total of 510 unique mutants in the *hisA* gene of *Salmonella enterica*. The DFE was analyzed in relation to several key determinants of protein function and the in silico predictability of fitness was assessed by 40 different methods. We also addressed the impact of accumulating mutations on the fitness and studied how epistatic effect between mutations will affect the fitness as mutations accumulate. The significance of protein level was studied to investigate how protein and organismal fitness are related.

## Results and Discussion

### Experimental System

HisA is an isomerase that catalyzes the fourth step in the L-histidine biosynthesis pathway, catalyzing $N'$-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) to $N'$-((5′-phosphoribulosyl) formimino)-5-aminoimidazole-4-carboxamide-ribonucleotide (PRFAR) (fig. 1A). *S. enterica* HisA was chosen for this study because of its conditional essentiality and because HisA function can be experimentally set to limit organism fitness (measured by growth rate) in a selective environment, that is, growth in minimal media lacking histidine. Furthermore, the $(\beta\alpha)_8$ barrel structure is common in enzymes, with ~10% of all enzymes having this structure (Höcker et al. 2001), making the findings for HisA potentially applicable to many enzymes. Mutations were introduced by error-prone PCR, resulting in a small bias toward transversions (48.7%) over transitions (46.8%), and a small amount of deletions (4.1%) and insertions (0.4%) (supplementary table S1, Supplementary Material online). The mutagenized gene variants were introduced through $\lambda$ red recombineering at the neutral (i.e., the insertions had no effect on growth under the tested conditions, data not shown) *cobA* locus and placed under the control of a strong constitutive promoter. A total of 510 mutants with 1 to 10 mutations per *hisA* gene were isolated and used for further analysis (fig. 1B). Eighty-one of the mutant clones had single amino acid substitutions in HisA and were chosen for a more detailed analysis under low expression conditions. This was achieved by placing them under the control of the inducible L-arabinose promoter ($P_{araBAD}$) and then organism fitness was measured by determining exponential growth rates in minimal media lacking histidine (fig. 1C). Selection coefficients were calculated as $s =$ relative growth rate$_{mutant} - 1$ (relative growth rate$_{wild-type}$). Mutants with a fitness cost smaller than the growth rate assay resolution (selection coefficient $|s| < 0.05$) could not be distinguished from wild-type fitness and were further examined by competition experiments which increased the assay sensitivity by 10-fold ($|s| < 0.005$).

### DFEs for All 510 Mutants

The most common method of presenting DFE data is to bin the data for plotting in a histogram. Several previous studies have presented DFE data in this way (Elena et al. 1998; Sanjuan et al. 2004; Eyre-Walker and Keightley 2007; Peris et al. 2010; Hietpas et al. 2011; Jacquier et al. 2013; Firnberg et al. 2014; Knight et al. 2015). However, this requires a
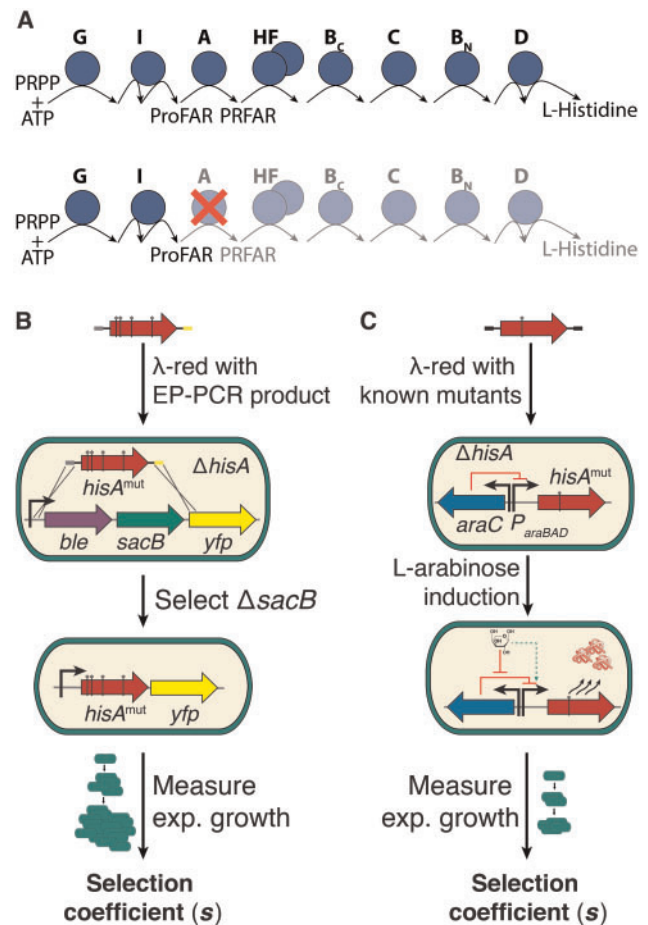


**FIG. 1.** Outline of the experiment. (*A*) HisA catalyzes the fourth step in the L-histidine biosynthesis pathway, the conversion of ProFAR to PRFAR. In minimal medium, deletion of *hisA* is lethal and mutations in *hisA* affect the growth rate. (*B*) Experimental system of this study. Error-prone PCR (EP-PCR) was used to introduce random mutations into the *hisA* gene of *Salmonella enterica*. The gene was introduced into the chromosome with $\lambda$ red recombineering by counter selecting *sacB*. The fitness of each mutant was measured by relative exponential growth rate in M9 minimal medium. The relative fitness (*s*) per generation was calculated (0 for neutral, positive for beneficial, and negative for deleterious mutations). (*C*) Mutated genes with only one mutation were placed under control of the L-arabinose inducible promoter $P_{araBAD}$ and exponential growth rate was measured under growth limiting concentrations of L-arabinose.

subjective selection of bin width and a starting point that can affect the appearance of the data (Silverman 1986). To minimize bias in describing the observed DFE we used a kernel density estimation, a less arbitrary representation of data (Silverman 1986; Fix and Hodges 1989; Silverman and Jones 1989) (fig. 2A, C, and E).

The DFE of all 510 mutants (containing between 1 and 10 mutations) was bimodal at high expression levels (fig. 2A and B). Similar distributions have been reported for mutations in an RNA virus (Sanjuan et al. 2004), TEM-1 $\beta$-lactamase (Jacquier et al. 2013), whole genome mutations in yeast (Wloch et al. 2001), Hsp90 in yeast (Jiang et al. 2013; Bank et al. 2014), and the *araC, D*, and *E* genes in *S. enterica* (Lind et al. 2016). In HisA, the fitness effects of many of the
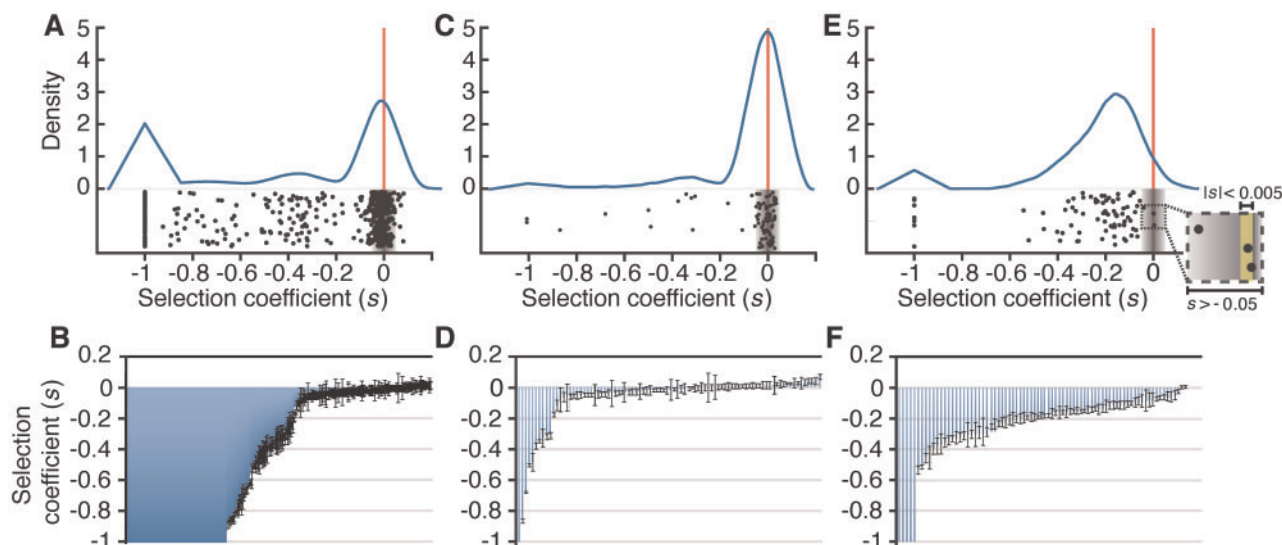
**Fig. 2.** Distribution of fitness effects (DFE) of mutations in HisA. (*A, C,* and *E*) Data are shown as black dots and their distribution as kernel density estimation (blue line). The gray shaded areas mark the neutral fitness effects (i.e., fitness effects smaller than the assay resolution) and the red bars mark $s = 0$. (*B, D,* and *F*) Data are represented by histograms with one bar per strain in the order of increasing fitness. Error bars represent the SD of eight replicates. (*A* and *B*) At high expression levels, the distribution of fitness effects for 510 mutants (containing between 1 and 10 mutations) is bimodal with one mode centered around neutral ($s = 0$) and one at lethal ($s = -1$). (*C* and *D*) At high expression levels, for mutants with only one mutation ($n = 81$), the distribution of fitness effects is unimodal centered around neutral. (*E* and *F*) At low expression levels, the fitness distribution of mutants with only one mutation ($n = 81$) is bimodal with one mode shifted toward deleterious effects ($s < 0$) and one mode at lethal effect. Mutants with exponential growth rate indistinguishable from wildtype levels (gray-shaded area around $s = 0$) were measured in a direct competition experiment for higher accuracy (yellow shaded area).

mutations (44%), were indistinguishable from neutral ($|s| < 0.05$), 26% were moderately deleterious and 30% were lethal. The average fitness costs per each nonsynonymous mutation calculated from the complete data set (510 mutants) was 20.5% and 10.5%, respectively when including and excluding lethal mutations (table 1). No beneficial mutations ($s > 0.05$) could be detected.

## DFEs for a Subset of 81 Mutants with Single Amino Acid Substitution Mutations at High and Low Expression

When analyzing only the 81 single amino acid substitution mutations (fig. 2C and D) at high expression levels, the distribution was unimodal with a majority of the neutral mutations, a tail toward highly deleterious mutations and an average fitness cost of amino acid substitutions of 9.5% and 4.8%, respectively, when including and excluding 4.9% (4/81) mutations that are lethal (table 1). We hypothesized that the observed apparent robustness of HisA at high expression (i.e., 66 out of the 81 [81%] mutants appeared neutral, fig. 2C and D), was because any reduction in protein activity caused by the mutation was masked by a high concentration of the enzyme. Thus, it is expected that a reduction in enzyme specific activity would not be detectable until the total activity (specific activity × enzyme concentration) falls below a certain threshold and only then a change in organism fitness (growth rate) would be observed. In other words, high expression can buffer a reduction in protein fitness, leaving the organismal fitness unaffected (fig. 3).

At high expression levels in our constructed strains, wildtype HisA total activity is not limiting for growth as compared with a *S. enterica* wild-type strain. To assess how protein fitness correlates with organism fitness, we made the activity of HisA strongly rate-limiting for growth. To achieve this, a wild-type copy of *hisA* was first placed under control of the L-arabinose inducible promoter $P_{araBAD}$ and then the growth rate of cells expressing wild-type *hisA* at different levels (conferred by varying the amount of the inducer compound L-arabinose) was measured. The relationship between organism growth and L-arabinose concentration was sigmoidal (supplementary fig. S1, Supplementary Material online), and at expression levels limiting for growth (approximately $\leq 0.1$ mg/ml L-arabinose), changes in protein activity would directly affect growth rate whereas at high expression levels (approximately $> 0.3$ mg/ml L-arabinose), changes in protein activity would give smaller changes in the growth rate. We then placed all genes carrying a single nonsynonymous mutation under the control of the same L-arabinose inducible promoter and measured the growth rate at a concentration of L-arabinose (0.1 mg/ml, see supplementary fig. S1, Supplementary Material online) that is limiting for growth with a wild-type copy of *hisA*. At this expression level, growth rate is reduced to 72% compared with the rate at high expression conditions, and any reduction in specific activity or concentration of the enzyme would result in a decreased growth rate. The resulting distribution with limiting expression (fig. 2E and F) was bimodal with most mutations resulting in reduced fitness (lethal 9%, deleterious 88.5%) and only a small fraction being indistinguishable from neutral (2.5%). Under

**Table 1.** Average Fitness Cost per Mutation.

| | | Including Lethal Mutations (%) | Excluding Lethal Mutations (%) |
|---|---|---|---|
| Based on all 510 mutants at high expression level | Both nonsynonymous and synonymous mutations | 20.0 | 6.9 |
| | Only nonsynonymous mutations[a] | 20.5 | 10.5 |
| Based on 81 single amino acid substitution mutants | High expression level | 9.5 | 4.8 |
| | Low expression level | 27.2 | 20.4 |

[a]Given fitness values assume epistatic effects between mutations and are calculated accordingly.
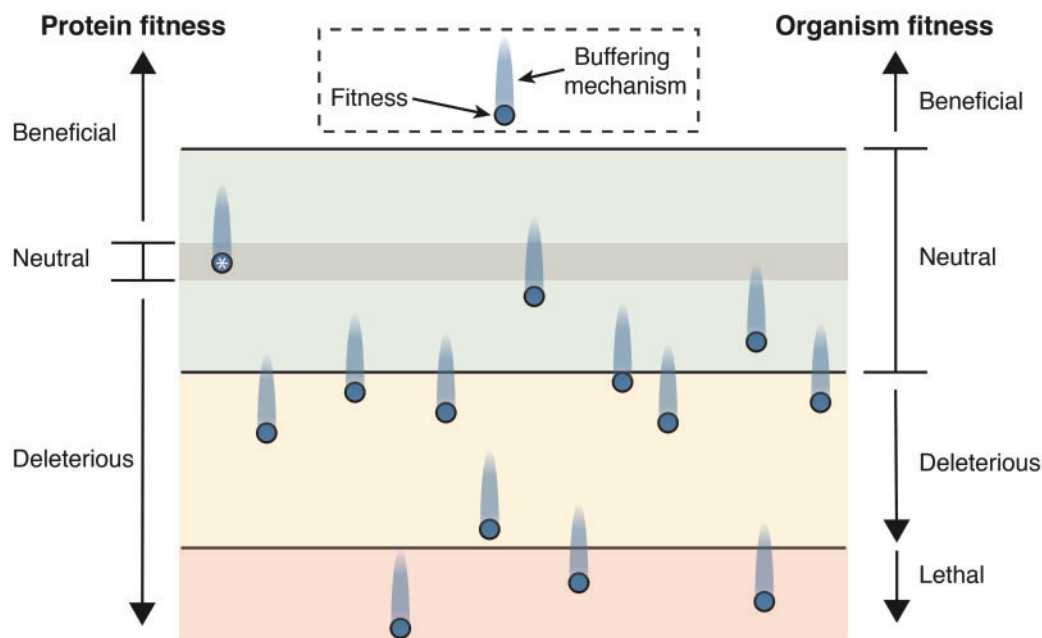


**FIG. 3.** Cartoon of the proposed relationship between protein and organism fitness. Blue dots represent the fitness of different mutants of the same enzyme, measured as protein fitness (left Y-axis) or organism fitness (right Y-axis). The total enzymatic activity (i.e., specific activity × enzyme concentration) of the wild-type protein (marked with an asterisk) is high enough such that it is not rate-limiting for growth, that is, there are no effects on organismal fitness (green area). When the total enzymatic activity of the protein falls below a threshold value (line between green and yellow area), the protein becomes rate-limiting for growth and organism fitness decreases (yellow area). At a protein fitness level, some mutations have no effect on fitness (neutral; within the gray area), many mutations are deleterious (below the gray area) and beneficial mutations (above the gray area) are expected to be rare. At an organism fitness level, various intrinsic buffering mechanisms (illustrated by the blue gradient area extending upward from protein fitness) can mask the effects of mutations. Expression of a surplus of the protein or regulatory feedback mechanisms can buffer against deleterious mutations, such that mutations that are deleterious for protein fitness can appear neutral for organism fitness (green area). Mutations with a large effect on the protein fitness, having a deleterious effect on organism fitness (yellow area), can for small effects on organismal fitness be buffered by the same mechanisms (dots in yellow area with a blue gradient extending into the green area) resulting in a neutral effect on original fitness (green area). The same applies to mutations lethal to organismal fitness (red area) that in some cases can be rescued by buffering mechanisms and be measured as strongly deleterious but not lethal (dots in red area with a blue gradient extending into the yellow area). Similarly, if enough of the protein is already expressed, mutations that are beneficial on protein fitness are unlikely to be beneficial on organism fitness. In addition, chaperonins that assist in folding of misfolded proteins can mask the effects of deleterious mutations.

these low-expression conditions the average fitness cost of amino acid substitutions was 27.2% and 20.4%, respectively, when including and excluding lethal mutations (table 1).

### Comparison of HisA DFE with Other Systems

Below we summarize and compare the mutational effects in HisA with studies of other experimental systems (see supplementary table S2, Supplementary Material online). Since the experimental set-ups differ quite extensively with regard to, for example, assay sensitivity and the extent to which the studied function is rate-limiting for growth, any observed differences could potentially result from

differences in assay conditions (discussed further in the section Why do DFEs differ in shape and magnitude between different proteins? below).

#### Beneficial Mutations

We found no fitness-increasing amino acid substitutions in HisA. This lack of beneficial mutations has previously been reported for other proteins, including the enzymes TEM-1 β-lactamase (Jacquier et al. 2013) and AraD (Lind et al. 2016), the fluorescent protein avGFP (Sarkisyan et al. 2016) and ribosomal proteins S20 and L1 (Lind et al. 2010).

However, in a study by Firnberg et al. (2014), 7% of the mutations in the TEM-1 $\beta$-lactamase were found to be beneficial. In two other Ara proteins, AraC and AraE, up to 8% of the mutations were found to be beneficial (Lind et al. 2016). Beneficial mutations have also been reported in both RNA and DNA viruses (Sanjuan et al. 2004; Peris et al. 2010), in *Pseudomonas fluorescens* (Kassen and Bataillon 2006; McDonald et al. 2011) and in yeast Hsp90 (Jiang et al. 2013; Bank et al. 2014), but they are thought to be very rare in humans (Zhang and Li 2005). Hence, beneficial mutations have been seen at various challenging conditions, both at the whole genome and protein levels.

### Lethal Mutations
The fraction of lethal mutations in HisA was 9%, similar to that reported in other bacterial proteins, 4–13% (Soskine and Tawfik 2010; Jacquier et al. 2013; Firnberg et al. 2014; Lind et al. 2016; Sarkisyan et al. 2016). However, the fraction of lethal mutations varies extensively between systems as exemplified by the absence of lethal mutations in ribosomal proteins S20 and L1 in *S. enterica* (Lind et al. 2010). While in viruses, lethal mutations can account for up to 40% of the mutations (Sanjuan et al. 2004).

### Neutral Mutations
Regarding neutral mutations, HisA showed a very low fraction with only 2.5% appearing indistinguishable from wild-type. This is similar to ribosomal proteins S20 and L1 in *S. enterica*, where 5% of the mutations were neutral (Lind et al. 2010). In other systems, the estimated fraction of neutral mutations for microorganisms is 26–56% (Sanjuan et al. 2004; Peris et al. 2010; Firnberg et al. 2014; Lind et al. 2016), apart from ribosomal proteins S20 and L1 from *S. enterica*, where only 5% neutral mutations were found (Lind et al. 2010). The fraction of neutral mutations at the whole genome level were 4% in enteric bacteria (Charlesworth and Eyre-Walker 2006), 16% in *Drosophila* (Eyre-Walker 2002), and 44–57% in humans (Charlesworth 2009; Bataillon and Bailey 2014).

### Why Do DFEs Differ in Shape and Magnitude between Different Proteins?
Combined these above studies show that the shape and magnitude of the DFE can differ extensively between experimental systems and species, raising the question of what the underlying reasons for these differences are. One important explanation is that the sensitivity of the assay systems used can vary at least 100-fold between studies (detection of $|s| = 0.005–0.5$ depending on study), which obviously can have a considerable impact on the apparent DFE. For example, with a less sensitive assay, the fraction of apparently neutral mutations and the robustness of the system would be overestimated (i.e., many of the mutations that are classified as neutral are in fact weakly deleterious or beneficial). Thus, it is likely that in many studies with limited assay sensitivity, the fraction of apparently neutral mutations is overestimated. Furthermore, if the readout of protein fitness is growth

rate, the expression level of the protein studied and the extent to which it is rate-limiting for growth will have a strong effect on the shape and magnitude of the DFE. The fitness cost of varying protein concentration is different between different proteins (Keren et al. 2016) and as a result the same change in protein specific activity might have different effects on organismal fitness for different proteins and organisms. That is, as shown in the present study under high-level expression of HisA, 86% (70/81) of the amino acid substitutions appeared neutral whereas at low expression level only 2.5% (2/81) of the mutations appeared neutral. Hence, an important conclusion from this work is that when measuring the distribution of effects of mutations on the function of a single protein (i.e., distribution of protein fitness effects), the concentration of the protein needs to be reduced such that it is rate-limiting for growth (fig. 3 and supplementary fig. S1, Supplementary Material online). In contrast, when assessing the distribution of effects of mutations on the fitness of an organism, the protein needs to be expressed at native levels, at the native locus and with native buffering mechanisms (fig. 3).

### Weak Negative Epistasis between Mutations
We analyzed the complete set of 510 mutants carrying 1 to 10 mutations and this analysis was done both including and excluding lethal and synonymous mutations. Without epistatic effects, an exponential decline in fitness should be expected when increasing the number of mutations. However, if there are epistatic interactions between the mutations, the data will fit to the equation $s = e^{-(am+bm^2)}$, where $m$ is the number of mutations, $a$ reflects the fraction of multiplicative deleterious mutations and $b$ is the epistasis parameter (Charlesworth 1990; Bershtein et al. 2006). For $b/a \approx 0$, the epistatic potential of the mutations is minimal whereas for $b/a \gg 0$ the mutations exhibit large negative epistatic effects (Bershtein et al. 2006). We tested exponential and epistatic curve fits, and observed that when including synonymous mutations, the $b/a \approx 0$ (0.04 and 0.02 when including and excluding synonymous mutations, respectively) and when excluding synonymous mutations, the ratio is $b/a > 0$ (0.40 and 0.14 when including and excluding synonymous mutations, respectively) (fig. 4; supplementary fig. S2 and table S3, Supplementary Material online). The epistasis parameter $b$ was positive for all studied cases (supplementary table S3, Supplementary Material online) suggesting that when assessing only nonsynonymous mutations, on an average there are negative epistatic effects between the mutations (Charlesworth 1990; Bershtein et al. 2006). This negative epistasis (on an average) has also been reported for other systems (Bershtein and Tawfik 2008; Perfeito et al. 2011; Jiang et al. 2013; Chou et al. 2014; Bank et al. 2016; Li et al. 2016; Sarkisyan et al. 2016).

### Fitness Cost per Mutation
The average fitness cost varied between 6.9% and 20.5% depending on which mutation types were included (table 1, fig. 4A–D; supplementary fig. S2 and table S3, Supplementary Material online). Regarding effects at the protein level, the
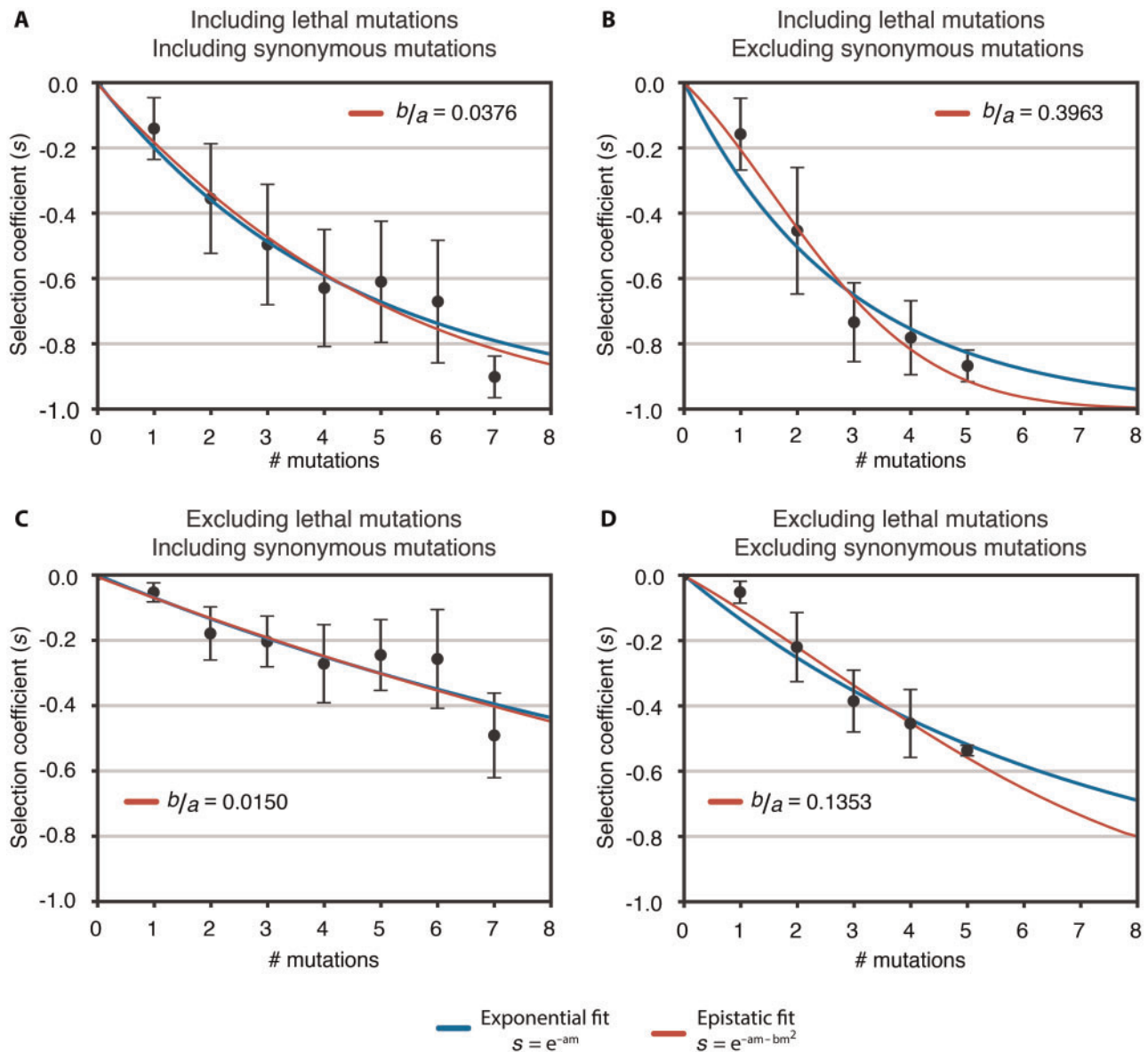
**Fig. 4.** Exponential (blue lines) and epistatic curve fits (red lines) of the average selection coefficient ($s$) with increasing number of mutations. Error bars represent the SD of all measured selection coefficients at each defined condition (A) Lethal mutations and synonymous mutations included in analysis. (B) Lethal mutations included in analysis and synonymous mutations excluded. (C) Lethal mutations excluded from analysis and synonymous mutations included. (D) Lethal mutations and synonymous mutations excluded from analysis. When synonymous mutations are included (A and C) the ratio $b/a \approx 0$ as compared with when synonymous mutations are excluded (B and D) where $b/a > 0$ suggesting negative epistatic interactions between nonsynonymous mutations leading to a higher combined cost compared with the cost mutations are expected to confer without epistasis effects.

average fitness costs for each random nonsynonymous mutation was 20.5% and already after accumulating approximately three nonsynonymous changes, fitness is on an average reduced to 34%, while at ten mutations, the fitness effect is expected to be lethal on an average. In proteins, the majority of mutations are assumed to reduce the protein stability, leading to decreased levels of functional protein (Green et al. 1992; Holder et al. 2001; DePristo et al. 2005; Tokuriki and Tawfik 2009a) and a lower fitness. However, since the $(\beta\alpha)_8$ barrel fold is considered to be a very stable scaffold (Höcker et al. 2001) the low mutational robustness of HisA is unexpected.

## No Restoration of Fitness by Chaperonin Overproduction

The GroEL/ES chaperonins have previously been reported to enable rescue of mutants (Gordon et al. 1994; Goyal and Chaudhuri 2015) and buffer the effect of deleterious mutations (Fares et al. 2002; Tokuriki and Tawfik 2009b). GroEL/ES is also known to be involved in the folding of many proteins with the $(\beta\alpha)_8$ barrel fold (Kerner et al. 2005; Fujiwara et al. 2010). Thus, we tested the rescuing capability of chaperonin overexpression on a subset of 26 HisA deleterious mutant strains. However, no rescue was observed for any of the mutant strains (supplementary fig. S3, Supplementary Material online, only
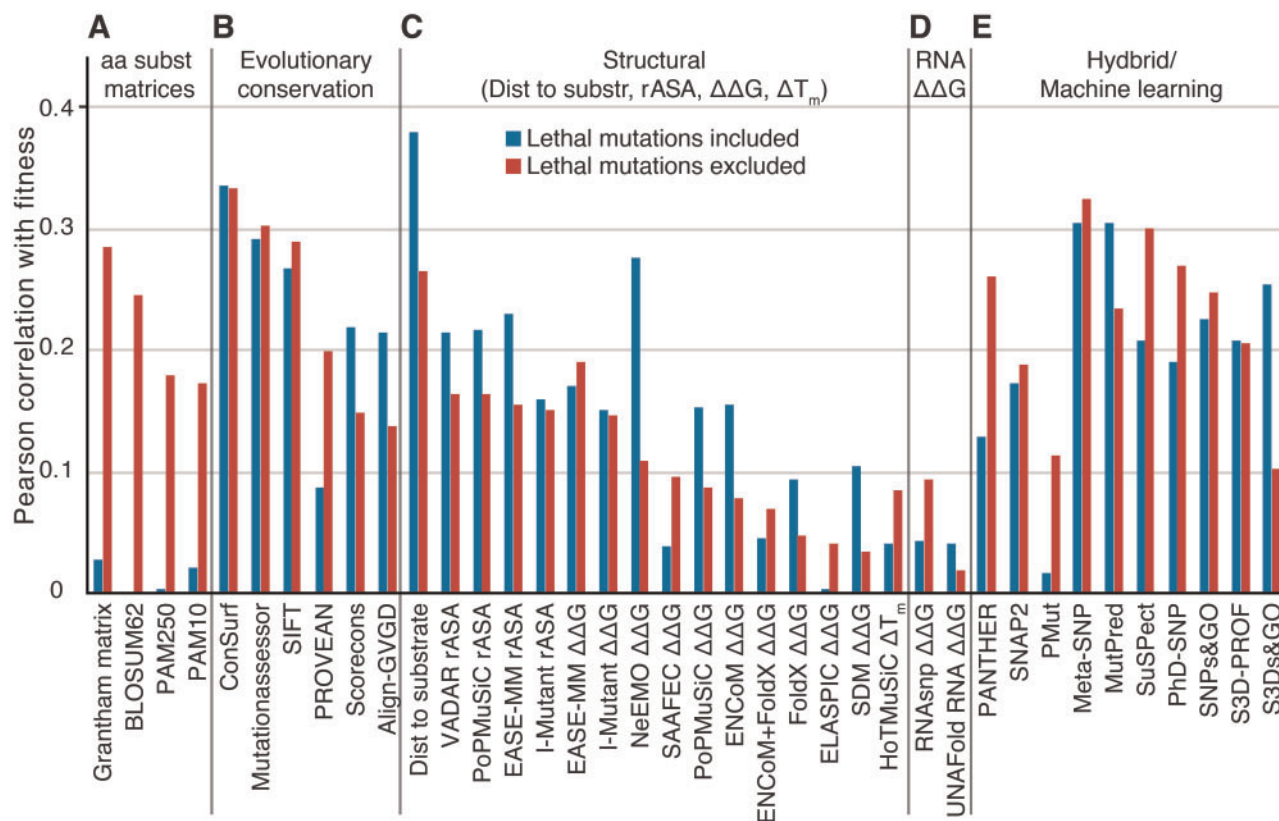
**Fig. 5.** Pearson correlation coefficients ($r$) for 38 different methods predicting the fitness effect of single amino acid mutations correlated to experimentally determined fitness at two different conditions all showed $r < 0.4$. Correlation coefficients were obtained from correlation predicted fitness from (A) amino acid (aa) substitution matrices, (B) evolutionary conservation scoring, (C) structural predictors: distance to substrate (Dist to substr), relative accessible surface area (rASA), change in folding free energy difference ($\Delta\Delta G$) and change in melting temperature ($\Delta T_m$), (D) change in RNA folding free energy difference (RNA $\Delta\Delta G$), and (E) hybrid/machine learning predictors.

8 mutants out of 26 are shown). Whether the lack of an effect is due to HisA protein not being a client for the GroEL/ES system or if there are other reasons remains unclear.

### Predicting Fitness Effects of Single Mutations

Accurate prediction of the effect of single mutations in an enzyme on organismal fitness is needed for understanding evolution and forecasting evolutionary pathways. The most direct way to assess the fitness of proteins would be to determine the specific activity (i.e., the kinetic properties of the mutant enzymes, $k_{cat}$ and $K_m$) as well as the in vivo level of enzyme and substrate, and then subsequently correlate these properties to organism fitness (Soskine and Tawfik 2010; Jiang et al. 2013; Boucher et al. 2016). However, such measurements are generally very labor intensive. Hence, researchers have tried to use various proxy measurements to estimate the effects of mutations (e.g., biophysical or phylogenetic characteristics of the studied protein). However, it is still unclear which factors and methods are most useful for predicting the effects of mutations on protein fitness. We applied 40 different prediction tools or measures to assess how well they predicted the effect of single mutations on protein fitness (summarized in fig. 5 and supplementary fig. S4, Supplementary Material online). The tools and measures span a wide range of properties such as changes in biochemical properties, evolutionary

distance, structural constraints, folding energy of the mRNA and machine learning tools for phenotypic prediction. Since we wanted to determine how well these methods predict effects on protein fitness, we used only fitness data obtained for the 81 single nonsynonymous mutants under low-level expression conditions where any mutational effect is directly detected as a change in organism fitness.

### Substitution Matrices Showed Varying Results

The Grantham, PAM, and BLOSUM amino acid substitution matrices (Grantham 1974; Dayhoff et al. 1978; Henikoff and Henikoff 1992) have previously been reported to correlate with fitness for some proteins, but not for others. Furthermore, though the correlations have been shown to be significant, they have been weak ($r = 0.05-0.5$) (Jacquier et al. 2013; Lind et al. 2016) and for HisA, only the Grantham matrix and the BLOSUM62 showed a significant but weak correlation when excluding the lethal mutations from the analyzed data set ($r = 0.21-0.28$, $P = 0.014-0.035$) (fig. 5A; supplementary fig. S4 and table S4, Supplementary Material online).

### Conservation Scoring

Phylogenetic conservation can be used as a measure of importance of each amino acid in a protein and the diversity

and conservation of amino acids in a certain position should reflect evolutionary constrains in that position (Mirny and Shakhnovich 1999; Franzosa and Xia 2009; Worth et al. 2009; Daudé et al. 2013; Gerek et al. 2013; Shahmoradi et al. 2014; Sikosek and Chan 2014; Yeh et al. 2014; Echave et al. 2016). The best prediction of fitness for HisA was accomplished with a conservation scoring method by ConSurf ($r = 0.33$) (fig. 5B; supplementary fig. S4 and table S4, Supplementary Material online), which was worse than the correlation reported previously for the enzyme AraD ($r = 0.49$), the AraCE proteins ($r_{AraC} = 0.58$, $r_{AraE} = 0.47$) (Lind et al. 2016), as well as the avGFP ($r = 0.40$) (Sarkisyan et al. 2016). However, the correlation was better than ribosomal proteins, in which no significant correlation was found (Lind et al. 2010).

### Structure-Based Methods
Biophysical properties (such as solvent accessibility, melting temperature, and thermodynamic stability) are predicted to have a big impact on protein evolution and fitness (Green et al. 1992; DePristo et al. 2005; Ramsey et al. 2011; Wylie and Shakhnovich 2011; Sikosek and Chan 2014; Boucher et al. 2016; Echave et al. 2016). Previous studies have found extensive support for the effect of mutations on thermodynamic stability ($\Delta\Delta G$) and rate of protein evolution (DePristo et al. 2005; Tokuriki et al. 2008; Tokuriki and Tawfik 2009c; Wylie and Shakhnovich 2011). However, it does not fully explain the variance observed in several studies, nor in HisA, with correlations varying between nonsignificant and $r = 0.4$ (Green et al. 1992; Holder et al. 2001; Lind et al. 2010, 2016; Jacquier et al. 2013) (fig. 5C; supplementary fig. S4 and table S4, Supplementary Material online). Neither did the difference in melting temperature ($\Delta T_m$), correlate with HisA fitness. The lack of linear correlation between fitness and $\Delta\Delta G$ values could be due to a stability threshold, where only mutations affecting the stability enough to reach a certain threshold value will have a negative effect on fitness (Jiang et al. 2013; Keren et al. 2016; Li et al. 2016; Sarkisyan et al. 2016). Even though the relative accessible surface area (rASA) have been reported to be a major determinant of evolutionary rate at single amino acid positions in proteins (Franzosa and Xia 2009), we observed only weak correlation with fitness in HisA ($r = 0.22$, $P < 0.05$; fig. 5C; supplementary fig. S4 and table S4, Supplementary Material online), similar to what has been reported for TEM-1 $\beta$-lactamase (Jacquier et al. 2013), AraCDE (Lind et al. 2016), and ribosomal proteins (Lind et al. 2010). One of the best (and simplest) predictors of mutational effects on organismal fitness when lethal mutations were included was the distance between the mutated amino acid and the HisA substrate in the crystallographic structure ($r = 0.38$, $P < 0.001$).

### RNA Stability Predictors
The stability of RNA can also affect the protein fitness. Mutations can interfere with the RNA structure and stability or disturb the translation of the protein, causing lower expression (Kudla et al. 2009; Lind et al. 2010; Goodman et al. 2013; Lind and Andersson 2013; Brandis et al. 2016). None of

the tested methods for determination of RNA folding $\Delta\Delta G$ correlated to fitness regardless of the data set considered ($r = 0.04$–$0.09$, $P > 0.4$) (fig. 5D; supplementary fig. S4 and table S4, Supplementary Material online).

### Machine Learning Methods and Disease Prediction Tools
To test the combination of conservation information and structural features as well as to include more advanced evolutionary deleteriousness scoring and take support from models built on previously known disease-causing mutations in proteins, we deployed a set of tools developed to predict protein fitness effects by machine learning (Thomas et al. 2003; Capriotti et al. 2006, 2013; Li et al. 2009; Yates et al. 2014; Hecht et al. 2015). However, none of the methods showed a strong correlation to observed organismal fitness, with Meta-SNP as the best performing predictor (fig. 5E; supplementary fig. S4 and table S4, Supplementary Material online), with a correlation of $r = 0.31$ ($P = 0.005$) and $r = 0.32$ ($P = 0.005$) when including and excluding lethal mutant strains, respectively. A correlation method based on the SNPs&GO 3D tool previously obtained a high correlation coefficient of $r = 0.71$ (Lind et al. 2016). For HisA however, we observed significant but weak correlation to organismal fitness when lethal mutations were included ($r = 0.25$, $P = 0.02$), but not when excluded ($r = 0.10$, $P = 0.38$) (fig. 5E; supplementary fig. S4 and table S4, Supplementary Material online). Combined, this suggested that the best methods to predict the effects of mutations are highly dependent on the specific protein analyzed.

### Effect of Mutations and Localization on the Structure
To further disentangle the effect of the mutations, we plotted the fitness effect of each mutation onto the crystal structure of the closed and substrate-bound structure of HisA (fig. 6). Many of the lethal mutations (purple) were in the catalytic face of the enzyme (6/7), suggesting that these mutations directly affect the binding of the substrate or the actual catalytic reaction. The average fitness for mutations in the catalytic face is $s = -0.37$ compared with average effect of $s = -0.18$ in the stability face (significant difference, Student's t-test $P < 0.01$) (fig. 6 and supplementary fig. S5, Supplementary Material online). In summary, several of the tested predictors gave weak but significant correlation where only six single predictors were significant for both data sets (with or without lethal mutations included). The best prediction methods were ConSurf, Mutationassessor, SIFT, Meta-SNP prediction, SNAP2, and the shortest distance from the mutated amino acid to the substrate.

### Multiple Linear Regression Models Explain More Variance
Clearly, no single method can predict the organismal fitness with high accuracy given a single protein mutation: none of the predictors used in this analysis could explain >11% of the observed variation (corresponds to simple linear regression $R^2 = 0.11$, Pearson correlation coefficient $r = 0.33$; ConSurf, supplementary table S5, Supplementary Material online). To test if the predictions can be improved by combining
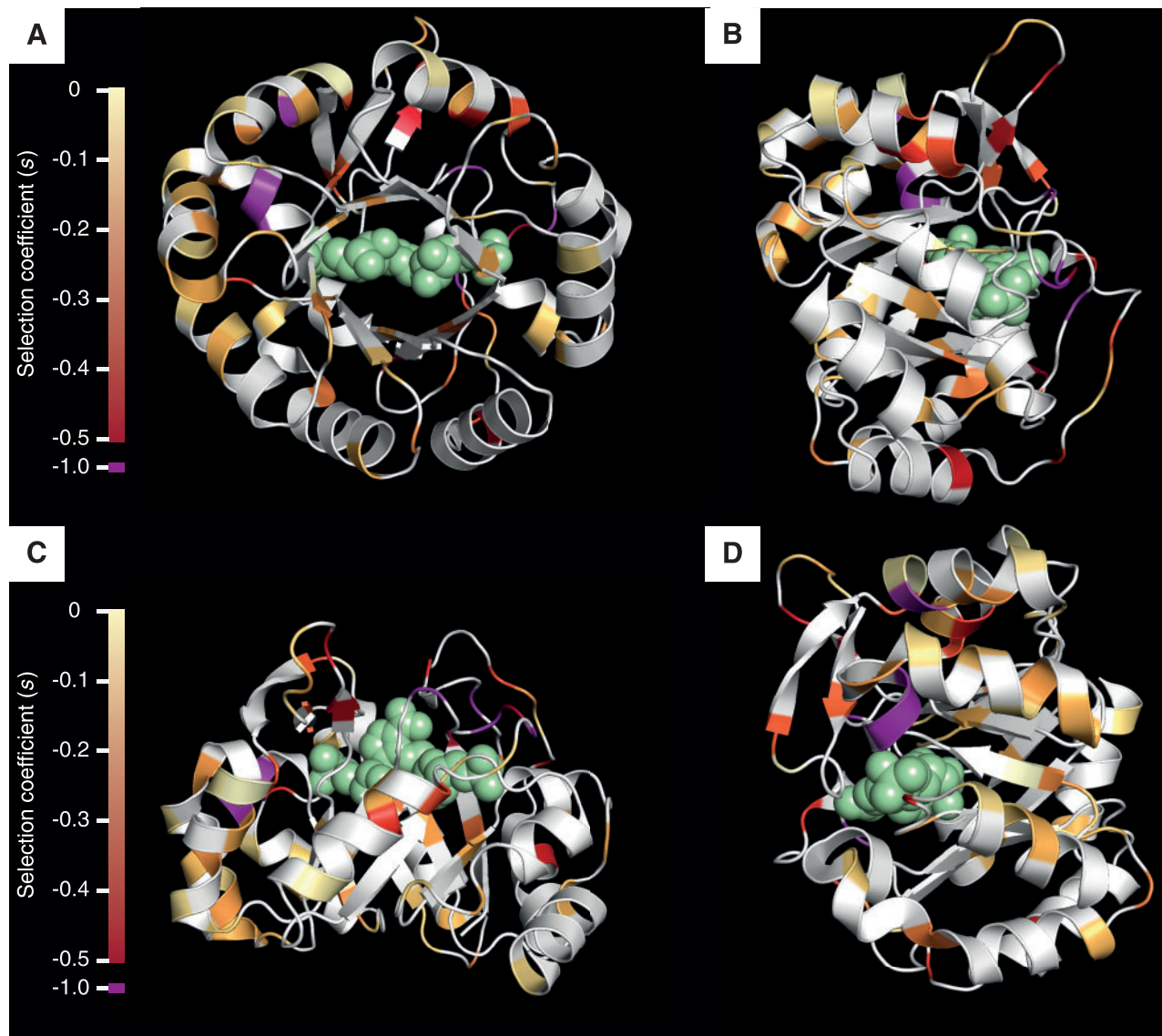
**Fig. 6.** Structure of HisA with ProFAR substrate in green. Amino acid residues are colored by the measured selection coefficient (*s*), ranging from pale yellow (neutral; *s* = 0) to dark red (strongly deleterious; *s* = −0.5), and purple (lethal; *s* = −1). The most deleterious amino acid changes are clustered close to the substrate and are mainly found in the catalytic face of the enzyme. HisA is shown from different viewing angles with views (*A*) facing the catalytic face, (*B*) side view with the catalytic site pointing rightward, (*C*) side view with the catalytic site pointing upward, and (*D*) side view with the catalytic site pointing leftward.

different methods, we fitted a multiple linear regression model with the selection coefficient as response variable and different subsets of the 40 predictors described above as dependent variables. The models with the lowest Mallow's $C_p$ value (Mallows 1973; Gilmour 1996) and highest adjusted $R^2$ (supplementary fig. S6, Supplementary Material online) were chosen for further testing. We found that a model combining 9 unique predictors (Grantham matrix, Mutationassessor, PROVEAN, EASE-MM rASA, PoPMuSiC ΔΔG, HoTMuSiC ΔTm, SNPs&GO, S3Ds&GO, and face location [catalytic or stability face]; supplementary fig. S6A and B, Supplementary Material online) could explain 47% of the observed variation when lethal mutations were excluded (henceforth referred to as the nonlethal model). A model built by including the lethal mutations

with 8 different predictors (EASE-MM ΔΔG, I-Mutant ΔΔG, NeEMO ΔΔG, PoPMuSiC ΔΔG, ENCoM ΔΔG, SDM ΔΔG, Secondary structure and face location [catalytic or stability face]; supplementary fig. S6C and *D*, Supplementary Material online; henceforth referred to as the lethal model) explained 50% of the observed variance. Thus, by combining different methods, a more powerful prediction of the fitness effects of a single mutation can be made.

*Multiple Linear Regression Models Are Able to Predict Selection Coefficients*
The resulting linear combination of predictors was then validated against the measured data. To build and assess our

models with independent data, we used a jack-knifing approach: the data set was randomly divided in half, a multiple linear regression model with the suggested predictors was fitted against one half of the data and used to predict the other half. The difference between the predicted and measured fitness was calculated (supplementary fig. S7, Supplementary Material online). The procedure was repeated 1000 times. The error of the prediction (calculated as the absolute difference between the value predicted on half the data and the experimentally measured value) when excluding lethal mutations has an average of $\Delta s = 0.09$ (supplementary fig. S7A, Supplementary Material online) and a SD of $\sigma^2 = 0.04$. When including lethal mutations, the error of the prediction deviates more frequently from zero and the variance of the selection coefficients is increased compared with when lethal mutations are excluded when fitting the model (supplementary fig. S7B, Supplementary Material online). The average error is $\Delta s = 0.17$ and the SD is $\sigma^2 = 0.07$. Further, in the nonlethal model, 50%, 80%, and 95% of the predictions have an average difference with the observed selection coefficient smaller or equal to 0.09, 0.12, and 0.16, respectively. In other words, 50% of the predictions were 0.09 or less from the observed selection coefficient and 80% were 0.12 or less from the observed value. In the lethal model, the predictions are not as accurate: 50%, 80%, and 95% of the predictions have an average difference smaller or equal to 0.17, 0.23, and 0.30, respectively.

## Conclusions

To forecast evolutionary trajectories, we need to understand how mutations influence the function of proteins, RNAs, and regulatory regions and how these changes in turn affect organism fitness. The present analysis of the fitness effects of mutations in the HisA biosynthetic enzyme contribute several new insights and illustrate the inadequacy of the available prediction tools.

First, our analysis shows that the average effect on fitness of single amino acid substitutions varies with expression levels. A reduction in organismal fitness will not be seen until the total protein activity (specific protein activity × protein concentration) drops below a threshold. As summarized in table 1, the average fitness costs for nonsynonymous mutations varied between 4.8% and 27.2% depending on the specific assay condition and method of calculation. These costs per amino acid substitution in HisA are similar to what has been observed previously for the AraC, D, and E proteins where the average cost was 12.3% (Lind et al. 2016). However, these costs are 10- to 100-fold higher than those observed for random genomic mutations where each mutation confers an average cost of ~0.15–1.5% (Kibota and Lynch 1996; Maisnier-Patin et al. 2005; Lind and Andersson 2008). This difference can be explained by the fact that in these genome-wide studies many of the random mutations occur in genes and chromosomal regions that are not rate-limiting for growth under the specific growth conditions used to measure organism fitness. In addition to this, any lethal or strongly deleterious mutations would not be detected as the analysis

requires growth. Interestingly, for the HisA protein, very few of the mutations (2/81 = 2.5%) appeared neutral with regard to protein fitness (even though they might appear neutral with regard to organism fitness if the protein is expressed at high level). This result is similar to what was observed for ribosomal proteins S20 and L1 (Lind et al. 2010) when using highly sensitive assays (6/126 = 4.8% neutral) but differs from the results obtained for the AraC, D, and E proteins where 33–56% of all amino acid substitution mutations appeared neutral (Lind et al. 2016). The fraction of neutral mutations is low in HisA (2.5%), which is essential under our assay conditions, and low in ribosomal proteins (4.8%), which can be deleted without lethal effect (Lind et al. 2010). Conversely, a large fraction of mutations are neutral in AraD, which is essential for growth with L-arabinose as carbon source (56%) and in AraC (33%), which is not essential (Lind et al. 2016). Since the studies of HisA, ribosomal proteins and AraCDE have been studied using similar assays with high sensitivity and under conditions where growth rate was limited by the functionality of these proteins, this suggests that the fraction of neutral mutations does not seem to be correlated to the relative importance of the protein in the conditions studied. It is also likely that this variation in the fraction of neutral mutations (from 2.5% to 56%) reflects actual biological differences in organismal robustness to mutations in these proteins rather than differences in the assay conditions and sensitivity. Why the robustness would differ up to 22-fold is unclear but, as proposed previously, one potential explanation might be differences in the strength of selection to reduce the costs of translational and transcriptional errors (Lind et al. 2016). Another potential explanation could be that different types of protein folds have intrinsic differences in their stability against amino acid substitutions. Furthermore, it is possible that the expression levels of these proteins differ, creating a difference in buffering capacity in organismal fitness for mutations that are deleterious to protein function (i.e., protein fitness) (Jiang et al. 2013; Keren et al. 2016; Li et al. 2016; Sarkisyan et al. 2016) (fig. 3).

Second, the shape of the fitness distribution is strongly dependent on the extent to which the studied protein is rate-limiting for growth. The overall activity of a protein is determined by its kinetic efficiency ($k_{cat}/K_m$) and the concentration of enzyme and substrate. If this activity is more than what is needed to sustain maximal growth, mutations that reduce either specific activity and/or concentration of the active enzyme will only influence organism fitness if the total activity is reduced below a certain threshold level. In line with this notion, with high expression levels, many of the mutations appeared neutral, whereas at low expression levels almost all mutations were deleterious. In other words, mutations deleterious to protein function can appear neutral for organism fitness if protein expression is sufficiently high. The buffering effect of the expression level also implies that mutations deleterious to the protein but not to the organism aid in protein evolution (fig. 3). Thus, a wider mutational landscape can be explored without the organism losing in fitness.

Third, the results show that as nonsynonymous mutations accumulated, fitness costs increased at an increasing rate, implying that random mutations interact such that their combined effect on fitness is amplified. Although weak, this negative epistasis is in line with what has been observed in other proteins (Bershtein and Tawfik 2008; Perfeito et al. 2011; Jiang et al. 2013; Chou et al. 2014; Bank et al. 2016; Li et al. 2016; Sarkisyan et al. 2016) but opposite to observations done at the genomic level in mutation accumulation experiments (Maisnier-Patin et al. 2005), even though the underlying mechanisms might be different when comparing HisA and other cases where negative epistasis was observed. For example, the results from Maisnier-Patin et al. (2005) suggest that random, genome-wide mutations interact such that their combined effect on fitness is strongly mitigated and that the genome is buffered against the fitness reduction caused by accumulated mutations. In this particular case, the proposed explanation for the buffering at the genome level is that in lineages that had accumulated many genome-wide mutations an upregulation of the heat shock chaperones DnaK and GroEL/ES occurs and buffer the deleterious effects (Maisnier-Patin et al. 2005). Similar stabilization effects of chaperones were also suggested to buffer the effects of highly deleterious mutations in different proteins (Bershtein et al. 2006; Tokuriki and Tawfik 2009c). Such a regulatory response could in principle also rescue the negative effects of mutation combinations in a single protein like HisA if the accumulated mutations destabilize the protein, and if there exists a suitable chaperone to refold the misfolded protein. The reduced protein fitness could hence be rescued and the organismal fitness remain unaffected (fig. 3), allowing for a greater mutational landscape to be explored without losing fitness. We observed, on an average, a weak negative epistatic response, but no rescue of the reduced fitness of the HisA mutants by GroEL/ES could be seen.

A fourth major conclusion is that all predictors for effects of mutations perform poorly if used individually, but a linear combination of predictors explains approximately half of the variance of the selection coefficient. Generally, mutations with large deleterious effect are most reliably predicted with several of the prediction tools whereas small effect mutations are more difficult to identify. It is notable that the simple measure of distance from the mutated amino acid to the bound substrate in the protein structure is one of the best predictors of effect on fitness. Combining different methods in a multiple linear regression model significantly improves the predictability, allowing 80% the predictions to fall within 12% of the experimentally observed value. It can still only explain half of the variation in fitness, indicating that we are still lacking fundamental knowledge on how to connect protein structure and function.

## Materials and Methods

### Strains and Media
All strains used in this study were derived from *Salmonella enterica* serovar Typhimurium strain LT2 (DA6192), and are listed in supplementary table S6, Supplementary Material

online, together with their corresponding s-values. Lysogeny broth (LB; 5 g yeast extract [Oxoid], 10 g Tryptone [Oxoid], 10 g NaCl [VWR], and 1 mM NaOH per liter) (Bertani 1951), and SOC (Hanahan 1983) were used as liquid media during the construction of the strains, with supplementation of 1.5% (w/v) agar (LA) as the solid medium when required. M9 minimal media (Miller 1992) was used for growth experiments, and for assessing viability or lethality of mutants. For selection for loss of *sacB*, sucrose selection plates (LA without sodium chloride, supplemented with 5% (w/v) sucrose) were used. When appropriate, media was supplemented with 25 μg/ml zeocin, 12.5 μg/ml chloramphenicol, 12.5 μg/ml (minimal media), or 50 μg/ml (rich media) kanamycin, 7.5 μg/ml tetracycline, 0.2% (w/v) glucose, 0.2% (v/v) glycerol, 0.1 mM L-tryptophan, or 0.1 mM L-histidine. For long-term storage, strains were grown overnight in LB, mixed with dimethyl sulfoxide (DMSO), at a final concentration of 10% (v/v), and frozen at −80 °C.

### Strain Constructions
Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific Inc.) was used for amplification of DNA cassettes for recombineering, except for when mutagenesis by error-prone PCR was performed, in which case GeneMorph II Random Mutagenesis Kit (Agilent Technologies) was used. For screening and generation of templates for Sanger sequencing, DreamTaq PCR Master Mix (Thermo Fisher Scientific Inc.) was used.

The construction methodology for the strains used is outlined in supplementary materials and methods, Supplementary Material online.

### Growth Measurements
Strains were recovered from −80 °C and streaked onto LA plates. Following overnight incubation at 37 °C, single colonies were picked and either restreaked on M9 minimal media agar or inoculated in M9 minimal media liquid cultures and incubated at 37 °C. All strains were tested on M9 minimal media agar supplemented with glucose and L-tryptophan. *hisA* mutants showing no colonies within 2 weeks were classified as lethal. Exponential growth rates were determined for all strains showing colonies on minimal media agar after 2 weeks. For growth rate experiments with constructs in *cobA*, four independent colonies of each construct were used to inoculate overnight cultures in M9 minimal media supplemented with L-tryptophan and were allowed to grow until no further change in optical density could be observed (maximum of 4 days). The cultures were diluted 1:1000 and 300 μl were added into two separate wells in a honeycomb plate. An isogenic strain with a wild-type copy of *hisA* was used as reference strain to enable calculations of relative growth rates. The cultures were allowed to grow in a Bioscreen C Analyzer at 37 °C with shaking for up to 4 days. The optical density at 600 nm wavelength ($OD_{600}$) was measured every 4 min. Exponential growth was observed in an $OD_{600}$ range of 0.02–0.055. Growth rates ($k$, $min^{-1}$) were obtained by fitting the curve $OD_{600} \propto N = N_0 e^{kt}$, where $t$ (min) is time, to the data. Relative growth rates were

calculated by dividing the growth rate of each replicate of the mutant by the average of the growth rate of the wild-type replicates (eight replicates each). Replicates with a growth rate much faster than the majority of replicates were assumed to have acquired media adaptation mutations and were removed from the analysis. If four or more replicates showed a growth rate much faster than the slowest replicates, the growth rate experiment was repeated. The measurement variation was calculated as the SD of all replicates included in the analysis. Student's $t$-test was used for testing assay resolution (i.e., maximum $|s|$ not significantly different from $s = 0$). For growth rate measurements of strains expressed from $P_{araBAD}$ promoter, overnight cultures were grown in M9 minimal media supplemented with glycerol and 0.05 mM histidine. Experiments showed that the histidine was depleted in the culture after overnight growth (data not shown). The cultures were diluted 1:1000 into M9 minimal media supplemented with glycerol and 0.1 mg/ml L-arabinose. 300 microliters of each culture were added into two separate wells in a honeycomb plate. An isogenic strain with a wild-type copy of $hisA$ was used as reference strain to enable calculations of relative growth rates. For test of GroEL/ES rescue of HisA mutants, pGro7-kan was transformed into a selection of 26 different strains with different amount of mutations and $hisA$ overexpressed from the $P_{tac}$ promoter in the $cobA$ locus showing reduced fitness. The pGro7-kan plasmid was also transformed into an isogenic strain with a wild-type copy of $hisA$ and the resulting strain was used as reference throughout the experiments. Overnight cultures were grown in M9 minimal media supplemented with glycerol, tryptophan, and kanamycin. The cultures were diluted 1:1000 into M9 minimal media supplemented with glycerol, tryptophan, and kanamycin and into separate cultures also supplemented with 0.5 mg/ml L-arabinose.

## Competition Experiments

To assess the fitness of mutants undistinguishable from wild-type in exponential growth rate experiments, we competed fluorescently marked mutant and wild-type strains against each other. Strains were taken up from the $-80\,°C$ freezer onto LA plates. Following overnight incubation at $37\,°C$, single colonies were picked and 1 ml M9 minimal media cultures supplemented with 0.1 mM histidine and 0.1 mg/ml L-arabinose were inoculated. Mutant and wild-type strains with different fluorescent markers were mixed 1:1 (1 µl each) in 1 ml M9 minimal media supplemented with 0.1 mg/ml L-arabinose and grown for 24 h at $37\,°C$. 1 µl was then transferred to 1 ml M9 minimal media supplemented with 0.1 mg/ml L-arabinose. At the same time, 50-fold dilutions were made in phosphate buffered saline (PBS) for flow cytometry. 20,000 cells were counted and the fraction of RFP and YFP positive cells was determined by flow cytometry (MACSQuant VYB, Miltenyi Biotec). After 24 h of growth at $37\,°C$ (approximately ten generations), cells were again prepared and fractions were counted by flow cytometry. The selection coefficients were determined using the regression model $s = [\ln(R(t)/R(0))]/[t]$ (Dykhuizen 1990) where $R$ is the ratio of mutant to wild-type and $t$ is number of generations. The fitness of each mutant

was measured on eight independent lineages (inoculated from independent colonies of mutant and wild-type) for both YFP and RFP markers. Four lineages had the mutant strain marked with RFP and the wild-type strain marked with YFP and in four lineages the fluorescent markers were swapped. The average fitness was calculated between the medians of the separate four + four replicates and is thereby also independent of any difference in cost between the fluorescent markers. The SD was calculated within the two sets of dye swaps as the SD between the two median values and was used as a measurement for variation. Student's $t$-test was used to determine if measured fitness was significantly different from $s = 0$ (i.e., neutral).

## Structure Analysis

For all mutation effect predictors that required a crystal structure to be used, the structure of *Salmonella enterica* HisA(D7N, D176A) with ProFAR (PDB ID: 5A5W (Söderholm et al. 2015)) was used. The flexible loops 1–8 were assigned as the catalytic face of the enzyme and the rest of the enzyme was assigned as the stability face. This structure was also used for visualization of fitness effects of mutants (fig. 6). The distance between the mutated amino acid and the HisA substrate ProFAR was defined as the shortest distance in the structure of HisA between any atom in the mutated amino acid to any atom in the cocrystallized substrate.

## Mutational Spectrum

There was no significant bias for the distribution of the mutations over the gene (Kolmogorov–Smirnov test $P = 0.40$). The mutational spectrum is provided in supplementary table S1, Supplementary Material online.

## Kernel Density Estimation

A linear kernel density function was used to illustrate the sharp peak at $s = -1$ and the wider fitness distribution over $s = 0$. The bandwidth was empirically set to 0.15 after analyzing several distributions.

## Fitness Predictions

Forty different tools and measures were used to generate predicted the fitness of a certain mutation. All single amino acid mutants in this work were manually entered into the webpage-based user interface of each tool and an effect value was acquired (except for the structural analysis of the shortest distance to ligand measure and face assignment). The correlation of the predicted values (i.e., predicted fitness effect) to the growth rate data (i.e., actual fitness effect) was assessed by calculating the Pearson correlation coefficient (Pearson's $r$) and served as a measurement of the ability of each method to predict the fitness effect. The methods used for prediction calculations are given in supplementary materials and methods, Supplementary Material online.

## Multiple Linear Regression Model

To attempt to better predict the effect of single mutations on selection coefficient, multiple linear regression models were

calculated. The full description of the experiment is described in the supplementary discussion, Supplementary Material online. Briefly, all possible regression models including all possible combinations of fitness predictions were assessed and the ones providing an optimal trade-off between sensitivity and number of predictors were selected. To test the predictive power of these selected models, half the data set (i.e., half the mutations) were randomly selected and the difference between the observed selection coefficient and the one predicted by the model was recorded. That procedure was repeated 1000 times, and the prediction power was estimated by averaging the absolute differences between the observed and the predicted values.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgment

## References

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.

Bank C, Matuszewski S, Hietpas RT, Jensen JD. 2016. On the (un)predictability of a large intragenic fitness landscape. *Proc Natl Acad Sci U S A.* 113:14085–14090.

Bataillon T, Bailey SF. 2014. Effects of new mutations on fitness: insights from models and data. *Ann N Y Acad Sci.* 1320:76–92.

Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. 2006. Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444(7121):929–932.

Bershtein S, Tawfik DS. 2008. Advances in laboratory evolution of enzymes. *Curr Opin Chem Biol.* 12(2):151–158.

Bertani G. 1951. Studies on lysogenesis I.: the mode of phage liberation by lysogenic *Escherichia coli. J Bacteriol.* 62(3):293–300.

Boucher JI, Bolon DNA, Tawfik DS. 2016. Quantifying and understanding the fitness effects of protein mutations: laboratory versus nature. *Protein Sci.* 25(7):1219–1226.

Brandis G, Bergman JM, Hughes D. 2016. Autoregulation of the *tufB* operon in Salmonella. *Mol Microbiol.* 100(6):1004–1016.

Capriotti E, Calabrese R, Casadio R. 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* 22(22):2729–2734.

Capriotti E, Calabrese R, Fariselli P, Martelli P, Altman RB, Casadio R. 2013. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics* 14(Suppl 3):S6.

Charlesworth B. 1990. Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet Res.* 55(3):199–221.

Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet.* 10(3):195–205.

Charlesworth J, Eyre-Walker A. 2006. The rate of adaptive evolution in enteric bacteria. *Mol Biol Evol.* 23(7):1348–1356.

Chou H-H, Delaney NF, Draghi JA, Marx CJ. 2014. Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLoS Genet.* 10(2):e1004149.

Daudé D, Topham CM, Remaud-Siméon M, André I. 2013. Probing impact of active site residue mutations on stability and activity of Neisseria polysaccharea amylosucrase. *Protein Sci.* 22:1754–1765.

Dayhoff MO, Schwartz RM, Orcutt BC. 1978. A model of evolutionary change in proteins. In: Margaret O. Dayhoff, editor. Atlas of Protein Sequence and Structure 5. Vol. 40. National Biomedical Research Foundation. p. 345–352.

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

DePristo MA, Weinreich DM, Hartl DL. 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet.* 6(9):678–687.

Dykhuizen DE. 1990. Experimental Studies of Natural Selection in Bacteria. *Annu Rev Ecol Syst.* 21(1):373–398.

Echave J, Spielman SJ, Wilke CO. 2016. Causes of evolutionary rate variation among protein sites. *Nat Rev Genet.* 17(2):109–121.

Elena SF, Ekunwe L, Hajela N, Oden SA, Lenski RE. 1998. Distribution of fitness effects caused by random insertion mutations in *Escherichia coli. Genetica* 102:349–358.

Eyre-Walker A. 2002. Changing effective population size and the McDonald-Kreitman test. *Genetics* 162(4):2017–2024.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet.* 8(8):610–618.

Eyre-Walker A, Woolfit M, Phelps T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173:891–900.

Fares MA, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E. 2002. Endosymbiotic bacteria GroEL buffers against deleterious mutations. *Nature* 417(6887):398–398.

Firnberg E, Labonte JW, Gray JJ, Ostermeier M. 2014. A comprehensive, high-resolution map of a gene's fitness landscape. *Mol Biol Evol.* 31(6):1581–1592.

Fix E, Hodges JL. 1989. Discriminatory analysis. Nonparametric discrimination: consistency properties. *Int Stat Rev.* 57(3):238.

Franzosa EA, Xia Y. 2009. Structural determinants of protein evolution are context-sensitive at the residue level. *Mol Biol Evol.* 26(10):2387–2395.

Fujiwara K, Nakahigashi K, Ishihama Y, Soga T, Taguchi H. 2010. A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO J.* 29(9):1552–1564.

Gerek ZN, Kumar S, Ozkan SB. 2013. Structural dynamics flexibility informs function and evolution at a proteome scale. *Evol Appl.* 6:423–433.

Gilmour SG. 1996. The interpretation of Mallows's $C_p$-statistic. *Statistician* 45(1):49–56.

Goodman DB, Church GM, Kosuri S. 2013. Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342(6157):475–479.

Gordon CL, Sather SK, Casjens S, King J. 1994. Selective in vivo rescue by GroEL/ES of thermolabile folding intermediates to phage P22 structural proteins. *J Biol Chem.* 269(45):27941–27951.

Goyal M, Chaudhuri TK. 2015. GroEL–GroES assisted folding of multiple recombinant proteins simultaneously over-expressed in *Escherichia coli. Int J Biochem Cell Biol.* 64:277–286.

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* 185:862–864.

Green SM, Meeker AK, Shortle D. 1992. Contributions of the polar, uncharged amino acids to the stability of staphylococcal nuclease: evidence for mutational effects on the free energy of the denatured state. *Biochemistry* 31(25):5717–5728.

Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol.* 166(4):557–580.

Hecht M, Bromberg Y, Rost B. 2015. Better prediction of functional effects for sequence variants. *BMC Genomics* 16(Suppl 8):1–12.

Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A.* 89(22):10915–10919.

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

Höcker B, Jürgens C, Wilmanns M, Sterner R. 2001. Stability, catalytic versatility and evolution of the ($\beta\alpha$) 8-barrel fold. *Curr Opin Biotechnol.* 12:376–381.

Holder JB, Bennett AF, Chen J, Spencer DS, Byrne MP, Stites WE. 2001. Energetics of side chain packing in staphylococcal nuclease assessed by exchange of valines, isoleucines, and leucines†. *Biochemistry* 40(46):13998.

Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit E, Poulain J, Barnaud G, et al. 2013. Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc Natl Acad Sci U S A.* 110(32):13067–13072.

Jiang L, Mishra P, Hietpas RT, Zeldovich KB, Bolon DNA, Serio TR. 2013. Latent effects of Hsp90 mutants revealed at reduced expression levels. *PLoS Genet.* 9(6):e1003600.

Kassen R, Bataillon T. 2006. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat Genet.* 38(4):484–488.

Keightley PD, Eyre-Walker A. 2010. What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Philos Trans R Soc Lond B Biol Sci.* 365(1544):1187–1193.

Keren L, Hausser J, Lotan-Pompan M, Slutskin IV, Alisar H, Kaminski S, Weinberger A, Alon U, Milo R, Segal E. 2016. Massively parallel interrogation of the effects of gene expression levels on fitness. *Cell* 166(5):1282–1294.e18.

Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang H-C, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, Hartl FU. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli. Cell* 122(2):209–220.

Kibota TT, Lynch M. 1996. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli. Nature* 381(6584):694–696.

Knight GM, Colijn C, Shrestha S, Fofana M, Cobelens F, White RG, Dowdy DW, Cohen T. 2015. The distribution of fitness costs of resistance-conferring mutations is a key determinant for the future burden of drug-resistant tuberculosis: a model-based analysis. *Clin Infect Dis.* 61(Suppl 3):S147–S154.

Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in *Escherichia coli. Science* 324: 255–258.

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 25(21):2744–2750.

Li C, Qian W, Maclean CJ, Zhang J. 2016. The fitness landscape of a tRNA gene. *Science* 352:837–840.

Lind PA, Andersson DI. 2008. Whole-genome mutational biases in bacteria. *Proc Natl Acad Sci U S A.* 105(46):17878–17883.

Lind PA, Andersson DI. 2013. Fitness costs of synonymous mutations in the rpsT gene can be compensated by restoring mRNA base pairing. *PLoS One* 8(5):e63373.

Lind PA, Arvidsson L, Berg OG, Andersson DI. 2016. Variation in mutational robustness between different proteins and the predictability of fitness effects. *Mol Biol Evol.* 34:408–418.

Lind PA, Berg OG, Andersson DI. 2010. Mutational robustness of ribosomal protein genes. *Science* 330(6005):825–827.

Maisnier-Patin S, Roth JR, Fredriksson A, Nyström T, Berg OG, Andersson DI. 2005. Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nat Genet.* 37(12):1376–1379.

Mallows CL. 1973. Some comments on C P. *Technometrics* 15(4):661–675.

McDonald MJ, Cooper TF, Beaumont HJE, Rainey PB. 2011. The distribution of fitness effects of new beneficial mutations in *Pseudomonas fluorescens. Biol Lett.* 7(1):98–100.

Miller JH. 1992. A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria. Plainview, N.Y: Cold Spring Harbor Laboratory Press.

Mirny LA, Shakhnovich EI. 1999. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J. Mol. Biol.* 291:177–196.

Perfeito L, Ghozzi S, Berg J, Schnetz K, Lässig M. 2011. Nonlinear fitness landscape of a molecular pathway. *PLoS Genet.* 7(7): e1002160.

Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuan R. 2010. Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. *Genetics* 185:603–609.

Ramsey DC, Scherrer MP, Zhou T, Wilke CO. 2011. The relationship between relative solvent accessibility and evolutionary rate in protein evolution. *Genetics* 188(2):479–488.

Sanjuan R, Moya A, Elena SF. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U.S.A.* 101:8396–8401.

Sanjuan R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 365:1975–1982.

Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, et al. 2016. Local fitness landscape of the green fluorescent protein. *Nature* 533(7603):397–401.

Shahmoradi A, Sydykova DK, Spielman SJ, Jackson EL, Dawson ET, Meyer AG, Wilke CO. 2014. Predicting Evolutionary Site Variability from Structure in Viral Proteins: Buriedness, Packing, Flexibility, and Design. *J Mol Evol.* 79:130–142.

Sikosek T, Chan HS. 2014. Biophysics of protein evolution and evolutionary protein biophysics. *J R Soc Interface* 11(100):1–35.

Silander OK, Tenaillon O, Chao L, Barton NH. 2007. Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biol.* 5(4):0922–0931.

Silverman BW, Jones MC. 1989. E. Fix and J.L. Hodges (1951): an important contribution to nonparametric discriminant analysis and density estimation: commentary on Fix and Hodges (1951). *Int Stat Rev.* 57:233–247.

Silverman BW. 1986. Density estimation for statistics and data analysis. London: Chapman and Hall.

Söderholm A, Guo X, Newton MS, Evans GB, Näsvall J, Patrick WM, Selmer M. 2015. Two-step ligand binding in a ($\beta\alpha$)$_8$ Barrel enzyme: substrate-bound structures shed new light on the catalytic cycle OF HisA. *J Biol Chem.* 290(41):24657–24668.

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

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome* 13(9):2129–2141.

Tokuriki N, Stricher F, Serrano L, Tawfik DS. 2008. How protein stability and new functions trade off. *PLoS Comput Biol.* 4(2):e1000002.

Tokuriki N, Tawfik DS. 2009a. Protein dynamism and evolvability. *Science* 324(5924):203–207.

Tokuriki N, Tawfik DS. 2009b. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* 459:668–673.

Tokuriki N, Tawfik DS. 2009c. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol.* 19(5):596–604.

Walkiewicz K, Benitez Cardenas AS, Sun C, Bacorn C, Saxer G, Shamoo Y. 2012. Small changes in enzyme function can lead to surprisingly large fitness effects during adaptive evolution of antibiotic resistance. *Proc Natl Acad Sci U S A.* 109(52):21408–21413.

Worth CL, Gong S, Blundell TL. 2009. Structural and functional constraints in the evolution of protein families. *Nat Rev Mol Cell Biol.* 10:709–720.

Wloch DM, Szafraniec K, Borts RH, Korona R. 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae. Genetics* 159:441–452.

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

Yates CM, Filippis I, Kelley LA, Sternberg MJE. 2014. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J Mol Biol.* 426(14):2692–2701.

Yeh S-W, Huang T-T, Liu J-W, Yu S-H, Shih C-H, Hwang J-K, Echave J. 2014. Local Packing Density Is the Main Structural Determinant of the Rate of Protein Sequence Evolution at Site Level. *Biomed Res Int.* 2014:1–10.

Yue P, Yue P, Li Z, Li Z, Moult J, Moult J. 2005. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol* 353(2):459–473.

Zhang L, Li W-H. 2005. Human SNPs reveal no evidence of frequent positive selection. *Mol Biol Evol.* 22(12):2504–2507.