A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape

Elad Firnberg,[1] Jason W. Labonte,[1] Jeffrey J. Gray,[1] and Marc Ostermeier*,[1]
[1]Department of Chemical and Biomolecular Engineering, Johns Hopkins University
*Corresponding author: E-mail: oster@jhu.edu.
Associate editor: Howard Ochman

Abstract

Mutations are central to evolution, providing the genetic variation upon which selection acts. A mutation's effect on the suitability of a gene to perform a particular function (gene fitness) can be positive, negative, or neutral. Knowledge of the distribution of fitness effects (DFE) of mutations is fundamental for understanding evolutionary dynamics, molecular-level genetic variation, complex genetic disease, the accumulation of deleterious mutations, and the molecular clock. We present comprehensive DFEs for point and codon mutants of the *Escherichia coli* TEM-1 β-lactamase gene and missense mutations in the TEM-1 protein. These DFEs provide insight into the inherent benefits of the genetic code's architecture, support for the hypothesis that mRNA stability dictates codon usage at the beginning of genes, an extensive framework for understanding protein mutational tolerance, and evidence that mutational effects on protein thermodynamic stability shape the DFE. Contrary to prevailing expectations, we find that deleterious effects of mutation primarily arise from a decrease in specific protein activity and not cellular protein levels.

Key words: protein evolution, fitness landscape, beta-lactamase.

Introduction

The fitness landscape model for protein evolution, as first conceptualized by Smith in 1970 (Smith 1970) and generalized by others (Orr 2005), imagines evolution as a process by which a sequence moves by stochastic processes from its wild-type sequence through fitter and fitter sequences until the sequence reaches a local fitness optimum. The nature of the fitness landscape determines the dynamics of evolution and fundamentally shapes what is and is not possible in evolution. Much has been learned from theoretical studies and small-scale interrogations of real fitness landscapes (Eyre-Walker and Keightley 2007). However, we still lack a systematic, assumption-free, experimental determination of the distribution of fitness effects (DFE) for all mutations of a gene performing its native function in its native host. The situation is akin to having a small set of aerial photographs of a geographical area versus having comprehensive satellite coverage such as provided by Google Earth.

We sought to provide a comprehensive, quantitative description of a fitness landscape corresponding to a gene and its nearest neighbors in both DNA and protein sequence space (i.e., the set of all sequences that differ by a single bp, codon, or amino acid) but avoid or ameliorate current limitations of large-scale measurements of fitness. Growth competition experiments or experiments in which alleles are enriched based on a threshold for function are the current state of the art (Fowler et al. 2010; Araya et al. 2012; Deng et al. 2012; McLaughlin et al. 2012; Schlinkmann et al. 2012; Whitehead et al. 2012; Roscoe et al. 2013; Starita et al. 2013). To varying degrees such experiments offer a direct "head-to-head" comparison of alleles but suffer four

significant limitations. First, most studies utilize nonnative reporter assays (e.g., phage display, cell surface display, and two-hybrid systems) in which the gene or gene fragment is removed from its native context and host and fused to another gene (but see Deng et al. 2012; Roscoe et al. 2013). Second, population size can affect the measured value of fitness due to stochastic effects. Third, these experiments have limited ability to measure fitness for low fitness alleles because such alleles are depleted during the course of the experiment. For example, Roscoe et al. (2013) were unable to reproducibly measure the fitness of ubiquitin point mutants with a fitness below approximately 40% of wild-type due to their rapid depletion in growth competition experiments. Thus, although such experiments tell us the location of valleys in the landscape, they cannot tell us anything about what the valleys look like. Fourth, the fitness measurements are subject to the extent and underlying form of genotype-by-environment interactions. For example, the fitness of an antibiotic resistance gene measured by a growth competition experiment will be a function of the arbitrary selective pressure used in the experiment (the antibiotic concentration). Alleles conferring resistance far above or below the level necessary for growth at one antibiotic concentration may show no fitness difference in that environment yet show significant differences at a different antibiotic concentration closer to their resistance limit. We desired to decouple fitness from genotype-by-environment interactions as much as possible to quantify the underlying landscape and thus better understand a gene's intrinsic evolutionary potential and limitations.

A key determinant of the fitness landscape is what we are defining here as "gene fitness." "Fitness" in the traditional

Article

Open Access

biological definition refers to the extent to which an organism is adapted to or able to produce offspring in a particular environment (often measured as growth rate in the fitness landscape of bacteria). Instead, we are referring to "gene fitness" as a phenotypic signature—the suitability of a gene to provide a particular function. Thus, a gene fitness landscape might also be referred to as a phenotypic landscape. Although to a large extent this landscape reflects a protein functional landscape, we use the term "gene fitness" because our landscape encompasses mutational effects at the DNA, RNA and protein level and is a holistic metric of the ability of an allele to provide a particular function. However, when we average the gene fitness values of synonymous alleles to examine the effect of missense mutations, we refer to this as a protein fitness landscape. "Protein fitness" is similarly defined as the suitability of a protein to provide a particular function. Unless otherwise specified, use of the word fitness in relation to a specific allele refers to "gene fitness," whereas its use in relation to a specific protein refers to "protein fitness." These landscapes define the relationship between DNA/protein sequence and biological function. The more that function is related to growth rate in particular environment, the more direct the relationship between gene fitness and organismal fitness. For example, in the case of antibiotic resistance genes, gene/protein fitness (as measured by minimum inhibitory concentration [MIC] of the antibiotic), and organismal fitness (as measured by growth rate) correlate to some extent (Deris et al. 2013; Jacquier et al. 2013), provided the antibiotic concentration is at the appropriate level for comparing the alleles. Here, we present a comprehensive gene fitness landscape for the TEM-1 β-lactamase gene and a protein fitness landscape for the TEM-1 protein.

Results and Discussion

Fitness Landscape of TEM-1 β-Lactamase

We chose to measure the DFE of the TEM-1 β-lactamase gene, a convenient model for the study of evolution and the fitness effects of mutations (Salverda et al. 2010; Soskine and Tawfik 2010). TEM-1 is native to Escherichia coli as a plasmid-borne gene (Medeiros 1984), and we examined TEM-1 in this context. TEM-1 confers high resistance to penicillin antibiotics such as ampicillin (Amp). Thus, when E. coli cells bearing TEM-1 are challenged to grow in the presence of Amp, alleles conferring an enhanced ability to degrade the antibiotic will enrich. Thus, Amp resistance is a key determinant of organismal fitness in the presence of Amp (Bershtein et al. 2006; Weinreich et al. 2006; Jacquier et al. 2013), although assessing TEM-1 fitness by measuring Amp resistance does not capture organismal fitness differences not associated with antibiotic resistance. Previous partial characterizations of the DFE of TEM-1 (Bershtein et al. 2006; Deng et al. 2012; Jacquier et al. 2013) suffer several significant limitations. These studies did not characterize the relationship between sequence and fitness (Bershtein et al. 2006), used error-prone PCR to generate mutations that were heavily biased to A/T to C/G transitions (80%) (Bershtein et al. 2006; Jacquier et al. 2013), and/or focused on the characterization of high fitness alleles with more

than one mutation and assumed additivity for predicting the effect of the individual mutations (Deng et al. 2012). In addition, fitness was either measured using either growth competition experiments (Deng et al. 2012), which suffer from limitations described in the Introduction, or in the coarse-grained manner of a MIC assay (Bershtein et al. 2006; Jacquier et al. 2013). MIC assays suffer the drawbacks of being low-throughput and low-resolution. Alleles with known mutations must be isolated and tested individually, and MICs are measured in discrete values (typically 2-fold increments). For example, the resolution of the MIC assay in a study of the amoxicillin resistance effects of 18% of the possible amino acid substitutions in TEM-1 was insufficient to capture the effects of synonymous mutations or to identify any beneficial mutations (Jacquier et al. 2013), both of which we readily achieve.

Here, we describe a synthetic biology approach to quantify fitness of TEM-1 in a single experiment that avoids or ameliorates the limitations of growth competition experiments and MIC assays and allows a comprehensive analysis of the DFE. A synthetic biology approach is by definition artificial in at least some aspects, but unlike several previous studies we measure the DFE of the gene in its native host and do not employ gene fusions or artificial reporters of fitness. Additionally, because TEM-1 increases the Amp resistance of E. coli cells over 1,000-fold, the combination of TEM-1 and Amp afforded the opportunity to determine the DFE over a wide range of fitness values. Our approach decouples genotype-by-environment interactions as far as Amp resistance is concerned. We quantify TEM-1's underlying fitness landscape and thus its intrinsic evolutionary potential and limitations. This gene fitness landscape is a very significant determinant in the organismal fitness landscape for growth of the bacteria in the presence of Amp (Jacquier et al. 2013). However, the two types of landscapes are not equivalent.

We determined the DFE for 98.2% (2,536/2,583) of all point mutations (i.e., all 1-bp changes) and 83.9% (15,167/18,081) of all codon substitutions in the TEM-1 gene (fig. 1). The latter includes all 1-, 2-, and 3-bp changes of the 287 codons of TEM-1. We also determined the DFE for 95.6% (5,212/5,453) of the possible single amino acid substitutions in the corresponding TEM-1 protein (fig. 2). We excluded insertions and deletions (indels) from our analysis of the DFE, an important class of mutations that generally have more deleterious effects on fitness (Toth-Petroczy and Tawfik 2013). The source of TEM-1 variants was a previously described library (CCM-2) designed to contain all possible single codon substitutions in the TEM-1 gene (i.e., each codon position in the gene could be changed to any of the other 63 codons but each allele had only one position changed) (Firnberg and Ostermeier 2012). To measure gene fitness, we first partitioned the CCM-2 library into 13 partially overlapping sublibraries based on relative Amp resistance using a synthetic gene circuit that functions as a tunable bandpass genetic selection for Amp resistance (Sohka et al. 2009) (supplementary fig. S1, Supplementary Material online). Next, we performed deep sequencing on each of the sublibraries, counting how many
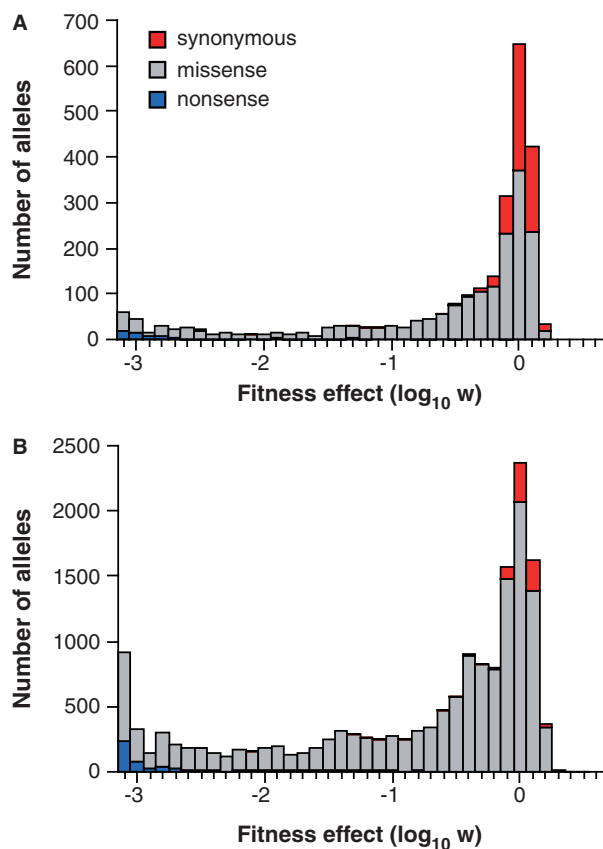
Fig. 1. Distribution of gene fitness effects (DFE) of mutations in *TEM-1*. (A) The DFE of point mutations (i.e., 1-bp changes in the gene). (B) The DFE of all possible codon substitutions (i.e., all 1, 2, and 3 base changes in the 287 codons of *TEM-1*). Gene fitness values for conferring ampicillin resistance are presented on a log scale with 0 corresponding to the fitness of *TEM-1*. The contributions of synonymous (red), missense (gray), and nonsense (blue) mutations to the DFE are indicated. Gene fitness as a function of codon substitution is provided as supplementary data S1 (Supplementary Material online).

times each allele appeared in each sublibrary and used these statistics to quantify each allele's conferred antibiotic resistance or gene fitness (*w*) relative to *TEM-1* (supplementary fig. S2, Supplementary Material online). We used the fitness effects of synonymous mutations to determine an upper limit on the error of our fitness measurements (supplementary fig. S3 and Materials and Methods, Supplementary Material online). Our method enables the accurate fitness quantification of any allele and avoids population size effects because the alleles are isolated on a plate. Unlike growth competition experiments, the probability of observing an allele in our experiment is predominantly independent of its conferred fitness. Additionally, the method decouples fitness from genotype-by-environment interactions, at least as far as the major environmental factor affecting fitness is concerned (i.e., the antibiotic concentration). We individually sequenced the alleles from 27 randomly selected colonies from two of the sublibraries and found these alleles' gene fitness values were in the expected range (supplementary fig. S4, Supplementary Material online).

TEM-1's DFE (fig. 1) indicated the gene was fairly robust to mutations as nearly one-half (47.3%) of all alleles retained at least 50% of the fitness of *TEM-1*. Among alleles with point mutations, 63.8% maintained at least 50% of the fitness of *TEM-1* (53.2% of the nonsynonymous and 97.2% of the synonymous point mutations), less than a previous estimation of 75% (Soskine and Tawfik 2010). Still, a sizable fraction of the mutants lost more than 90% of their fitness (19.6% of the point mutations and 30.3% of all codon substitutions), roughly in line with previous estimates of the frequency of mutations having a severe deleterious effect (Camps et al. 2007). Among point mutants, only 6% of the alleles completely lost the ability to provide any Amp resistance and 33% of those were nonsense mutations (fig. 1A), which is similar to the approximately 8% inactivating mutations found in a previous study of an error-prone PCR library of *TEM-1* (Soskine and Tawfik 2010). Only 7.1% (1,074/15,167) of the alleles and 7.0% (367/5,212) of the missense mutations increased fitness above that of *TEM-1* outside the range of the error. The bimodal distribution was qualitatively similar to the DFE of randomly chosen point mutations in DNA and RNA viruses (Sanjuan et al. 2004; Peris et al. 2010), the DFE of induced mutations in yeast (Wloch et al. 2001), the DFE of missense mutations of ubiquitin (Roscoe et al. 2013), a sampling of *TEM-1*'s DFE for amoxicillin resistance (Jacquier et al. 2013), and estimations of *TEM-1*'s DFE for Amp resistance (Soskine and Tawfik 2010).

Benefits of the Genetic Code's Architecture

TEM-1's DFE provides evidence that the standard genetic code's architecture minimizes the deleterious effects of mutations and enriches for adaptive mutations. The adaptive theory on the origin of the genetic code states that the genetic code is arranged to minimize the deleterious effects of mutations and mistranslations (Sonneborn 1965; Woese 1965). This theory predicts that point mutations would be less deleterious than 2- or 3-bp substitutions. We have recently shown this prediction held true for mutations in two small genes (*HB36* and *HB80*; Whitehead et al. 2012) that were reengineered for a new function in a nonnative organism (Firnberg and Ostermeier 2013). Here, we find that this prediction is also true of a wild-type gene in its native host. The median changes in relative gene fitness for 1-, 2-, and 3-bp substitutions at a codon position were −0.36, −0.52, and −0.63, respectively. More significantly, the frequency of point mutations among the alleles with a fitness less than 0.1 was 35.3% less than that expected if the point mutations were evenly distributed across all fitness values ($P = 1.1 \times 10^{-40}$ based on comparison with a hypergeometric distribution). For *HB36* and *HB80*, point mutations were 56.4% and 53.8% depleted from clones with a fitness less than 0.1, respectively ($P = 1.35 \times 10^{-18}$ and 1.27×10^{-21}) (Firnberg and Ostermeier 2013). We interpret this result as evidence that the code's arrangement minimizes the fitness cost of amino acid substitutions. An alternative explanation is that *TEM-1*, as a product of millions of years of evolution under the standard genetic code (Hall and Barlow 2004),
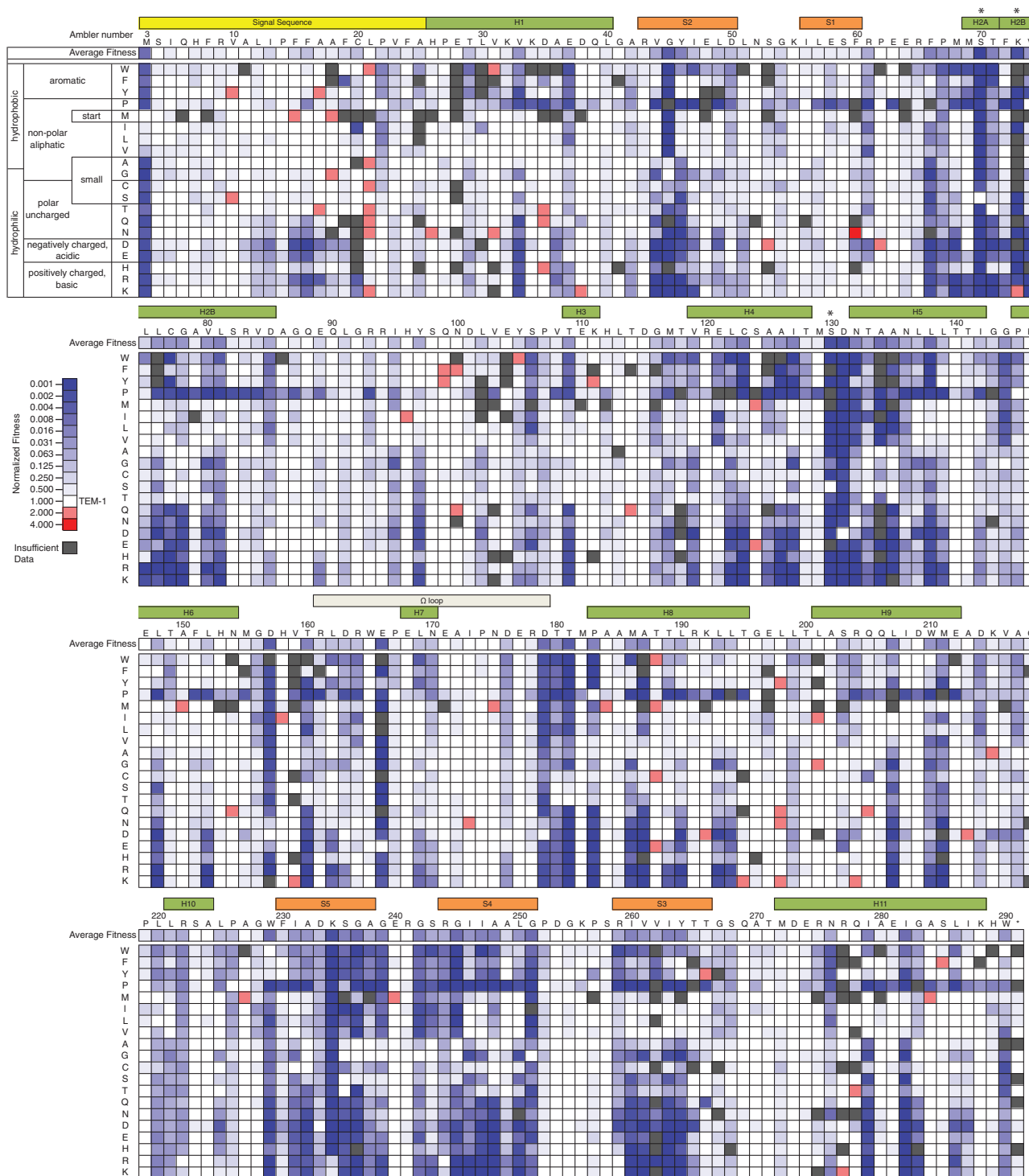
Fig. 2. The sequence-function landscape of TEM-1. The heat map indicates the protein fitness values for ampicillin resistance of the indicated amino acid substitution. The Ambler consensus numbering system (Ambler et al. 1991) for class A β lactamases is used. An asterisk indicates key active site residues. For the start codon, fitness values correspond to the average of the codons for the indicated amino acid though methionine is expected to be the amino acid incorporated. Protein fitness as a function of missense mutation is provided as supplementary data S2 (Supplementary Material online).

evolved to minimize the deleterious effects of mutations under the rules of this code.

We also find further support for our hypothesis that the standard genetic code's architecture enriches for adaptive mutations (Firnberg and Ostermeier 2013). Among the 367 beneficial missense mutations in TEM-1, 41.1% can be

achieved by point mutations, 32.5% higher than the 31.0% expected if 367 missense mutations were chosen at random ($P = 8.8 \times 10^{-6}$ based on comparison to a hypergeometric distribution). Our comprehensive analysis of beneficial mutations in a natural gene in its native host is the strongest evidence yet supporting the hypothesis that the code's

arrangement makes adaptive mutations more likely. The role, if any, such enrichment played in the origin of the genetic code and whether the enrichment is a side effect of the code's error minimization bias are difficult questions to answer (Firnberg and Ostermeier 2013).

The Effects of Synonymous Mutations

The effects of synonymous mutations on protein synthesis and fitness have important implications for evolution and biotechnology. However, despite an abundance of plausible hypotheses, we lack a mechanistic understanding of these effects (Plotkin and Kudla 2011). Our systematic strategy provides an assumption-free approach for testing and generating these hypotheses. We first examined the gene fitness values of 725 alleles synonymous to TEM-1. Beneficial and deleterious synonymous mutations distributed differently across the sequence of TEM-1 (fig. 3A). Beneficial mutations occur primarily in positions 15–30 and 130–260, whereas deleterious mutations appeared in clusters in the first half of the gene and were almost absent from the second half of the gene. No trend in the types of substitutions for either beneficial or deleterious effect was apparent other than eight of the ten beneficial mutations at Arg codons being to the rare *E. coli* codons AGA (2/10) and AGG (6/10). The pattern of beneficial and deleterious synonymous codons indicates the existence of regions of TEM-1 with suboptimal and less robust mRNA properties, respectively.

We next analyzed the effects of 14,055 synonymous substitutions among the set of 15,167 alleles with gene fitness measurements (supplementary fig. S3, Supplementary Material online). Over the length of the entire gene, CUA (Leu), AGG (Arg), and UCG (Ser) provided an average fitness advantage over some of their synonymous codons (supplementary fig. S5, Supplementary Material online), but the advantage was only approximately 5%. Interestingly, CUA and AGG are rare codons in *E. coli*. Codon usage often differs in the beginning of the gene from the rest of the gene, which has been hypothesized to result from a selection against 5′ mRNA structure and/or a selection for rare codons that provide a slower elongation time at the 5′ end (Plotkin and Kudla 2011). Our data address both these hypotheses. Positions 2–10 in TEM-1 had an almost 2-fold broader distribution of synonymous effects compared with any other section of the gene (supplementary fig. S3F, Supplementary Material online). Within these nine positions, we observed 26–85% mean fitness increases for certain codons of Ala, Arg, Gly, Leu, Pro, and Ser relative to select synonyms (supplementary fig. S6, Supplementary Material online). These synonymous fitness differences distributed differently among the nine positions (supplementary fig. S7, Supplementary Material online). Contrary to the slow elongation hypothesis, favored codons tended to appear more frequently in the *E. coli* genome than their corresponding disfavored codon (fig. 3B) (Hilterbrand et al. 2012). However, none of the 16 observed codon preferences were between the most and least frequently used codons within a synonym set suggesting that codon usage was an inadequate explanation for the observed preferences
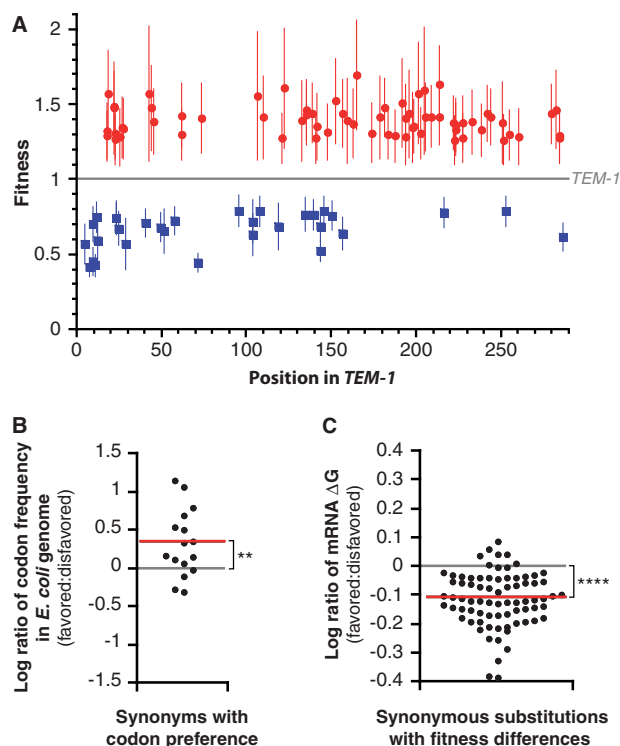

Fig. 3. Effects of synonymous substitutions. (A) Beneficial and deleterious synonymous mutations in TEM-1 are not evenly distributed. Alleles synonymous to TEM-1 with a gene fitness significantly higher (red circles) or lower (blue squares) than that of TEM-1 are shown. The criteria for significance was that the error did not extend into the range fitness = 1 ± 0.1. Error bars provide an upper estimate on error on the fitness measurements as described in the text. (B) An analysis of pairs of synonymous alleles with mutations in codon positions 2–10 of the gene (supplementary figs. S6 and S7, Supplementary Material online) revealed that codon preferences tended to be for codons with a higher frequency in the *Escherichia coli* genome. (C) Preferred codons at positions 2–10 of the gene were predicted to result in mRNA with less stable structures around the initiation codon. Red bar is the mean. $**P < 0.01$, $****P < 0.0001$ by Student's t-test.

(e.g., as a result of tRNA abundance). We next calculated the folding energy of the mRNA around the initiation codon for alleles exhibiting gene fitness differences (Hofacker 2009). In almost all cases, favored codons reduced mRNA stability around the translation start site compared to disfavored codons (fig. 3C). Our findings reinforce recent studies that undermine the slow elongation hypothesis (Supek and Smuc 2010; Charneski and Hurst 2013) and support the theory that mRNA structure at the beginning of genes determines the translation rate (Bentele et al. 2013; Goodman et al. 2013). Like the most recent of these studies (Goodman et al. 2013), our study shows how systematic analyses of large synthetic libraries is a powerful approach for testing competing hypotheses.

Exceptions to the Standard Genetic Code

Among the three stop codons, UAG (amber) exhibited nonsense suppression (supplementary fig. S8A, Supplementary Material online). A 3′ flanking purine after the UAG enhanced

this suppression (supplementary fig. S8B, Supplementary Material online), as has been observed with the amber suppressor tRNA allele supE (Bossi 1983; Miller and Albertini 1983). We sequenced the seven tRNAs known to serve as amber nonsense suppressors and found that E. coli strain SNO301 harbors the supE44 allele, which consists of a duplicate copy of the glnV tRNA gene, glnX, with the expected anticodon mutation (thereby inserting glutamine at UAG codons) (Singaravelan et al. 2010). This allele suppressed a UAG with a 3′ flanking purine at a mean efficiency of 7–10% (supplementary fig. S8C, Supplementary Material online).

Substitutions for the AUG start codon that provided significant antibiotic resistance (>5% of that of TEM-1) included seven of the nine point mutants of AUG (supplementary fig. S9, Supplementary Material online), consistent with known native and nonnative alternative initiation codons in E. coli (Sacerdot et al. 1996; Sussman et al. 1996). In addition, we observed that GUA, GUC, and GUU could serve as weak initiation codons (7–14% as efficient AUG). Initiation from GUA in E. coli has been previously reported (Haggerty and Lovett 1997), but initiation from GUC and GUU has not.

Mutational Tolerance

As the effects of nonsynonymous mutations dwarfed that of synonymous mutations, we combined the gene fitness data of synonymous codons to determine the DFE of missense mutations in the TEM-1 protein (fig. 2). This protein fitness landscape of TEM-1 broadly matched what is known about protein structure in general and TEM-1 in particular. For example, proline was the least tolerated substitution (supplementary fig. S10, Supplementary Material online, which displays TEM-1's amino acid substitution matrix for Amp resistance), especially in alpha helices, and key TEM-1 active site residues did not tolerate mutation (fig. 2). TEM-1's signal sequence is required for export via the Sec pathway to the periplasm. The signal sequence (fig. 2) tolerated most mutations consistent with the pathway's broad specificity (Gierasch 1989). However, the hydrophobic core of the signal sequence did not tolerate substitution of charged residues, consistent with typical export-defective mutants in Sec pathway signal sequences (Gierasch 1989). Signal sequence residue L21 was a hot spot for beneficial mutations, and L21F is found in some extended-spectrum–resistant TEM alleles (Sougakoff et al. 1989).

The comprehensive protein fitness landscape of missense mutations enables a rigorous determination of a protein's mutational tolerance in its biological context. We determined the effective number of amino acids at a position (k^*), which derives from the fitness entropy that is calculated from the distribution of protein fitness values for the 20 amino acids at that position. This measure of tolerance is more informative than establishing an arbitrary fitness cutoff for deciding whether a mutation is tolerated. Our approach is analogous to how information-theoretical entropy is used to measure variability at a position in a set of aligned sequences (Shenkin et al. 1991). However, our measure of tolerance is specific for

TEM-1 and the effect of single amino acid substitutions. This tolerance is a measure of TEM-1's ability to move a Hamming distance of one on the amino acid level and thus does not include epistatic effects. A k^* value of 1 corresponds to a position at which all missense mutations completely inactivate the protein and a k^* value of 20 means that all 19 amino acid substitutions provide the same fitness as the wild-type amino acid.

The distribution of k^* was strongly biased toward high values (fig. 4A). Half of all positions accepted 15.5 or more amino acid substitutions. Under the simple assumption of a linear correlation, percent solvent-accessible surface area accounted for 49% of the k^*'s variance (fig. 4B) and predicted k^* better than distance from the active site (fig. 4C) or a k^* determined from a sequence alignment of 156 class A β-lactamases (Deng et al. 2012) (fig. 4D). Both a k^* based on a multiple sequence alignment (fig. 4D) and previous calculations of k^* for TEM-1 (Deng et al. 2012) (supplementary fig. S11, Supplementary Material online) greatly underestimated TEM-1's mutational tolerance to single amino acid substitutions presumably because epistatic constraints will further limit what sequence combinations are seen naturally, the set of known functional sequences is only a small subset of all possible functional sequences, and a high stringency was used in selecting functional sequences in the case of the later study. The tolerance of amino acid position i weakly correlated with positions $i + 1$, $i + 3$, and $i + 4$ (correlation coefficient 0.25–0.28, $P \leq 1.8 \times 10^{-5}$) but not $i + 2$ or $i > 4$. This correlation primarily occurred at residues with high k^* values. The eight positions with $k^* < 2.5$ include the four strictly conserved residues involved in the catalytic mechanism (S70, K73, S130, and E166) and four other highly conserved residues (fig. 4E). In contrast, the 42 most tolerant positions ($k^* > 19$) predominantly appeared away from the active site in surface loops and at position 2 in alpha helices (fig. 4E). Alpha helices (mean $k^* = 13.5 \pm 5.4$) tolerated substitutions better than beta strands (mean $k^* = 9.89 \pm 4.8$) ($P = 0.0005$ Student's t-test), perhaps a reflection of the buried nature of the beta strands.

Substitution matrices, such as BLOSUM (Henikoff S and Henikoff JG 1992), score the likelihood of substituting one amino acid for another based on alignments of conserved regions of related proteins. A recent study found that the BLOSUM62 matrix best predicted the effect of nonsynonymous mutations in TEM-1 and explained 16% of the variance in the MIC for amoxicillin (Jacquier et al. 2013). We find that BLOSUM62 matrix scores predict 16% of the variance in protein fitness caused by nonsynonymous mutations.

Determinants of Mutational Effects on Protein Fitness

What basic phenomena underlie the DFE? For an enzyme, fitness (w) will strongly depend on the total catalytic activity in the cell (v_t), which is a product of the enzyme's specific catalytic activity (v_{sp}) and the protein abundance (P), which is how much protein is present in the cell in a correctly folded, soluble form.
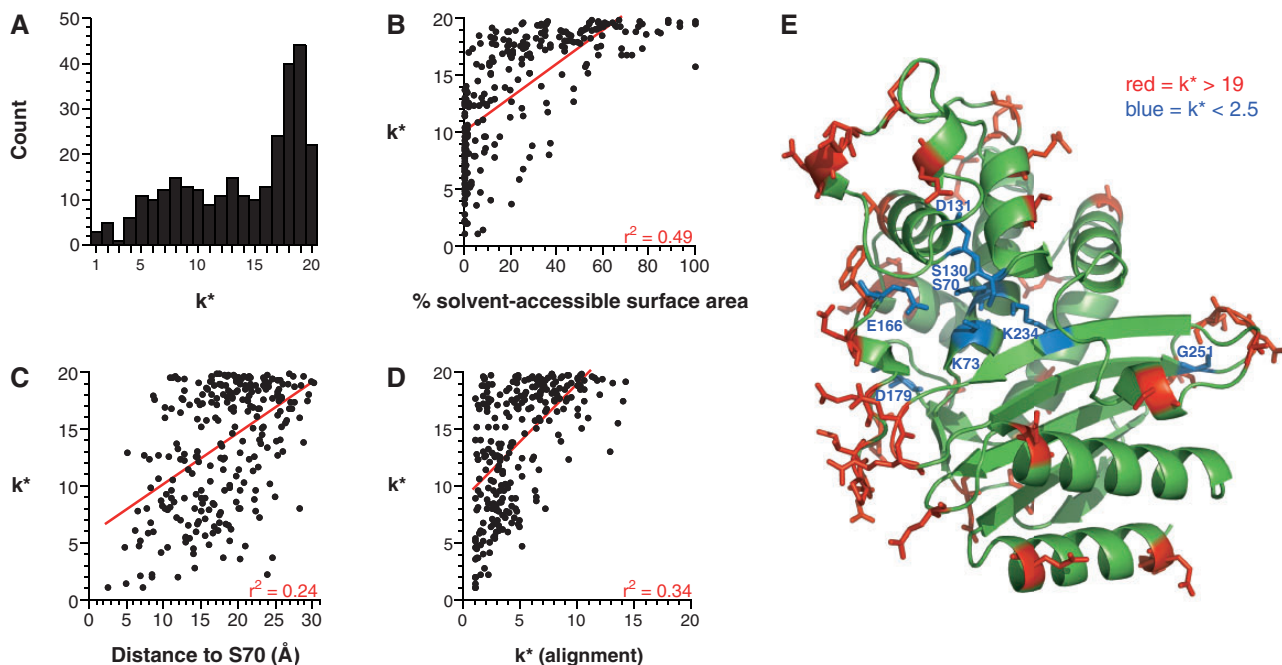
$$w = f(v_t) \qquad (1)$$

FIG. 4. Tolerance of TEM-1 to missense mutation. Tolerance was measured by the effective number of amino acids at a position (k^*), which derives from the distribution of protein fitness values for the 20 amino acids at that position. k^* ranges in value from 1 (position is completely intolerant to substitution) to 20 (position tolerates all possible amino acids with no loss in fitness). (A) The distribution of k^* values in TEM-1. (B) Correlation of k^* correlates with percent solvent-accessible surface. (C) Correlation of k^* with distance from the active site. (D) Correlation of k^* with a sequence alignment of 156 class A β lactamases (Deng et al. 2012). (E) Model of TEM-1 (PDB ID 1XPB [Fonze et al. 1995]) indicating the least tolerant positions ($k^* < 2.5$, shown in blue), which include the key active site residues S70, K73, S130, and E166, and the most tolerant positions ($k^* > 19$, shown in red).

$$\nu_t = \nu_{sp}P \qquad (2)$$

For many genes, especially essential genes, the functional form of equation (1) is likely complex. For example, an increase in ν_t may be detrimental for fitness if it negatively perturbs metabolic flux in the cell. In addition, essential genes are likely to have evolved to be buffered against the deleterious effects of mutation (i.e., they possess a ν_t that is well above a level that would compromise fitness). A study of 27 mutants of the essential enzyme dihydrofolate reductase (DHFR) supports this idea (Bershtein et al. 2012). The mutations were chosen to have little effect on specific catalytic activity but a range of effects on thermostability. The authors found that organismal fitness (i.e., growth rate) only weakly correlated with protein abundance, and large decreases in protein abundance generally had marginal effects on fitness. A follow-up study estimated that DHFR could sustain an 80% cut in protein abundance with little effect on organismal fitness and that the dependence of fitness on protein abundance exhibited Michaelis–Menten-like behavior (Bershtein et al. 2013). This study illustrates the challenge of gaining insight into the basic phenomena underling the DFE without knowledge of the form of equation (1). *TEM-1* offers a simple case for addressing this issue, as TEM-1 fulfills a single cellular role (inactivation of β-lactam antibiotics), and the reaction's substrate and product are not part of any native *E. coli* metabolic or signaling pathway. As a result, fitness for *TEM-1* is directly proportional to the total antibiotic hydrolysis

activity in the cell (Soskine and Tawfik 2010), as shown in equation (3):

$$w \propto \nu_t = \nu_{sp}P \qquad (3)$$

and this facilitates an analysis of the relative effects of mutation on both ν_{sp} and P.

Experimental evolution studies have shown that ν_{sp} and P are equally important targets for adaptive evolution (Counago et al. 2008; Walkiewicz et al. 2012). We expect protein abundance to be a function of the thermodynamic stability (ΔG) as well as protein production rates (i.e., arising from mRNA properties, interactions with chaperones) and degradation rates (i.e., proteolytic susceptibility). Both computational and experimental studies show that, on average, missense mutations decrease thermodynamic stability (Tokuriki et al. 2007). A prevailing hypothesis on the origin of deleterious fitness effects of mutation states that thermodynamic stability is the primary determinant of the DFE through its effect on protein abundance (DePristo et al. 2005; Camps et al. 2007; Tokuriki et al. 2007; Wylie and Shakhnovich 2011). Although the hypothesis is intuitive and appealing, experimental evidence for a significant correlation between protein stability and fitness via an effect on protein abundance is scant (Bershtein et al. 2012). Mutations that reduce function often show decreased protein abundance (Pakula et al. 1986; Schultz and Richards 1986); however, mutations that increase stability can reduce specific activity (Shoichet et al. 1995) and reductions in protein
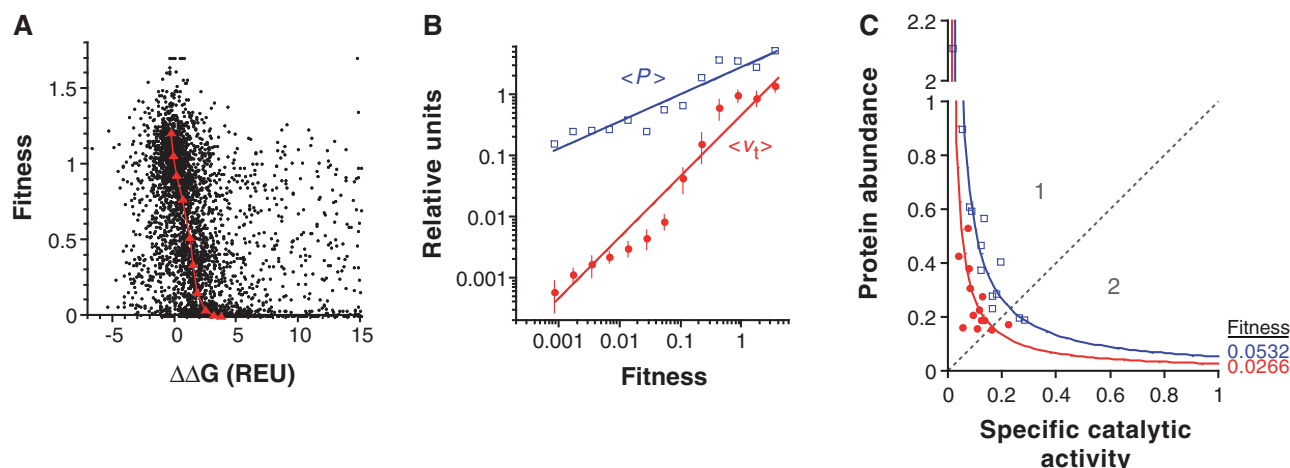
Fig. 5. The determinants of protein fitness. (*A*) Loss of fitness correlates with loss of thermodynamic stability. Protein fitness is shown as a function of change in ΔG as predicted by Rosetta (Chaudhury et al. 2010) for 4,783 missense mutations in wild-type TEM-1. The median predicted $\Delta\Delta G$ for fitness deciles is shown in red triangles. Predicted changes > 15 REU are not shown and are not considered in the median calculation. The distribution of $\Delta\Delta G$ for select fitness deciles can be found in supplementary figure S12A, Supplementary Material online. (*B*) Total cellular catalytic activity determines TEM-1 protein fitness. The average total cellular catalytic activity $<v_t>$ and the average protein abundance $<P>$ were experimentally measured for 13 sublibraries of ~15,000 unique TEM-1 alleles partitioned based on relative fitness. The values of $<v_t>$ and $<P>$ are relative to that of TEM-1. The slight sigmoidal form of $<v_t>$ is an expected artifact of the methodology (supplementary fig. S14, Supplementary Material online). The error bars represent the standard deviation of six assays from two independent experiments. The lines are guides for the eye. (*C*) Protein fitness phase space defined by equation (3). The protein abundance and specific catalytic activity (relative to TEM-1) of 26 randomly selected members of sublibraries 6 (red solid circle) and 7 (blue open square) is shown. The dotted line corresponds to an equal decrease in protein abundance and specific catalytic activity. In region 1, a mutation affects specific catalytic activity more than protein abundance. The solid lines are of constant fitness at the average expected fitness values of the two sublibraries from which the alleles were randomly selected.

stability often accompany adaptive mutations (Wang et al. 2002). In the aforementioned study of 27 mutations that destabilized DHFR (most of which were buried in the hydrophobic core of the protein) the authors found that although protein abundance correlated with thermostability ($r^2 = 0.41$ at 37 °C), organismal fitness changed very little with protein abundance (Bershtein et al. 2012). The study could not address how the deleterious effects of mutations partition between effects on specific catalytic activity and protein abundance because the mutations were selected to be those that do not affect catalytic activity. A study of 990 missense mutations in TEM-1 found that 15–19% of the variance in amoxicillin resistance could be explained by the computationally predicted change in protein stability caused by the introduction of the mutation in TEM-116 (TEM-116 is TEM-1 with the V84I and A184V mutations) (Jacquier et al. 2013); however, the study did not address the mutations' effect on protein abundance or specific catalytic activity. A comprehensive, systematic study of 1) the relationship between fitness and thermostability, and 2) the relative contributions of protein abundance and specific activity to the deleterious effects of mutations would more fully address the fundamental phenomena underlying the DFE.

We predicted $\Delta\Delta G$ ($\Delta G_{wild-type} - \Delta G_{mutant}$) using Rosetta (Das and Baker 2008; Chaudhury et al. 2010) for 4,783 missense mutations of TEM-1, allowing limited backbone flexibility (fig. 5A). Variants that were predicted to be more stable tended to have higher protein fitness (fig. 5A; supplementary fig. S12A, Supplementary Material online). The larger a mutation's deleterious effect on fitness, the higher the probability

that the mutation produced a very large predicted energy score (supplementary fig. S12A, Supplementary Material online). Predictions of $\Delta\Delta G$ using PoPMuSiC (Dehouck et al. 2011), a more empirical approach than Rosetta to predicting changes in protein stability, produced similar results and indicated that 18% of the variance in protein fitness can be explained by thermostability (supplementary fig. S12B, Supplementary Material online). We compared fitness with experimentally measured melting temperatures of 36 TEM-1 alleles and found a positive correlation with an r^2 of 0.53 (supplementary fig. S12D, Supplementary Material online). This suggests that limitations in the computational prediction of $\Delta\Delta G$ result in an underestimation of the degree to which thermodynamic stability determines fitness. The lack of a positive correlation between melting temperature and fitness observed in a previous study (Bershtein et al. 2012) underscores the fact that protein stability effects on fitness will be observed only when the fitness function of equation (1) is in a regime where changes to v_t affect fitness. Our systematic and comprehensive approach to examining the relationship between protein stability and protein fitness for a single protein, combined with our observed correlation between melting temperature and fitness provides strong experimental evidence that effects on protein stability significantly shape the DFE.

Whether mutational reductions in protein abundance as opposed to specific activity are the major cause of loss of fitness has not previously been experimentally addressed. We experimentally addressed this question by analyzing the soluble fraction of cell lysates of the 13 sublibraries and

randomly selected alleles from our *TEM-1* library. We first established that w and ν_t are directly proportional as predicted by equation (3) by measuring the mean total hydrolysis activity of the cell $<\nu_t>$ for the 13 sublibraries of CCM-2 (fig. 5*B*). We measured protein abundance by quantitative western blot of the soluble fraction of cell lysates (supplementary fig. S13, Supplementary Material online). We assumed all soluble protein was folded and active. The mean protein abundance $<P>$ of the sublibraries did not decrease nearly as rapidly with decreasing protein fitness as $<\nu_t>$ did (fig. 5*B*). In addition, an increase in the mean amount of aggregated protein did not accompany a loss of fitness (supplementary fig. S15, Supplementary Material online). These findings suggest that mutational effects on ν_{sp} rather than on P may play the larger role in the deleterious effects of mutation. As this interpretation hinges on the distribution of values of P in the sublibraries, we measured P for 27 randomly selected alleles with a protein fitness of about 0.025–0.05 (i.e., the alleles of supplementary fig. S4, Supplementary Material online). We chose this fitness range so that the mutational effects were substantial, but not inactivating. This ensured that our conclusions would not depend on small changes in w and P. We excluded the I13E allele from analysis since this mutation in the signal sequence caused a defect in normal export/proteolytic processing (supplementary fig. S13, Supplementary Material online). The remaining 26 alleles exhibited a decrease in both protein abundance and predicted thermodynamic stability relative to TEM-1 with the exception of the R244E allele, which showed an increase in both (supplementary fig. S12C, Supplementary Material online). From w and P, we calculated ν_{sp} using equation (3) and examined the protein fitness phase space by plotting P versus ν_{sp} (fig. 5*C*). We find that the deleterious effects of mutations, at least for mutations with large deleterious effects, arise more from a decrease in specific activity than from a decrease in protein abundance. Despite the large negative effects on specific activity, the mutated residues of the 26 alleles were not clustered around the active site but were scattered throughout the protein (supplementary fig. S4C, Supplementary Material online). Thus, the dominant effect of mutation on specific activity does not arise because the 26 mutations were biased to be proximal to the active site. We postulate that mutational effects on specific activity may be as important to the DFE at high fitness as at low fitness, but this postulate requires experimental investigation.

We do not interpret the diminished role of mutational effects on protein abundance as reducing the role of thermodynamic stability in fitness. Protein stability, in addition to its effect on protein abundance, may exert its effect on fitness through a decrease in a protein's specific activity. Perhaps, this manifests by perturbing the conformational ensemble away from more active states or by increasing the number of states (i.e., altering protein dynamics). Protein abundance's relative resilience to decreases in thermodynamic stability is striking but fits the growing appreciation that the cellular environment is not a passive solution at equilibrium (Bershtein et al. 2013). Rather, the cellular environment acts as a buffer for deleterious mutational effects on protein abundance through

the effect of chaperones, proteins that facilitate the proper folding of other proteins (Rutherford 2003). Chaperone overexpression can compensate for the deleterious mutational effects on protein abundance (Tokuriki and Tawfik 2009) and fitness (Bershtein et al. 2013). This theory offers an explanation for TEM-1's stability threshold that buffers the effect of mutations on fitness (Bershtein et al. 2006). We suspect that the relative contribution of protein abundance to fitness may increase with the number of mutations as the protein's stability margin is exhausted by the cumulative effect of mutations, an effect that is characterized by negative epistasis (Bershtein et al. 2006). As such, negative epistasis may arise in part as a consequence of the beneficial properties of the cellular environment in addition to a protein's intrinsic stability margin.

Conclusions

The application of synthetic biology to the study of fundamental biological questions, as we have done in this study of gene and protein fitness landscapes, offers a well-defined, systematic approach for testing and generating hypotheses. Our comprehensive determination of the fitness effects of mutation of *TEM-1* provides the first detailed maps of fitness landscapes corresponding to a gene and its nearest neighbors at the basepair, codon, and amino acid level. To the extent that *TEM-1* is a representative gene, our study provides several important insights. Evolution must traverse fitness landscapes under the constraints of the genetic code—constraints that minimize the effect of mutation and enrich for beneficial mutations. The small fitness effects of synonymous mutations have complex determinants including regional proclivities for synonymous fitness effects in the gene. At the beginning of the gene, fitness effects of synonymous mutations strongly correlate with mRNA stability. Missense mutational effects on thermodynamic stability shape the DFE, but their deleterious effect on specific protein activity tends to exceed that on protein abundance, at least for mutations with large deleterious effects. We hypothesize that TEM-1's high mutational tolerance may in part derive from the cell's buffering capacity to mediate the deleterious effects of lost stability on protein abundance, a phenomena that might give rise to negative epistasis. Further inquiry into the fundamental determinants of the landscape's topology is necessary to address this hypothesis and substantiate these findings.

Materials and Methods

Fitness Determination

Escherichia coli SNO301 (*ampD1*, *ampA1*, *ampC8*, *pyrB*, *recA*, and *rpsL*) cells harboring the comprehensive codon mutagenesis library CCM2 (Firnberg and Ostermeier 2012) were plated on LB-agar plates supplemented with 13 different Amp concentrations (2-fold increments ranging from 0.25 µg/ml to 1,024 µg/ml) to partition the library into 13 partially overlapping sublibraries based on relative Amp resistance using a synthetic gene circuit that functions as a tunable band-pass genetic selection for Amp resistance (Sohka et al. 2009) (supplementary figs. S1 and S2, Supplementary Material online).

Barcoded PCR amplicons were prepared from each plate and subjected to 454 deep-sequencing. The 1,325,979 sequencing reads were filtered for quality and for reads that only contained one codon substitution. We tabulated the number of sequencing counts for each allele at each Amp concentration and determined the fitness w relative to TEM-1 from the statistics. As the distribution of growth as a function of Amp is roughly symmetric when plotted as the $\log_2$(Amp concentration) (Sohka et al. 2009), we determined the unnormalized fitness f of allele i as

$$f_i = \frac{\sum_{p=1}^{13} c_{i,p} \log_2(a_p)}{\sum_{p=1}^{13} c_{i,p}} \qquad (4)$$

in which $c_{i,p}$ is the number of counts of allele i on sublibrary plate p in the deep sequencing data and a_p is the concentration of Amp on sublibrary plate p in μg/ml. We normalized all fitnesses by the fitness of wild type as follows:

$$w_i = \frac{2^{f_i}}{2^{f_{WT}}}. \qquad (5)$$

This result is a normalized fitness w_i that is 1.0 for wild-type TEM-1, greater than 1.0 for beneficial mutations, and between 0 and 1.0 for deleterious mutations. We determined the fitness of wild-type TEM-1 (f_{WT}) using equation (4) using the counts of all alleles with a synonymous substitution in TEM-1, because the fitness of these varied very little. As a check, we compared this value with the fitness determined by equation (4) using the counts of all sequencing reads that lacked a mutation. The two values differed by only 2.5%. We determined an upper limit on the error in our fitness measurements from the DFEs of synonymous mutations as a function of the number of times an allele was observed (supplementary fig. S3, Supplementary Material online). Fitness values and error estimates are tabulated in supplementary data S1 and S2 (Supplementary Material online).

Prediction of mRNA Stability at the Transcript Start

The RNAfold utility of the Vienna RNA Package (version 2.1.2) was used to predict the minimum free energy of RNA sequences (Hofacker 2009). For each allele in supplementary figure S7 (Supplementary Material online), the Gibbs free energy was calculated as the average free energy of every 39 nt window centered on nucleotides from −5 to +10 of the gene start as described (Bentele et al. 2013).

Mutational Tolerance

The observed effective number of amino acids k_o^* at a position was determined from the protein fitness values of the n missense mutations with fitness data at that position using equations (6) and (7).

$$S = -\sum_{i=1}^{n} w_i \log_2 w_i \qquad (6)$$

$$k_o^* = 2^S \qquad (7)$$

We obtained the effective number of amino acids k^* by normalizing k_o^* to be based on 20 amino acids by equation (8).

$$k^* = \frac{20 k_o^*}{n} \qquad (8)$$

A table of k_o^* and k^* is provided as supplementary data S4 (Supplementary Material online).

Prediction of Protein Thermodynamic Stability

PyRosetta v3.4.0 r55307 (Chaudhury et al. 2010) was used to compute the difference in score (in Rosetta energy units [REU]) between the mature structures (lacking the signal sequence) of each amino acid mutant and wild-type TEM-1 (Protein Data Bank identifier 1XPB; Fonze et al. 1995). PopMusic predictions of $\Delta\Delta G$ (supplementary fig. S12B, Supplementary Material online) were determine online at http://babylone.ulb.ac.be/popmusic (last accessed February 17, 2014) (Dehouck et al. 2011) using 1XPB.

Protein Abundance and Total Catalytic Activity

Relative protein abundance was quantified by using Quantity One 1-D analysis software (Bio-Rad) of Western blots of the soluble fraction of cell lysates in comparison with a standard curve. Representative westerns are shown in supplementary figure S13, Supplementary Material online. Catalytic activity of the sublibraries and clones was determined by measuring the rate of hydrolysis (at 486 nm) of 50 μM nitrocefin in 10 mM phosphate buffer pH 7.4 at 37 °C. The initial rate was normalized by the total amount of protein added for each sample.

Supplementary Material

Supplementary materials and methods, figures S1–S15, and data S1–S4 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

References

Ambler RP, Coulson AF, Frere JM, Ghuysen JM, Joris B, Forsman M, Levesque RC, Tiraby G, Waley SG. 1991. A standard numbering scheme for the class A beta-lactamases. Biochem J. 276(Pt 1):269–270.

Araya CL, Fowler DM, Chen W, Muniez I, Kelly JW, Fields S. 2012. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. Proc Natl Acad Sci U S A. 109:16858–16863.

Bentele K, Saffert P, Rauscher R, Ignatova Z, Bluthgen N. 2013. Efficient translation initiation dictates codon usage at gene start. Mol Syst Biol. 9:675.

Bershtein S, Mu W, Serohijos AW, Zhou J, Shakhnovich EI. 2013. Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. Mol Cell. 49:133–144.

Bershtein S, Mu W, Shakhnovich EI. 2012. Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc Natl Acad Sci U S A.* 109:4857–4862.

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

Bossi L. 1983. Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. *J Mol Biol.* 164:73–87.

Camps M, Herman A, Loh E, Loeb LA. 2007. Genetic constraints on protein evolution. *Crit Rev Biochem Mol Biol.* 42:313–326.

Charneski CA, Hurst LD. 2013. Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol.* 11:e1001508.

Chaudhury S, Lyskov S, Gray JJ. 2010. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* 26:689–691.

Counago R, Wilson CJ, Pena MI, Wittung-Stafshede P, Shamoo Y. 2008. An adaptive mutation in adenylate kinase that increases organismal fitness is linked to stability-activity trade-offs. *Protein Eng Des Sel.* 21:19–27.

Das R, Baker D. 2008. Macromolecular modeling with Rosetta. *Annu Rev Biochem.* 77:363–382.

Dehouck Y, Kwasigroch JM, Gilis D, Rooman M. 2011. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics* 12:151.

Deng Z, Huang W, Bakkalbasi E, Brown NG, Adamski CJ, Rice K, Muzny D, Gibbs RA, Palzkill T. 2012. Deep sequencing of systematic combinatorial libraries reveals beta-lactamase sequence constraints at high resolution. *J Mol Biol.* 424:150–167.

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

Deris JB, Kim M, Zhang Z, Okano H, Hermsen R, Groisman A, Hwa T. 2013. The innate growth bistability and fitness landscapes of antibiotic-resistant bacteria. *Science* 342:1237435.

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

Firnberg E, Ostermeier M. 2012. PFunkel: efficient, expansive, user-defined mutagenesis. *PLoS One* 7:e52031.

Firnberg E, Ostermeier M. 2013. The genetic code constrains yet facilitates Darwinian evolution. *Nucleic Acids Res.* 41:7420–7428.

Fonze E, Charlier P, To'th Y, Vermeire M, Raquet X, Dubus A, Frere JM. 1995. TEM1 beta-lactamase structure solved by molecular replacement and refined structure of the S235A mutant. *Acta Crystallogr D Biol Crystallogr.* 51:682–694.

Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, Fields S. 2010. High-resolution mapping of protein sequence-function relationships. *Nat Methods* 7:741–746.

Gierasch LM. 1989. Signal sequences. *Biochemistry* 28:923–930.

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

Haggerty TJ, Lovett ST. 1997. IF3-mediated suppression of a GUA initiation codon mutation in the recJ gene of *Escherichia coli. J Bacteriol.* 179:6705–6713.

Hall BG, Barlow M. 2004. Evolution of the serine beta-lactamases: past, present and future. *Drug Resist Updat.* 7:111–123.

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

Hilterbrand A, Saelens J, Putonti C. 2012. CBDB: the codon bias database. *BMC Bioinformatics* 13:62.

Hofacker IL. 2009. RNA secondary structure analysis using the Vienna RNA package. *Curr Protoc Bioinformatics.* Chapter 12:Unit12.12.

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.* 110:13067–13072.

McLaughlin RN Jr, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. 2012. The spatial architecture of protein function and adaptation. *Nature* 491:138–142.

Medeiros AA. 1984. Beta-lactamases. *Br Med Bull.* 40:18–27.

Miller JH, Albertini AM. 1983. Effects of surrounding sequence on the suppression of nonsense codons. *J Mol Biol.* 164:59–71.

Orr HA. 2005. The genetic theory of adaptation: a brief history. *Nat Rev Genet.* 6:119–127.

Pakula AA, Young VB, Sauer RT. 1986. Bacteriophage lambda cro mutations: effects on activity and intracellular degradation. *Proc Natl Acad Sci U S A.* 83:8829–8833.

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.

Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet.* 12:32–42.

Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DN. 2013. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J Mol Biol.* 425:1363–1377.

Rutherford SL. 2003. Between genotype and phenotype: protein chaperones and evolvability. *Nat Rev Genet.* 4:263–274.

Sacerdot C, Chiaruttini C, Engst K, Graffe M, Milet M, Mathy N, Dondon J, Springer M. 1996. The role of the AUU initiation codon in the negative feedback regulation of the gene for translation initiation factor IF3 in *Escherichia coli. Mol Microbiol.* 21:331–346.

Salverda ML, De Visser JA, Barlow M. 2010. Natural evolution of TEM-1 beta-lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiol Rev.* 34:1015–1036.

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.

Schlinkmann KM, Honegger A, Tureci E, Robison KE, Lipovsek D, Pluckthun A. 2012. Critical features for biosynthesis, stability, and functionality of a G protein-coupled receptor uncovered by all-versus-all mutations. *Proc Natl Acad Sci U S A.* 109:9810–9815.

Schultz SC, Richards JH. 1986. Site-saturation studies of beta-lactamase: production and characterization of mutant beta-lactamases with all possible amino acid substitutions at residue 71. *Proc Natl Acad Sci U S A.* 83:1588–1592.

Shenkin PS, Erman B, Mastrandrea LD. 1991. Information-theoretical entropy as a measure of sequence variability. *Proteins* 11:297–313.

Shoichet BK, Baase WA, Kuroki R, Matthews BW. 1995. A relationship between protein stability and protein function. *Proc Natl Acad Sci U S A.* 92:452–456.

Singaravelan B, Roshini BR, Munavar MH. 2010. Evidence that the supE44 mutation of *Escherichia coli* is an amber suppressor allele of glnX and that it also suppresses ochre and opal nonsense mutations. *J Bacteriol.* 192:6039–6044.

Smith JM. 1970. Natural selection and the concept of a protein space. *Nature* 225:563–564.

Sohka T, Heins RA, Phelan RM, Greisler JM, Townsend CA, Ostermeier M. 2009. An externally-tunable bacterial band-pass filter. *Proc Natl Acad Sci U S A.* 106:10135–10140.

Sonneborn TM. 1965. Degeneracy of the genetic code: extent, nature, and genetic implications. In: Bryson V, Voge HJ, editors. Evolving genes and proteins. New York: Academic Press. p. 377–397.

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

Sougakoff W, Petit A, Goussard S, Sirot D, Bure A, Courvalin P. 1989. Characterization of the plasmid genes blaT-4 and blaT-5 which encode the broad-spectrum beta-lactamases TEM-4 and TEM-5 in enterobacteriaceae. *Gene* 78:339–348.

Starita LM, Pruneda JN, Lo RS, Fowler DM, Kim HJ, Hiatt JB, Shendure J, Brzovic PS, Fields S, Klevit RE. 2013. Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. *Proc Natl Acad Sci U S A.* 110:E1263–E1272.

Supek F, Smuc T. 2010. On relevance of codon usage to expression of synthetic and natural genes in *Escherichia coli. Genetics* 185:1129–1134.

Sussman JK, Simons EL, Simons RW. 1996. *Escherichia coli* translation initiation factor 3 discriminates the initiation codon in vivo. *Mol Microbiol.* 21:347–360.

Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS. 2007. The stability effects of protein mutations appear to be universally distributed. *J Mol Biol.* 369:1318–1332.

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

Toth-Petroczy A, Tawfik DS. 2013. Protein insertions and deletions enabled by neutral roaming in sequence space. *Mol Biol Evol.* 30: 761–771.

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.* 109:21408–21413.

Wang X, Minasov G, Shoichet BK. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol.* 320:85–95.

Weinreich DM, Delaney NF, Depristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111–114.

Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, et al. 2012. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol.* 30:543–548.

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.

Woese CR. 1965. On the evolution of the genetic code. *Proc Natl Acad Sci U S A.* 54:1546–1552.

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

A comprehensive, high-resolution map of a gene's fitness landscape

Elad Firnberg[1], Jason W. Labonte[1], Jeffrey J. Gray[1], and Marc Ostermeier[1]

[1]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218 USA
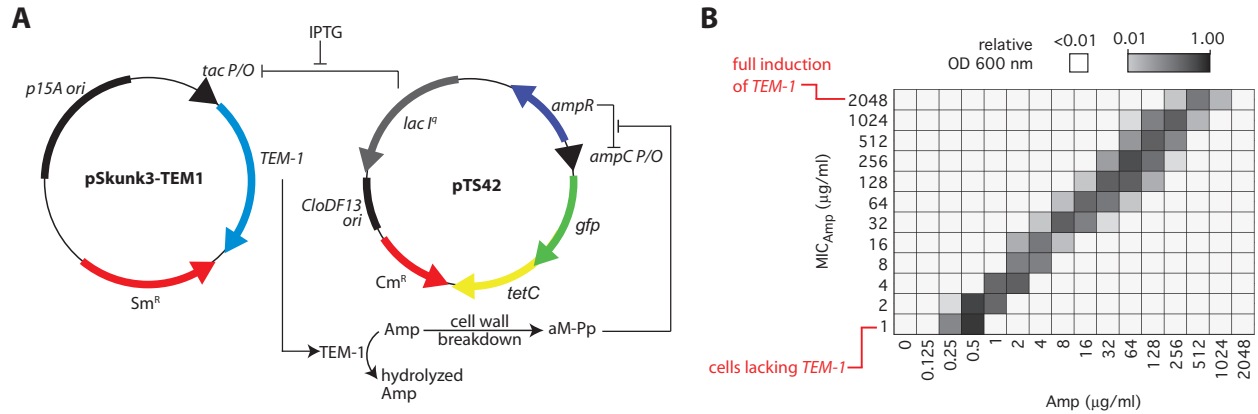
Figures S1-S15

Fig. S1. Bacterial band-pass filter for beta-lactamase activity. (A) Essential components of the genetic circuit for band-pass selection (Sohka, et al. 2009). In the absence of sufficient cellular TEM-1 β-lactamase activity for hydrolysis of ampicillin (Amp), cell wall synthesis is compromised and cells cannot proliferate. Cell wall breakdown results in the accumulation of aM-pentapeptide (aM-Pp), which induces the *ampC* promoter via interactions with AmpR (Dietz, et al. 1997; Valtonen, et al. 2002) resulting in the production of TetC (which confers tetracycline resistance) and the green fluorescent protein (GFP). The level of Amp necessary to induce *ampC* is lower than the level that prevents the growth of *E. coli* cells (Valtonen, et al. 2002). Thus, cells that hydrolyze Amp too efficiently cannot grow in the presence of Tet. As a result, in the presence of tetracycline and Amp, cells will proliferate only if they possess an intermediate amount of Amp hydrolysis activity. The level of Amp hydrolysis activity necessary for growth increases linearly with Amp concentration (Valtonen, et al. 2002). *TEM-1* expression is regulated through IPTG-induction of the LacI-repressed *tac* promoter. (B) Demonstration of band-pass selection for BLA activity in *E. coli* SNO301 cells (Sohka, et al. 2009). The growth of cells in liquid LB media containing 20 µg/ml Tet was detected by measuring the OD at 600 nm of the liquid culture and is presented in the form of a heat map as a function of Amp concentration (x-axis) for cells with different levels of β-lactamase activity (y-axis, as measured by the minimum inhibitory concentration (MIC) for Amp).
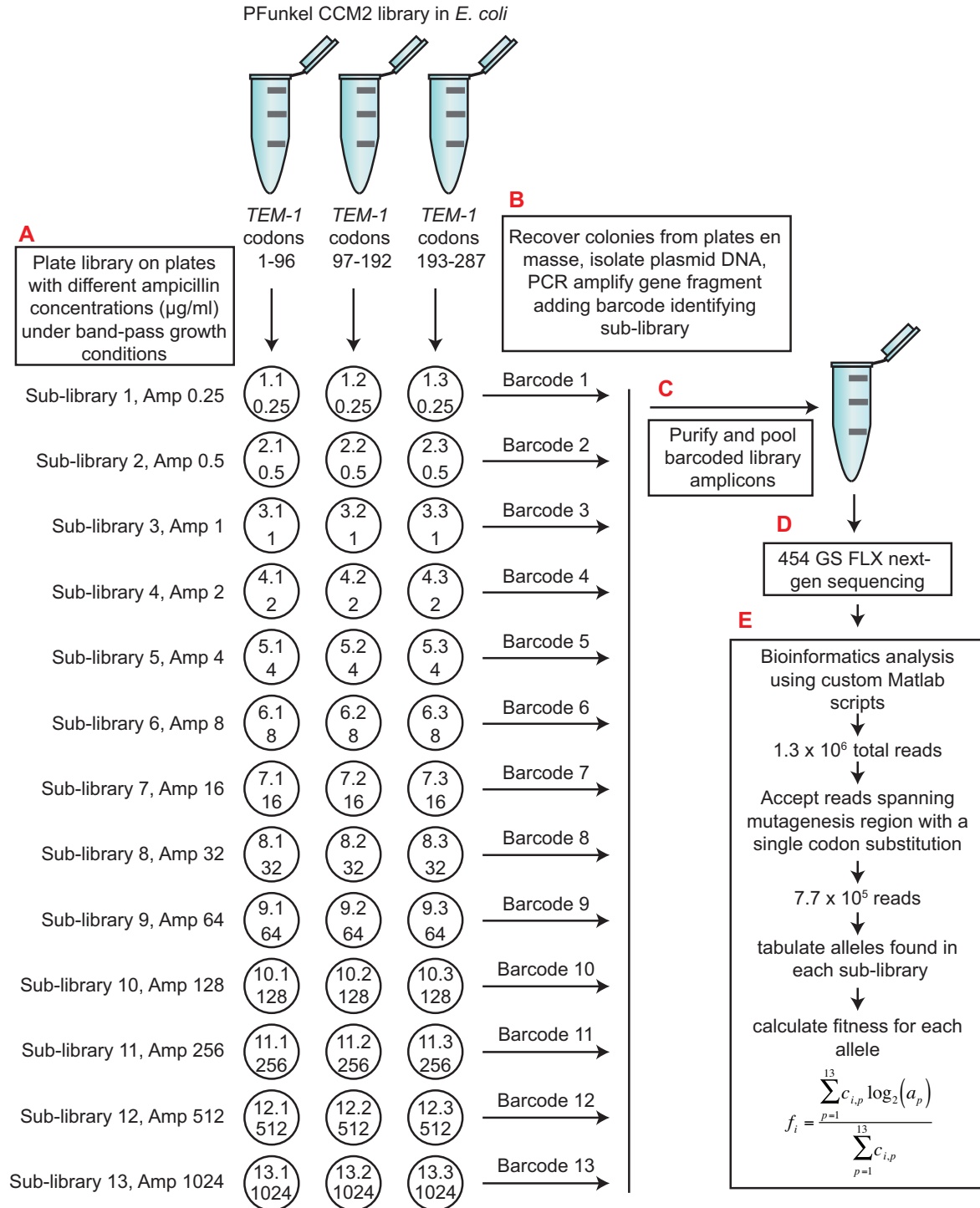
Fig. S2. System for measuring gene fitness of *TEM-1* alleles. The CCM2 library in strain SNO301 (separate libraries for each third of the *TEM-1* gene) was plated on media containing Tet and different concentrations of Amp. PCR products with barcodes identifying the plating conditions were subjected to 454 GS FLX DNA sequencing and the counts of alleles on each of the growth conditions was used to quantify fitness (see Methods).
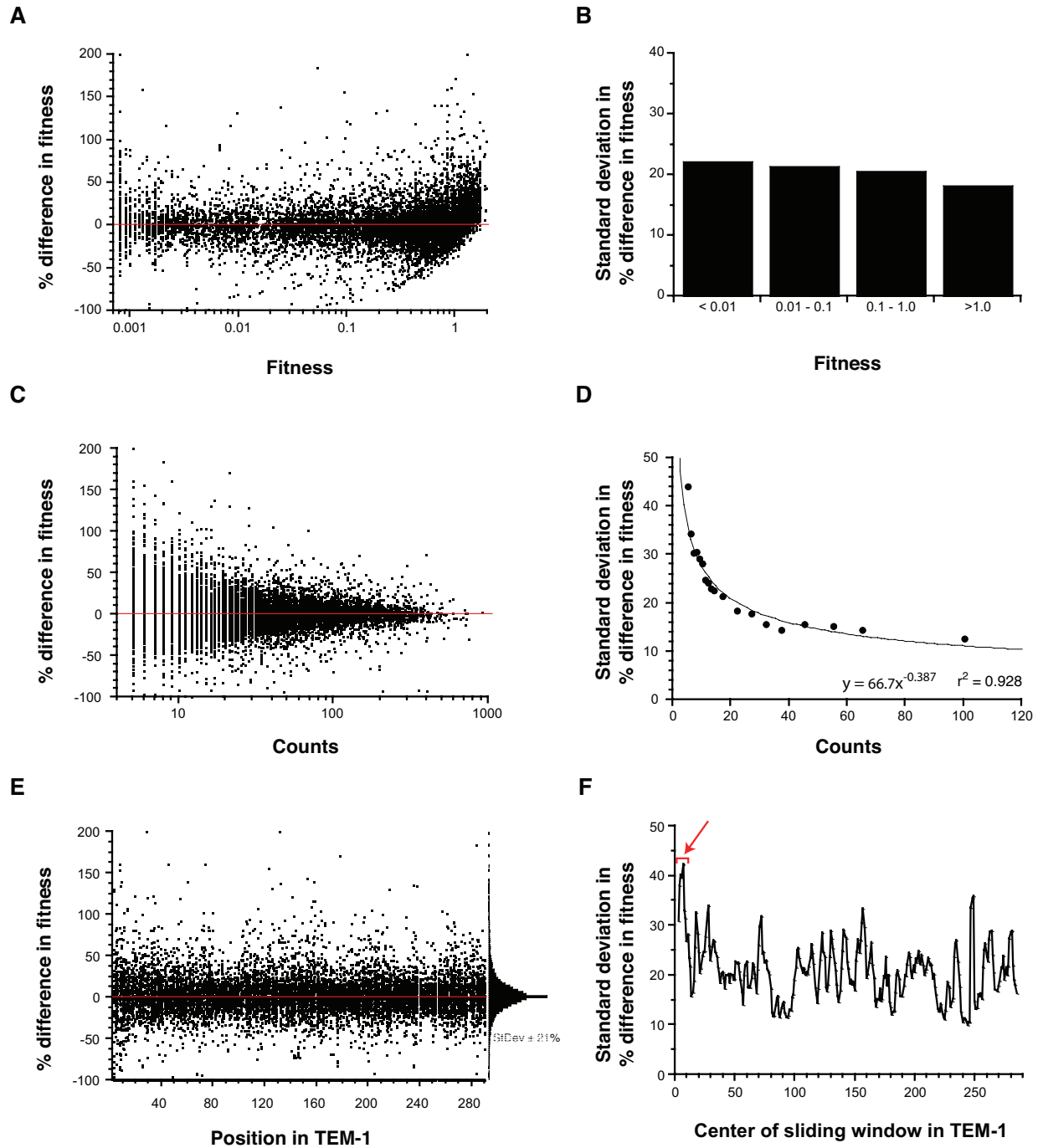
A

B

C

D

$y = 66.7x^{-0.387}$ $r^2 = 0.928$

E

StDev ± 21%

F

Fig. S3. Distribution of synonymous effects. We determined the percent difference in gene fitness between the fitness of allele *i* with a mutation at codon *j* of the gene and the mean fitness of all alleles with a synonymous mutation at codon *j* (including allele *i*). (A) The distribution of percent fitness difference as a function of fitness indicates that the fitness measurement is equally precise at low and high fitness values. (B) This observation is further illustrated by the standard deviation of percent fitness differences for different fitness ranges. (C) The percent fitness difference decreases with the number of times an allele is encountered in the deep sequencing experiment (i.e. the

counts).   (D) The standard deviation in percent fitness difference as a function of counts, defined here, was used as an upper limit of error in the fitness measurements. (E) The percent fitness difference is fairly uniform across the *TEM-1* sequence. (F) However, a broader distribution in the first ten codons of the gene is apparent. A sliding window of three positions was used.
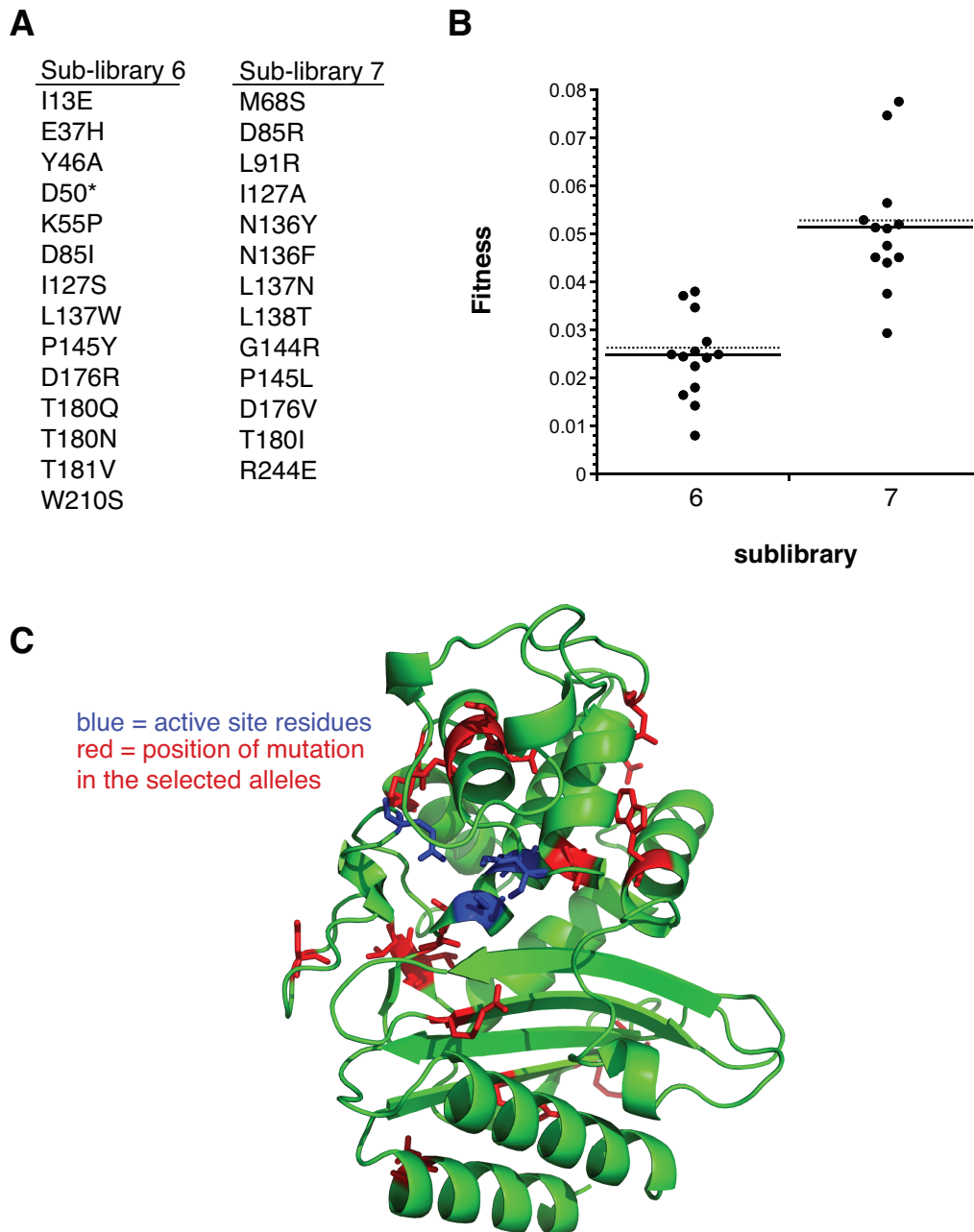
A

Sub-library 6	Sub-library 7
I13E	M68S
E37H	D85R
Y46A	L91R
D50*	I127A
K55P	N136Y
D85I	N136F
I127S	L137N
L137W	L138T
P145Y	G144R
D176R	P145L
T180Q	D176V
T180N	T180I
T181V	R244E
W210S	

B

C

blue = active site residues
red = position of mutation
in the selected alleles

Fig. S4. Randomly selected members of sub-libraries 6 and 7. These 27 members were selected on plates with 8 µg/ml Amp (sub-library 6) or 16 Amp (sub-library 7). (A) Mutations in the 27 alleles (each allele had one mutation). * indicates the mutation is to the amber stop codon (UAG).   (B) Gene fitness values. The solid line indicates the mean of the randomly selected members. The dotted line indicates the expected mean of the sub-library based on the Amp concentration used in the genetic selection to obtain the sub-library. (C) The distribution of mutational sites on the structure of TEM-1 (Fonze, et al. 1995). Red indicates the mutational sites and blue indicates the four key active site residues (S70, K73, S130 and E166).
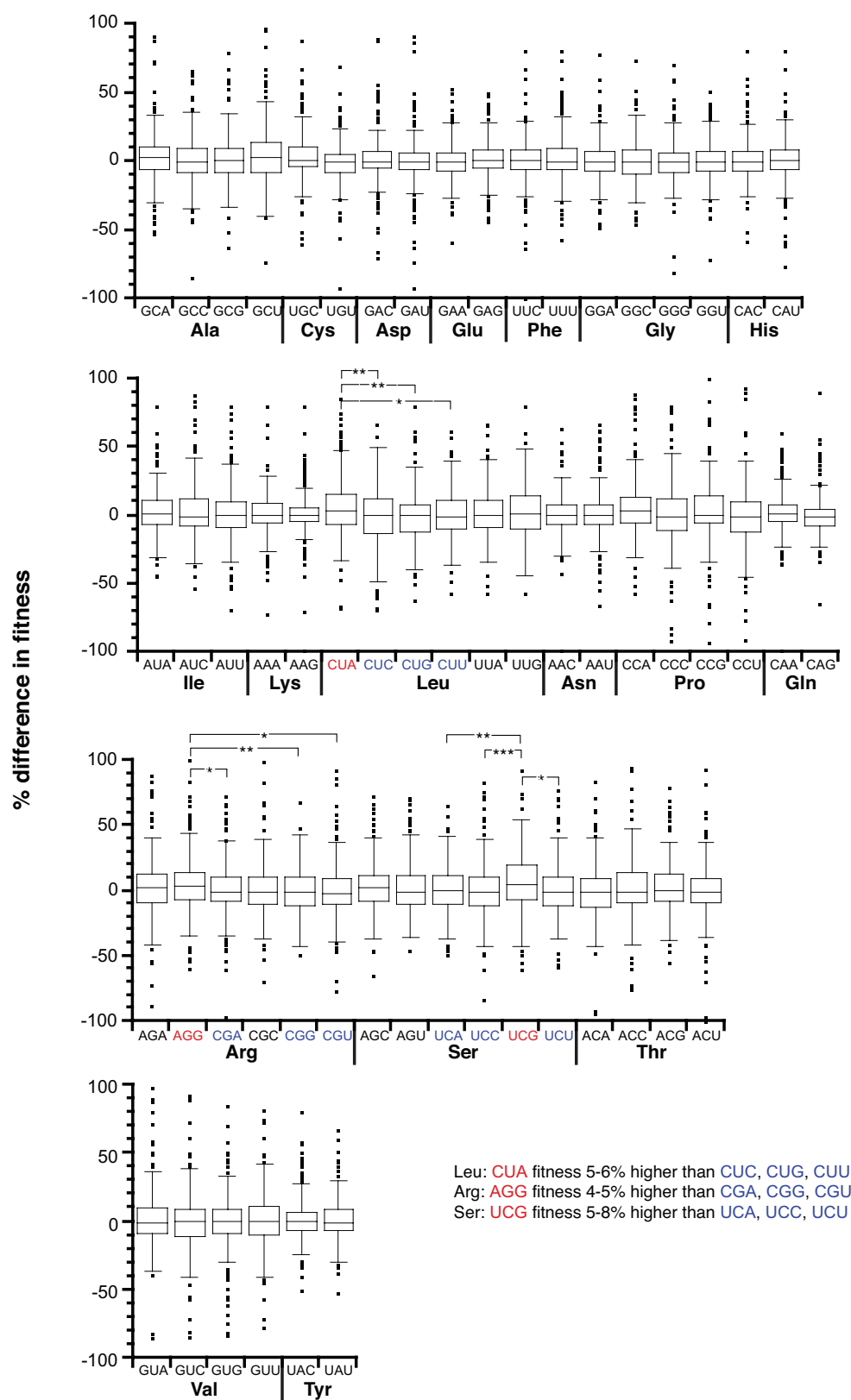
Fig. S5. (legend on next page)

Fig. S5. Global gene fitness effects of codon usage in *TEM-1*. The percent fitness differences of 14,055 synonymous substitutions among the 15,167 alleles with fitness measurements were analyzed as a function of the codon substituted.   The global bias along the entire *TEM-1* gene for any particular codon, if there is any, is on the order of 5% or less. Our analysis for global codon bias is more likely to be able to identify smaller differences in mean fitness for codon sets with a greater number of codons. Thus, although we identify about 5% fitness differences between certain codons for Leu, Arg, and Ser, the fact that these are the three amino acids with six codons made the identification of significant but small differences in these codon sets more likely.   Equal differences may or may not exist within other codon sets but we lack the statistical significance to identify them. *P* values were determined by Student's *t*-test.
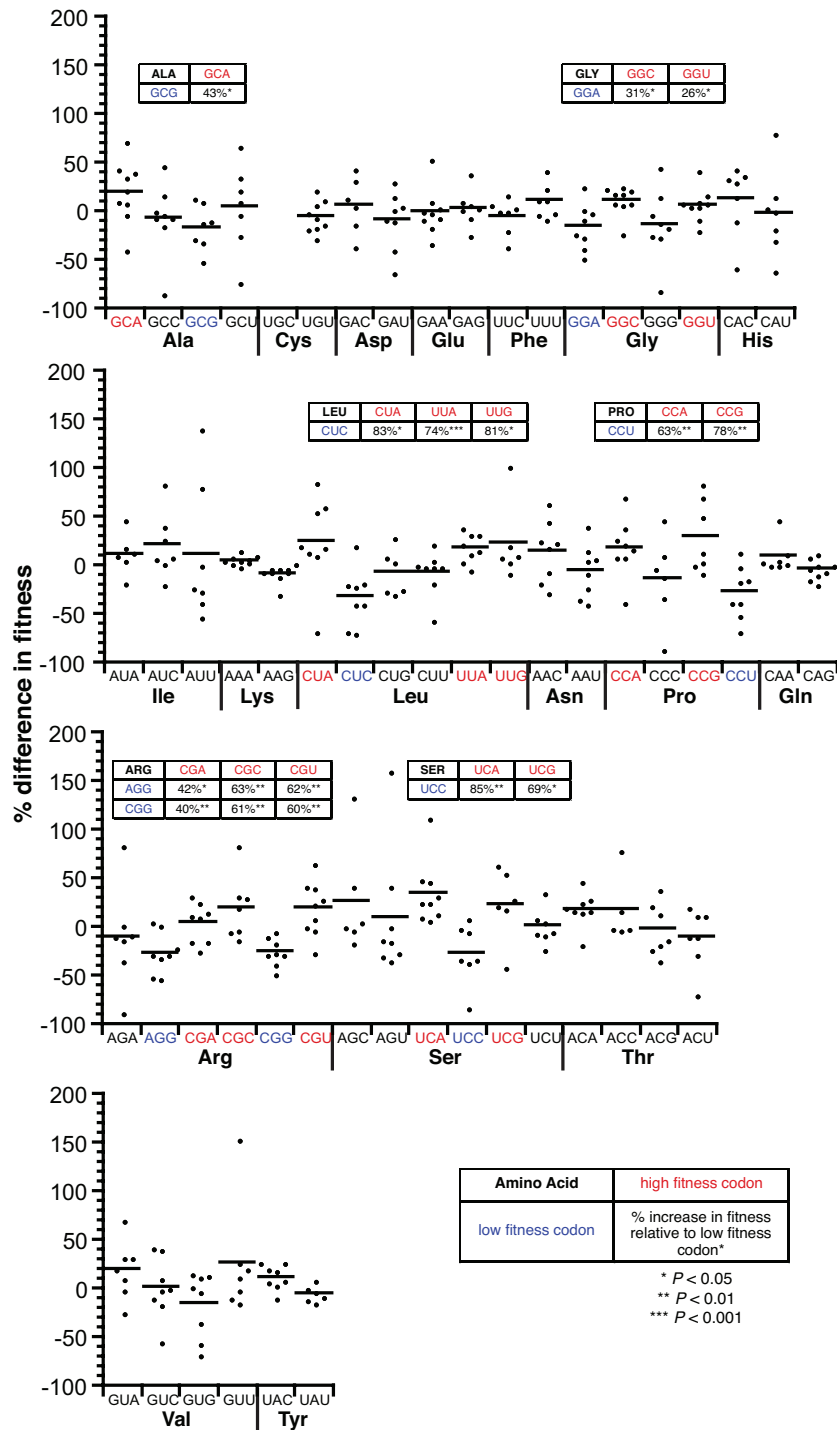
Fig. S6. Gene fitness effects of codon usage at positions 2-10 in *TEM-1*. The percent fitness differences of synonymous substitutions within positions 2-10 in *TEM-1* were analyzed as a function of the codon substituted. *P* values were determined by Student's *t*-test.
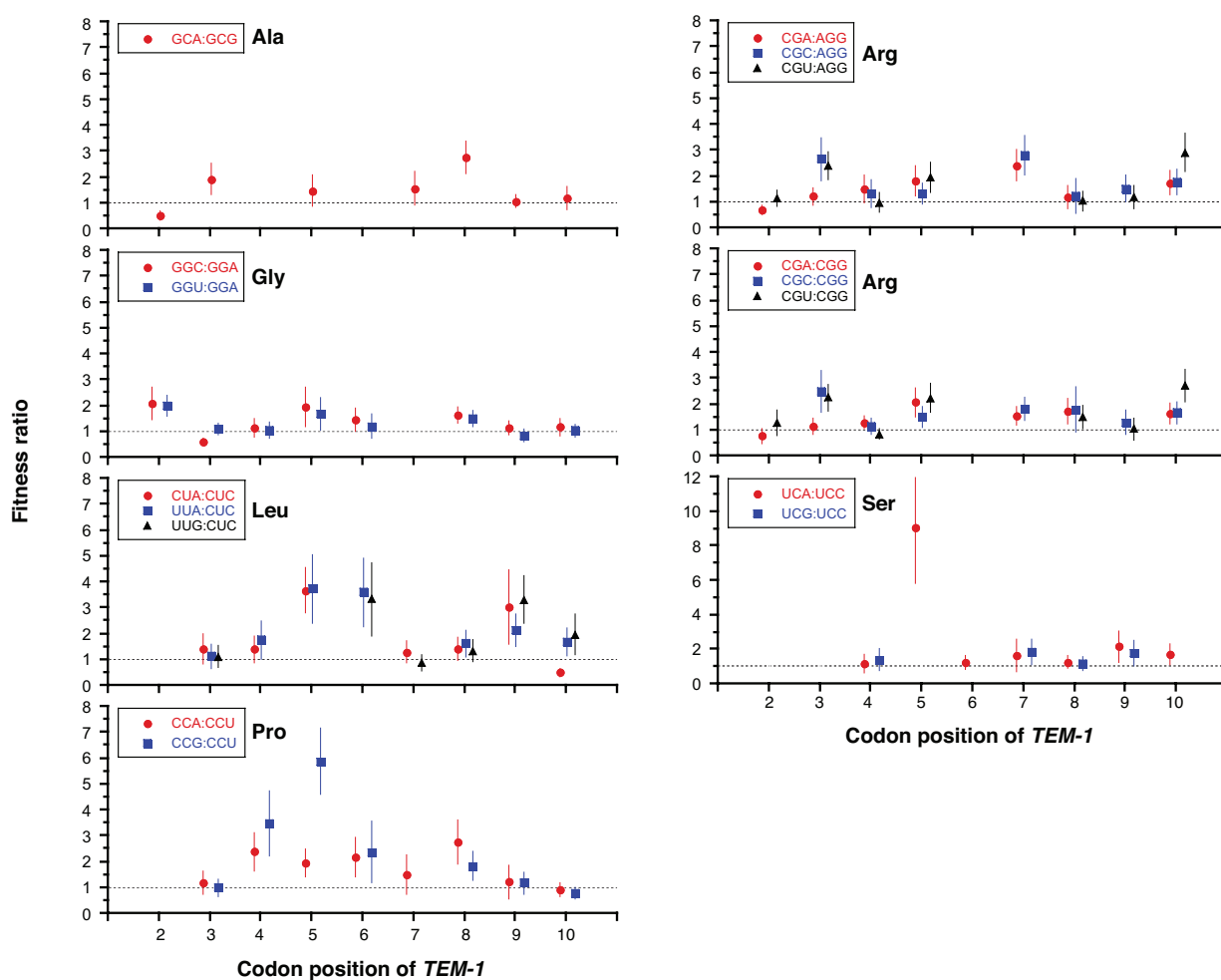
Fig. S7. Positional dependence of synonymous fitness effects at positions 2-10 in *TEM-1*. For the synonymous codon pairs determined to have significant differences in fitness within positions 2-10 of the gene (Fig. S6), the ratio of fitnesses for the two alleles is shown as a function of position in the gene.
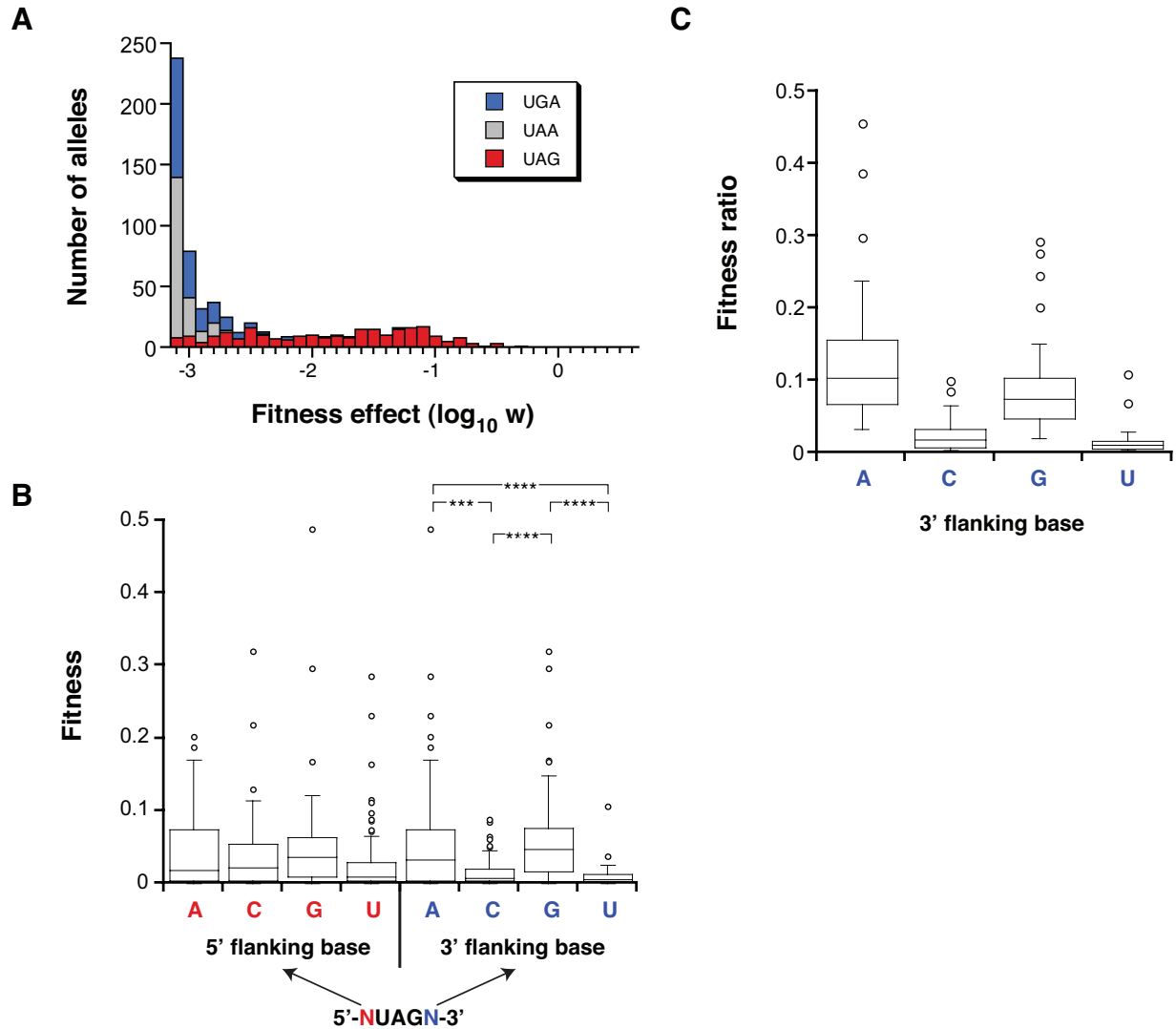
Fig. S8. Gene fitness effects of nonsense mutations in *TEM-1*. (A) The DFE of nonsense mutations as a function of the three nonsense codons. Gene fitness values are presented on a log scale with 0 corresponding to the fitness of *TEM-1*. (B) The efficiency of nonsense suppression at UAG is higher if an A or a G is at the 3' flanking position. *** $P = 0.0002$, **** $P < 0.0001$ by Wilcoxon–Mann–Whitney test. (C) The efficiency of nonsense suppression at UAG is strongly determined by the 3' flanking nucleotide. The fitness ratio compares the fitness of an allele with a mutation to UAG to the fitness of an allele with a missense mutation to glutamine at the same position. Only glutamine missense mutations with $w > 0.25$ were considered. The median efficiencies were 10.4% (3' A) 1.8% (3' C), 7.5% (3' G) and 1.1% (3' U). Differences between A/G and C/U have a P value of <0.0001 (Wilcoxon–Mann–Whitney test).
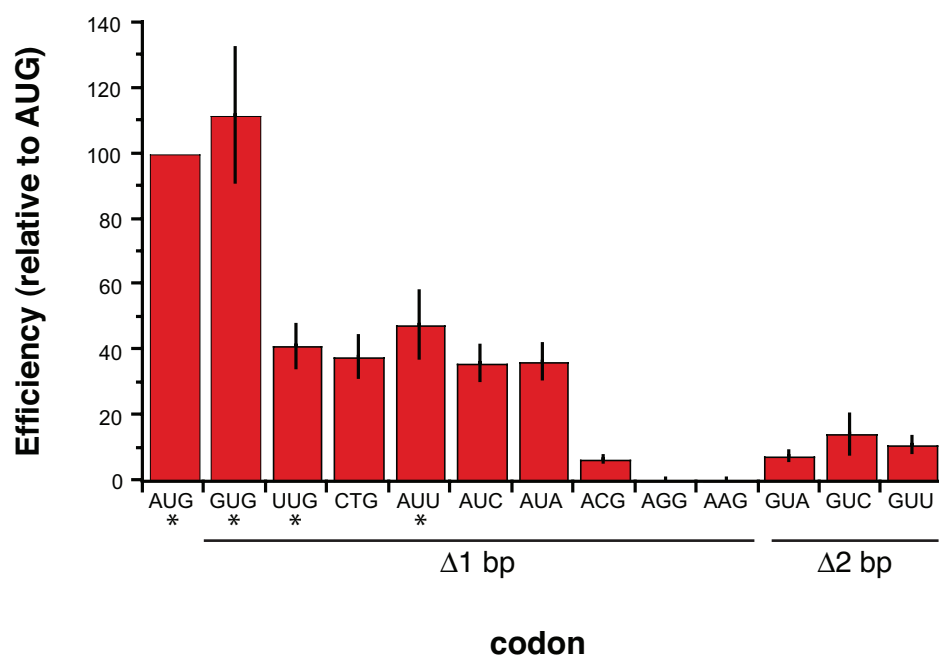
Fig. S9. Relative efficiency at which select codons serve as initiation codons. The efficiency was determined by dividing the gene fitness of the allele with the indicated codon at position 1 in the gene by the fitness of *TEM-1* (i.e. with AUG in position 1). Asterisks indicate known native initiation codons in the *E. coli* genome. All codons that differ from AUG by 1 bp are shown as are the three codons that differ by more than 1 bp that exhibited >1% efficiency.
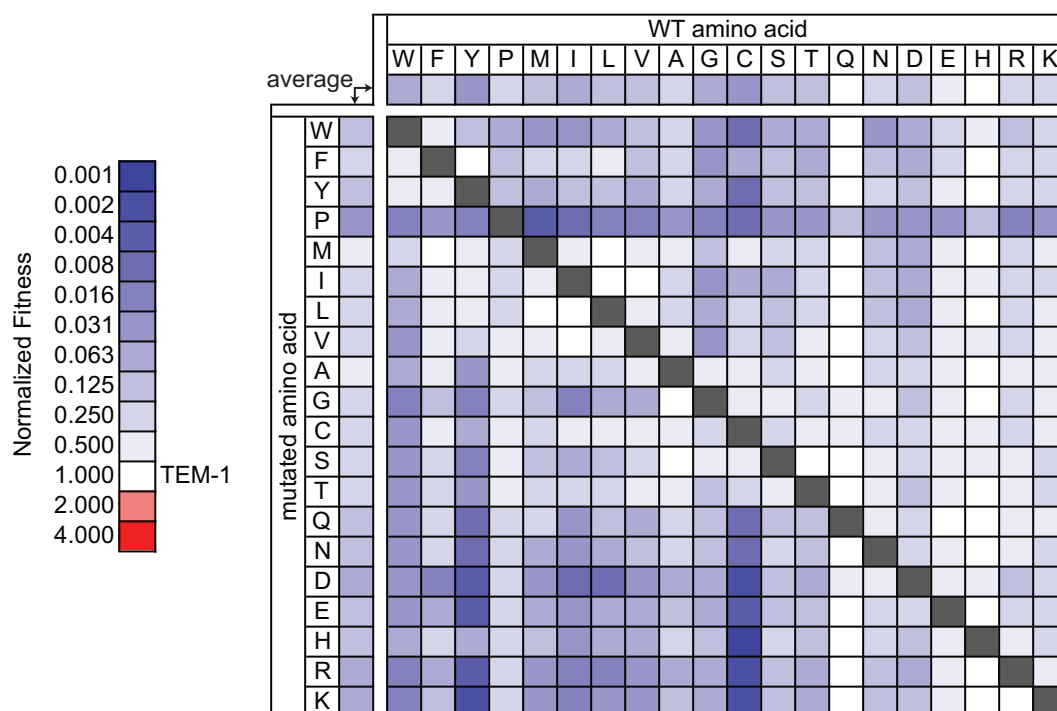
Fig. S10. Amino acid substitution matrix for TEM-1. The heat map indicates the average protein fitness for the indicated substitution. Tabulated data for this heat map is provided in Data S3.
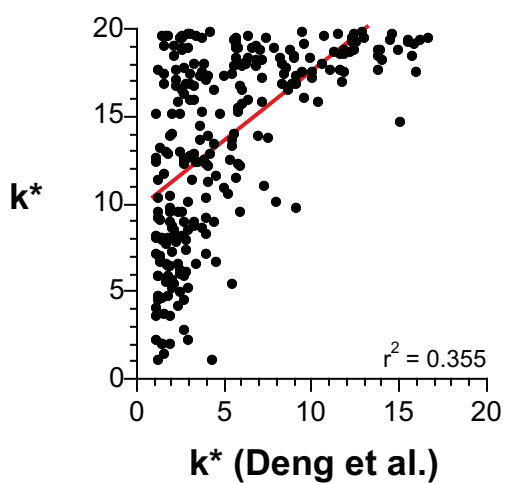

Fig. S11. Comparison of k* of this study with that determined from *TEM-1* alleles with multiple mutations by Deng et al. (Deng, et al. 2012).
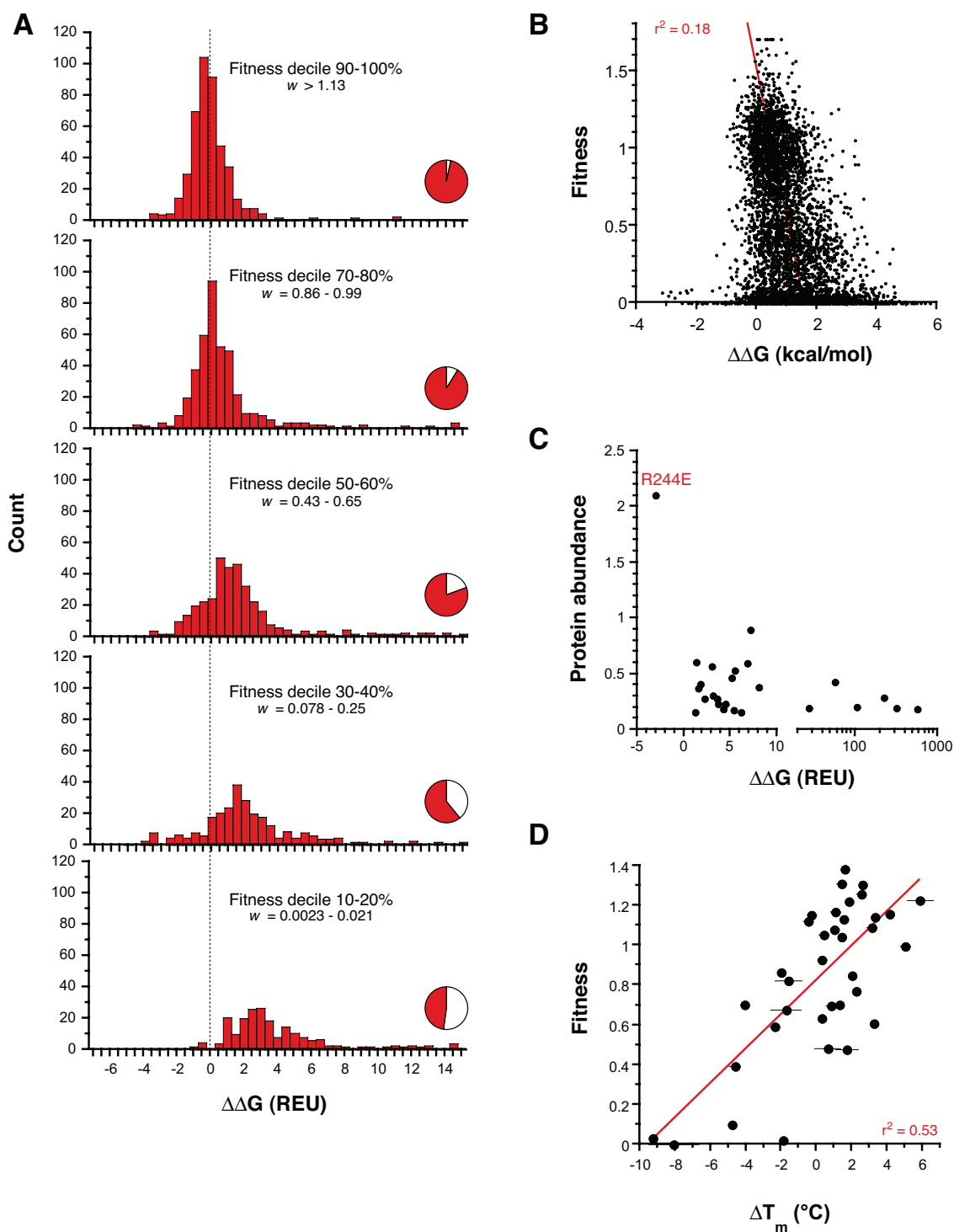
Fig. S12. The correlation between protein fitness and protein stability. (A) Distribution of predicted ΔΔG (by Rosetta) for select fitness deciles of 4783 missense mutations of TEM-1 (i.e. the data of Fig. 5A). In general, proteins with reduced fitness are predicted to have decreased stability. The fitness decile (i.e. 90-100% indicates the fittest 10%)

and the decile's corresponding fitness range are indicated. The pie graphs indicate in red the fraction of ΔΔG values for a decile that are <15 Rosetta energy units (REU). Limitations in the accurate prediction of $\Delta\Delta G$ (Potapov, et al. 2009) including the constraints on backbone movement contribute to the high $\Delta\Delta G$ values of some variants. (B) Protein fitness is shown as a function of change in ΔG as predicted by PoPMuSiC (Dehouck, et al. 2011) for 4783 missense mutations of TEM-1. (C) Protein abundance as a function of predicted ΔΔG (by Rosetta) for 26 randomly selected alleles with low fitness (i.e. the alleles of Fig. S4 with the exception of I13E). Protein abundance is expressed relative to TEM-1. All alleles were predicted to lose thermodynamic stability, the exception being R244E, which had a 2.1 higher protein abundance than TEM-1 that is likely the result of the increase stability. (D) Protein fitness as a function of experimentally measured change in the melting temperature (T_m). We compared the fitness and melting temperature (T_m) of 36 TEM-1 mutants (Bershtein, et al. 2008; Brown, et al. 2010; Deng, et al. 2012; Guillaume, et al. 1997; Kather, et al. 2008; Raquet, et al. 1995; Wang, et al. 2002a, b). The least fit sequences tended to be those few sequences that had lost more than 4 °C in T_m. The dataset is biased towards stabilizing mutations obtained by functional selection that generally had a <2-fold effect on fitness.
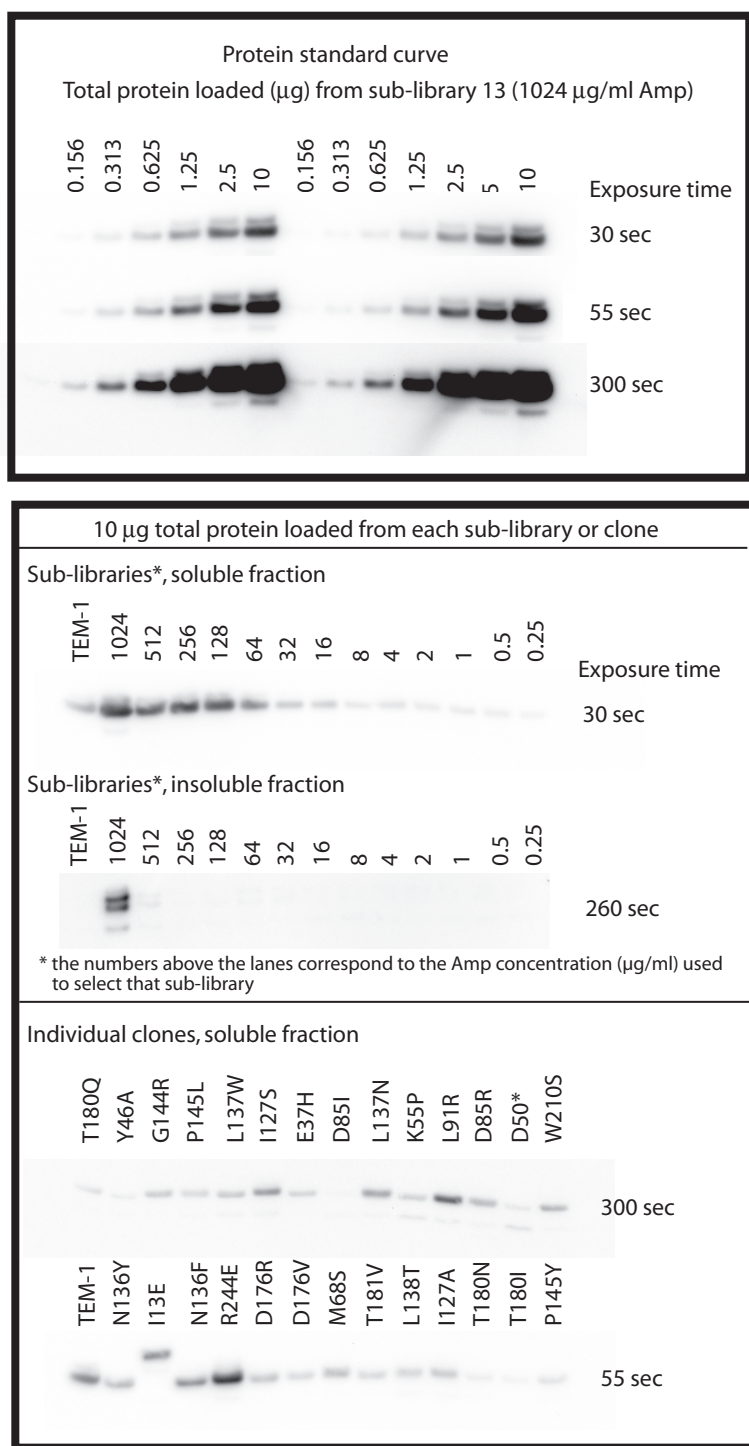
Fig. S13. Representative western blots from protein abundance quantification. The I13E allele (a mutation in the signal sequence) has a higher molecular weight band that corresponds to the size of the protein if the signal sequence has not been removed.
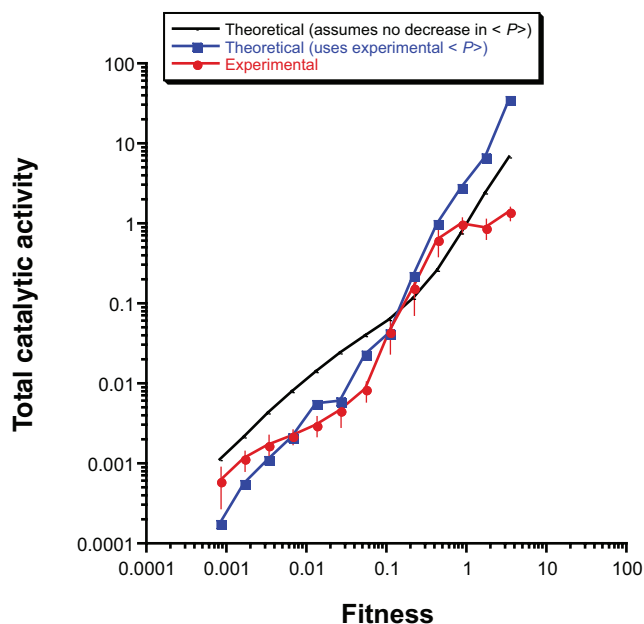
Fig. S14. Expected relationship between protein fitness for Amp resistance and total cellular catalytic activity as measured by nitrocefin hydrolysis. The slight sigmoidal relationship is illustrated by the theoretical calculation that assumes that protein abundance does not change with fitness (black line). By instead using the actual measures of protein abundance (blue squares), the inflection point shifts to lower values and the theoretical curve more closely matches the experimental data (red circles). The calculations generating both theoretical curves are described in Extended Experimental Procedures.
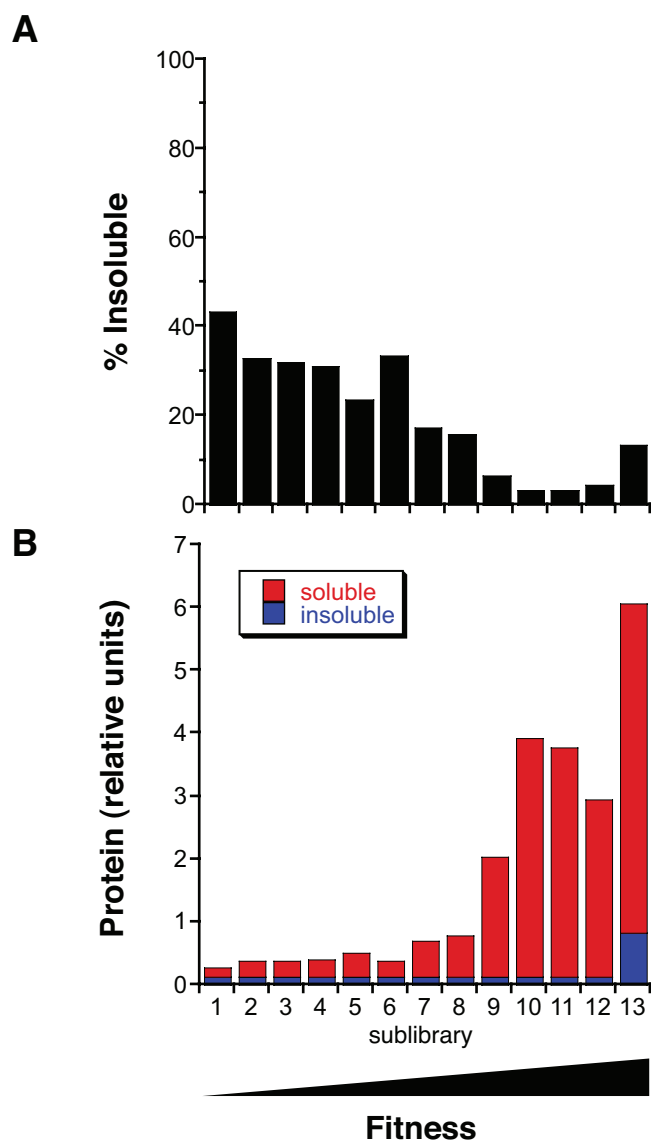
Fig. S15. A decrease in protein fitness is not accompanied by an increase in insoluble TEM-1. (A) The percentage of TEM-1 protein in an insoluble state increased as fitness decreased. (B) However, this trend results from a decrease in the soluble TEM-1 and not from an increase in insoluble TEM-1. A representative western blot of the insoluble fraction is shown in Fig. S13.

A comprehensive, high-resolution map of a gene's fitness landscape

Elad Firnberg[1], Jason W. Labonte[1], Jeffrey J. Gray[1], and Marc Ostermeier[1]

[1]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218 USA

Supplemental Materials and Methods
Supplemental References

SUPPLEMENTAL MATERIALS AND METHODS

Description of the band-pass selection system.

The band-pass genetic selection for β-lactamase activity can select for *E. coli* cells exhibiting any desired level of β-lactamase activity (Sohka, et al. 2009) (Fig. S1).   The two plasmids that comprise the system and the interactions of their components are shown in Supplementary Fig. 1A. In this system, cells with too little β-lactamase activity (relative to the amount of β-lactam antibiotic such as ampicillin) cannot grow due to the β-lactam's inhibitory effect on cell wall synthesis. However, the β-lactam also serves to confer resistance to the antibiotic tetracycline (Tet) via induction of the *ampC* promoter. Cells with too much β-lactamase activity (relative to the amount of β-lactam antibiotic) cannot grow in the presence of Tet since rapid degradation of the β-lactam prevents sufficient induction of TetC expression. Thus cells challenged to grow in the presence of Tet and Amp will grow only if the amount of β-lactamase activity is balanced between having enough activity to degrade the β-lactam to allow cell wall synthesis but not too much to prevent induction of TetC (Fig. S1B). Since increased β-lactam concentration shifts the amount of β-lactamase activity necessary for growth to higher levels, our cells can be used to select for any level of β-lactamase activity (from the absence of activity to full activity) simply by adding the necessary concentration of the β-lactam. A more in depth characterization of this system is provided in Sokha et al. (Sohka, et al. 2009) including the linear relationship between a cell's minimum inhibitory concentration (MIC) of Amp in the absence of Tet and the concentration of Amp in the presence of Tet that provides the best growth for that cell.   This linear relationship holds over three orders of magnitude of Amp concentration. In the experiments reported here, gene fitness corresponds to the Amp concentration providing the best growth in the presence of Tet, which is roughly 25% of the MIC_{Amp}. Although in our system *TEM-1* is under the IPTG

inducible *tac* promoter instead of its native promoter, the level of Amp resistance is similar to that of its native promoter (only 4-fold higher) and very little protein aggregates (Figure S15B).   Thus, the expression of TEM-1 from the *tac* promoter is comparable that from its native promoter and does not have overexpression artifacts.

Gene fitness determination.

The method for measuring fitness makes use of the band-pass genetic selection system (Fig. S1) and is schematically depicted in Fig. S2. Comprehensive codon mutagenesis library CCM2 comprises three separately constructed libraries, one for each third of the gene (Firnberg and Ostermeier 2012). Collectively, they are designed to contain all possible single codon substitutions in the *TEM-1* gene. Each library was plated on LB-agar plates containing 50 µg/ml spectinomycin, 50 µg/ml chloramphenicol, 300 µM IPTG, 20 µg/ml tetracycline and 13 different ampicillin concentrations (2-fold increments of ampicillin ranging from 0.25 µg/ml to 1024 µg/ml) at a cell density of 1700 CFU/cm^2. Plates were incubated at 37°C for 20 hrs. Colonies were recovered from the 39 plates with LB broth, and plasmid DNA isolated using the Qiagen QIAprep Spin Miniprep kit (27106). The plasmid DNA was linearized by restriction endonuclease digestion with SphI and purified using the Zymo DNA Clean & Concentrator kit. Because the plasmid miniprep also contained the band-pass plasmid, pTS42, the concentration of the library plasmid, pSkunk3-CCM2, was determined by running a sample of the linearized DNA on an ethidium bromide agarose gel and analyzing the band intensities of the two respective linearized plasmids. The mass ratio of pTS42:pSkunk3-CCM2 was determined to be ~12.5.

PCR amplicons of each of the 39 sub-libraries were created using Titanium Lib-A "A" and "B" fusion primers that included a 10-base MID barcode identifying the sub-library from which the DNA originated. In order to minimize the rate of recombination or

"PCR-jumping" between DNA template strands in the PCR amplification, experiments were performed to find the minimum amount of template DNA and minimum number of PCR cycles necessary to obtain sufficient PCR product. Each 25 µl PCR reaction had 22.4 pg (5×10^6 molecules pSkunk3-CCM2) linearized template DNA, 0.5 µM each barcoded primer, 200 µM each dNTP, 1X HF Phusion buffer, and 0.5 units Phusion high-fidelity DNA polymerase. Cycler conditions were $98°C$ for 30 sec, 25 cycles of $98°C$ for 10 sec, $61.9°C$ for 15 sec, $72°C$ for 3 min, and then $72°C$ for 5 min. PCR product DNA concentration for each reaction was determined using the Quant-iT Picogreen dsDNA Assay kit (P7589). The barcoded amplicons were then mixed together in molar proportion to the number of colonies that grew on their respective sub-library selection plate. A total of 2.4 µg of the amplicon mixture was electrophoresed on a TAE 0.7% agarose gel and then gel purified using the QIAquick Gel Extraction Kit (28706). Further purification was performed using the Agencourt AMPure XP PCR Purification kit (A63880) to remove short DNA fragments, primers, and primer dimers. The final purified amplicon DNA concentration was determined using picogreen and diluted to 10^9 molecules/µl in 1X TE and then further diluted to 10^7 molecules/µl in DI water. 454 sequencing was performed by Tufts University Core Facility on a Roche 454 GS FLX+ instrument.

We created custom MATLAB scripts to analyze the raw 454 sequencing reads. From a combination of five full or partial plate runs, we obtained a total of 1,325,979 reads. We aligned the reads to the template sequence, sorted the reads by barcode, and mapped all codon substitution mutations, ignoring indels (which are a common sequencing error). Reads were filtered out if the average base call quality score was less than 30, the read did not span the entire mutagenesis region, or if there was more than one codon mutation per read. We obtained 772,296 reads that passed our filtering requirements and had a single codon substitution.

We tabulated the number of sequencing counts for each allele in each sub-library (see Data S1). Since the distribution of growth as a function of Amp is roughly symmetric when plotted as the $\log_2$(Amp concentration) (Sohka, et al. 2009), we determined the unnormalized fitness f of allele i as

$$f_i = \frac{\sum_{p=1}^{13} c_{i,p} \log_2\left(a_p\right)}{\sum_{p=1}^{13} c_{i,p}}$$

Equation 2

in which $c_{i,p}$ is the number of counts of allele i on sub-library plate p in the deep sequencing data and a_p is the concentration of Amp on sub-library plate p in µg/ml . Counts of a particular allele can be expected to appear on 3-4 adjacent sub-library plates (Sohka, et al. 2009), corresponding to an 8-16 fold window of Amp concentration. In a minority of cases however, counts were observed in more than four sub-libraries or counts were observed to cluster in non-adjacent sub-libraries. Two phenomena can account for this: sequencing errors and the presence of an unintended, fitness-altering mutation outside the sequencing region. The frequency of sequencing errors is low relative to that of mutations, since 87% of the library members contain a mutation, and sequencing errors should occur with equal frequency among all sub-libraries. We implemented several measures to facilitate appropriate fitness assignment. First, an allele was assigned a fitness value only if its sequence was observed at least five times. Second, the window of the four adjacent sub-libraries with the highest combined sequencing counts was identified and only these values were included in the fitness calculation. Third, if an allele presented with multiple clusters of counts, the cluster corresponding to the fitness closest to the average fitness of the other synonymous codons was selected. For alleles with mutations in the start codon, knowledge about

alternative *E. coli* start codons was used to assign fitness. For alleles with nonsense

mutations with multiple clusters of counts, the lowest cluster was used to assign fitness.

For the fitness of missense mutations, we combined the sequencing counts of the

corresponding synonymous codons (Data S2) and recalculated the fitness using the

same methods as above for codon substitutions.

We normalized all fitnesses by the fitness of wildtype as follows:

$$w_i = \frac{2^{f_i}}{2^{f_{WT}}}$$ Equation 3

This result is a normalized fitness w_i that is 1.0 for wildtype *TEM-1*, > 1.0 for beneficial

mutations and between 0 and 1.0 for deleterious mutations. We determined the fitness

of wildtype *TEM-1* (f_{WT}) using Eq. 2 using the counts of all alleles with a synonymous

substitution in *TEM-1*, since the fitness of these varied very little. As a check, we

compared this value to the fitness determined by Equation 2 using the counts of all

sequencing reads that lacked a mutation.   The two values differed by only 2.5%. Gene

fitness values are tabulated in Data S1. Protein fitness values were found by averaging

the gene fitness values of synonymous genes and are tabulated in Data S2.

We determined an upper limit on the error in our fitness measurements. We

assumed (solely for the purpose of this error determination) that synonymous mutations

have no fitness effect. We compared an allele's gene fitness to the mean of all alleles

with a synonymous mutation at the same position (i.e. the protein fitness) and expressed

this difference as a percentage of the mean. The distribution of values for this 'percent

difference in fitness' did not vary with fitness, indicating our fitness measurements are

equally accurate at low and high fitness values (Fig. S3A,B), unlike in growth competition

experiments. As expected, the width of the distribution narrowed with the number of

times the allele was observed in the deep sequencing results (the 'allele count') (Fig.

S3C,D). We used the standard deviation of this distribution as a function of allele count

as an estimate of the error in gene fitness.   This error is an upper limit since synonymous mutations can have fitness effects (Plotkin and Kudla 2011).

Prediction of protein thermodynamic stability.

PyRosetta v3.4.0 r55307 (Chaudhury, et al. 2010) was used to compute the difference in score (in Rosetta Energy Units, REU) between the mature structures (lacking the signal sequence) of each amino acid mutant and wild type TEM-1 (Protein Data Bank identifier 1XPB (Fonze, et al. 1995))). The score is designed to capture the change in thermodynamic stability caused by the mutation ($\Delta\Delta G$) (Das and Baker 2008). First, all side chains were repacked (sampling from the 2010 Dunbrack rotamer library (Shapovalov and Dunbrack 2011)) and minimized for the wild type structure using the talaris2013 scoring function. Next, each missense mutation was introduced and all residues within a 10 Å distance of the mutated residue's center were repacked followed by a linear minimization of the backbone and all side chains. This procedure was performed 50 times, and the predicted $\Delta\Delta G$ is the average of the three lowest scoring structures. The PyRosetta script used has been included as part of the standard PyRosetta installation (http://www.pyrosetta.org/dow) and can be found in the apps directory of the PyRosetta installation as delta_score_per_mutation.py. PoPMuSiC predictions of $\Delta\Delta G$ (Fig S12B) were determine online at http://babylone.ulb.ac.be/popmusic (Dehouck, et al. 2011).

Preparation of samples for protein abundance and total catalytic activity assays.

For analysis of the sub-libraries, the three libraries of CCM2 were combined and plated on 13 different Amp concentrations as in the gene fitness measurement experiment described above. Individual clones were plated on the same Amp concentration upon which they were found (i.e. 8 or 16 µg/ml) at the same colony density at which the

original library grew (i.e. the numbers of the colonies on the plates were the same). Colonies were recovered by sweeping with LB broth containing 15 v/v% glycerol and 2 w/v% glucose. Aliquots comprising cultures with equal cell density were pelleted and the supernatant removed. After freezing at -80 ^{o}C for 10 min, the cells were lysed in 250 µL BugBuster Protein Extraction Reagent (Novagen 70584-3) containing ≥0.25 U/µL benzonase nuclease (Sigma E1014-25KU), 3 units/µL rLysozyme (Novagen 71110), and 1% protease inhibitor cocktail (Sigma P8849). Samples were incubated with gentle shaking at 4°C for 30 min, then centrifuged at 14,000 rpm for 30 min at 4°C. The supernatant was recovered as the soluble protein fraction and the pellet resuspended in 250 µL 8M urea as the insoluble fraction. Samples were aliquoted and stored at -80°C. Total protein concentration of each lysate sample was measured using the DC protein assay (Bio-Rad 500-0111) with a BSA standard.

Protein abundance.

Western blots for each sub-library were performed with 10 µg total protein of each lysate sample. A standard curve was prepared by diluting the lysate from sub-library 13 in 2-fold increments, with increasing additions of control lysate from cells not expressing any *TEM-1* allele to maintain a constant amount of 10 µg total protein. SDS-PAGE gels (Novex NP0323BOX) were electrophoresed for 45 min at 190 V, and transferred to a PVDF membrane (Bio-Rad #162-0177) for 30 min at 15 V. The membrane was blocked with a 4% milk solution in PBST (1X PBS, 0.05% tween 20) for 1 hr at room temperature with shaking. The primary anti-TEM-1 mouse monoclonal antibody (Thermo MA1-20370, 500-fold dilution in blocking solution) was incubated for 2 hrs at room temperature or overnight at 4°C with shaking. The membrane was then washed three times with PBST for 5 min with shaking. The secondary goat anti-mouse antibody (Bio-Rad 170-5047, 20,000-fold dilution in blocking buffer) was incubated for 1 hr with shaking at RT. The

membrane was then washed three times with PBST for 5 min with shaking, and a final wash with 1X PBS. A total of 1 ml of chemiluminescence detection reagent (Bio-Rad #170-5070) was then applied to the membrane, incubated for 1 min, and then imaged on a Bio-Rad Gel Doc XR system, recording exposures at 5 sec interval. Representative westerns are shown in Supplementary Fig. 13. Quantity One 1-D analysis software (Bio-Rad) was used to quantify the band intensity for BLA sub-library or clone samples. For each western blot, the image of longest exposure before any band reached detector saturation was selected for quantification. The local adjusted volume parameter was used as the measure of band intensity because it subtracts the local background around each band. Intensity was then converted to relative protein concentration using the standard curve correlation with the same corresponding exposure time.

Total catalytic activity.

Catalytic activity of the sub-libraries and clones was determined by measuring nitrocefin hydrolysis rates. A solution of 50 μM nitrocefin in 10 mM phosphate buffer pH 7.4 was incubated in a 96-well plate (BD Falcon 353072) at 37°C for 2 min. Then 1.0 to 2.5 μL of the soluble fraction of the lysate was added to a final volume of 200 μL and mixed by pipetting up and down briefly. The rate of hydrolysis was measured as the initial slope of absorbance as a function of time as measured at 486 nm at 37°C in a SpectraMax Plus 384 Microplate reader. The initial rate was normalized by the total amount of protein added for each sample.

Theoretical calculation of total catalytic activity vs. fitness.

Allele fitness results from the ability to hydrolyze Amp at the concentration of Amp present on the plate.   Our experimental measure of total catalytic activity uses the initial rate of nitrocefin hydrolysis at 50 μM nitrocefin. These will not correlate precisely 1:1

owing to the differences in catalytic constants for hydrolysis of the two substrates and the fact that nitrocefin hydrolysis is measured at a set nitrocefin concentration but fitness is evaluated at different Amp concentrations. However, we can predict the form of the relationship. We assumed Michaelis-Menten kinetics and the following values for k_{cat} and K_m for Amp (1187 s^{-1} and 42.7 µM) and nitrocefin (917 s^{-1} and 128 µM), which are average values for TEM-1 of that measured in several previous studies. We first calculated the reduction of total Amp hydrolysis activity expected for the expected mean fitness of the sub-library. We then assumed that this reduction in catalytic activity comes from equal percent reductions in the k_{cat} and K_m. We then assumed that this same percent reduction in the catalytic constants will occur for nitrocefin hydrolysis. We then calculated the predicted relative initial rate of nitrocefin hydrolysis at 50 µM nitrocefin using the Michaelis-Menten equation. We performed these calculations (a) assuming the protein abundance was the same for all fitness values and (b) using the experimentally observed measurements of protein abundance shown in Fig. 5B. The results are presented as Fig. S14.

SUPPLEMENTAL REFERENCES

Bershtein S, Goldin K, Tawfik DS. 2008. Intense neutral drifts yield robust and evolvable consensus proteins. J Mol Biol 379: 1029-1044.

Brown NG, Pennington JM, Huang W, Ayvaz T, Palzkill T. 2010. Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum TEM beta-lactamases. J Mol Biol 404: 832-846.

Chaudhury S, Lyskov S, Gray JJ. 2010. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics 26: 689-691.

Das R, Baker D. 2008. Macromolecular modeling with rosetta. Annu Rev Biochem 77: 363-382.

Dehouck Y, Kwasigroch JM, Gilis D, Rooman M. 2011. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. BMC Bioinformatics 12: 151.

Deng Z, Huang W, Bakkalbasi E, Brown NG, Adamski CJ, Rice K, Muzny D, Gibbs RA, Palzkill T. 2012. Deep sequencing of systematic combinatorial libraries reveals beta-lactamase sequence constraints at high resolution. J Mol Biol 424: 150-167.

Dietz H, Pfeifle D, Wiedemann B. 1997. The signal molecule for beta-lactamase induction in Enterobacter cloacae is the anhydromuramyl-pentapeptide. Antimicrob Agents Chemother 41: 2113-2120.

Firnberg E, Ostermeier M. 2012. PFunkel: efficient, expansive, user-defined mutagenesis. PLoS ONE 7: e52031.

Fonze E, Charlier P, To'th Y, Vermeire M, Raquet X, Dubus A, Frere JM. 1995. TEM1 beta-lactamase structure solved by molecular replacement and refined structure of the S235A mutant. Acta Crystallogr D Biol Crystallogr 51: 682-694.

Guillaume G, Vanhove M, Lamotte-Brasseur J, Ledent P, Jamin M, Joris B, Frere JM. 1997. Site-directed mutagenesis of glutamate 166 in two beta-lactamases. Kinetic and molecular modeling studies. J Biol Chem 272: 5438-5444.

Kather I, Jakob RP, Dobbek H, Schmid FX. 2008. Increased folding stability of TEM-1 beta-lactamase by in vitro selection. J Mol Biol 383: 238-251.

Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet 12: 32-42.

Potapov V, Cohen M, Schreiber G. 2009. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. Protein Eng Des Sel 22: 553-560.

Raquet X, Vanhove M, Lamotte-Brasseur J, Goussard S, Courvalin P, Frere JM. 1995. Stability of TEM beta-lactamase mutants hydrolyzing third generation cephalosporins. Proteins 23: 63-72.

Shapovalov MV, Dunbrack RL, Jr. 2011. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure 19: 844-858.

Sohka T, Heins RA, Phelan RM, Greisler JM, Townsend CA, Ostermeier M. 2009. An externally-tunable bacterial band-pass filter. Proc Natl Acad Sci USA 106: 10135-10140.

Valtonen SJ, Kurittu JS, Karp MT. 2002. A luminescent Escherichia coli biosensor for the high throughput detection of beta-lactams. J Biomol Screen 7: 127-134.

Wang X, Minasov G, Shoichet BK. 2002a. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J Mol Biol 320: 85-95.

Wang X, Minasov G, Shoichet BK. 2002b. Noncovalent interaction energies in covalent complexes: TEM-1 beta-lactamase and beta-lactams. Proteins 47: 86-96.