Variation in Mutational Robustness between Different Proteins and the Predictability of Fitness Effects

Peter A. Lind,1,2 Lars Arvidsson,1 Otto G. Berg,3 and Dan I. Andersson*,1
1Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden
2Department of Molecular Biology, Umeå University, Umeå, Sweden
3Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

*Corresponding author: E-mail: dan.andersson@imbim.uu.se
Associate editor: Csaba Pal

Abstract
Random mutations in genes from disparate protein classes may have different distributions of fitness effects (DFEs) depending on different structural, functional, and evolutionary constraints. We measured the fitness effects of 156 single mutations in the genes encoding AraC (transcription factor), AraD (enzyme), and AraE (transporter) used for bacterial growth on L-arabinose. Despite their different molecular functions these genes all had bimodal DFEs with most mutations either being neutral or strongly deleterious, providing a general expectation for the DFE. This contrasts with the unimodal DFEs previously obtained for ribosomal protein genes where most mutations were slightly deleterious. Based on theoretical considerations, we suggest that the 33-fold higher average mutational robustness of ribosomal proteins is due to stronger selection for reduced costs of translational and transcriptional errors. Whereas the large majority of synonymous mutations were deleterious for ribosomal proteins genes, no fitness effects could be detected for the AraCDE genes. Four mutations in AraC and AraE increased fitness, suggesting that slightly advantageous mutations make up a significant fraction of the DFE, but that they often escape detection due to the limited sensitivity of commonly used fitness assays. We show that the fitness effects of amino acid substitutions can be predicted based on evolutionary conservation, but those weakly deleterious mutations are less reliably detected. This suggests that large-effect mutations and the fraction of highly deleterious mutations can be computationally predicted, but that experiments are required to characterize the DFE close to neutrality, where many mutations ultimately fixed in a population will occur.

Key words: mutation, fitness, protein, robustness, arabinose operon, bacteria.

Introduction
The distribution of fitness effects (DFE) of random mutations for a whole genome or individual genes is a key factor in determining how evolution progresses in response to mutation, selection, and drift. Fitness constraints can be present at many levels, including protein and mRNA expression levels, protein stability, and structure as well as interactions with other cellular components, which makes prediction of the DFE extremely complex (Travisano and Shaw 2013; de Visser and Krug 2014). Several methods have been used to gain insight into the parameters of the DFE, including mutation accumulation experiments, fitness measurements of single induced mutations, and DNA sequence data (Frey-Walker and Keightley 2007; Bataillon and Bailey 2014). Distributions inferred from sequence analysis concern mutational effects small enough that SNPs can be detected in populations whereas experimental approaches can only detect mutations with relatively large effects, with the limit set by the sensitivity of the fitness assay used. Whereas many studies focus on the structure–function relationship and phenotypic effects of mutations (Suckow et al. 1996; Jacquier et al. 2013; Jimenez et al. 2013; Finnberg et al. 2014; Sarkisyan et al. 2016), sometimes providing data for thousands of mutations, the evolutionary relevance of such mutational landscapes is uncertain. For example, Kondrashov and co-workers (Sarkisyan et al. 2016) measured the effects of mutations in Green Fluorescent Protein (GFP). In this case, the gene was moved from its original organism, the jellyfish Aequorea victoria, into E. coli and mutated, codon optimized, fused to another protein, and expressed from a plasmid at high level. “Gene fitness” was then measured with low sensitivity as fluorescence, with excitation at a wavelength not primarily selected for in its original organism, in an external environment and temperature very different from where the gene had evolved, and with no connection between the function of the gene and cell fitness. Whereas this and similar studies certainly contribute to understanding constraints on the molecular level, their relevance for evolution or for organismal fitness landscapes is debatable.

Relatively few studies have directly measured the fitness effects of random single mutations in protein-coding genes in their native context with high sensitivity (Lind et al. 2010; Sanjuan 2010; Hietpas et al. 2011) and provided explanations for which factors influence the shape of DFEs. We previously reported that two ribosomal protein (RP) genes (Lind et al. 2010) had unimodal DFEs peaked at slightly deleterious
mutations, with an average fitness effect of $s = -0.012$, no complete loss of function, and very few neutral mutations. This finding contrasts with distributions reported for viruses, $\beta$-lactamase, and GFP where a high fraction of complete loss-of-function mutations was observed (Sanjuan 2010; Hietpas et al. 2011; Jacquier et al. 2013; Firnberg et al. 2014; Sarkisyan et al. 2016). To address these divergent results and their underlying evolutionary rationale, we used highly sensitive competition experiments to measure the DFE of random mutations for three genes from different protein classes in their native context to investigate if the high robustness of RPs is an exception and if any general properties of the DFE can be established. Specifically we wanted to determine the fraction of deleterious, neutral and advantageous mutations, investigate the molecular basis for mutational effects and the potential for in silico prediction, and examine if there is variation depending on differences in protein function.

Results and Discussion

Experimental System

Three genes in *Salmonella typhimurium* LT2, required for rapid growth on L-arabinose, were chosen for random mutagenesis based on their different molecular functions, conditional dispensability, and large fitness costs in a selective environment, i.e., growth on arabinose as the only carbon and energy source, when disrupted (fig. 1A). We introduced 156 single base pair substitutions, excluding stop codons, into *araC* ($n = 55$), *araD* ($n = 47$), and *araE* ($n = 54$) in their native chromosomal position in *S. typhimurium* and estimated their fitness ($s$) in one-to-one competition with wild-type controls (fig. 1B) in the presence of arabinose minimal medium. This set-up allowed detection of fitness differences of $|s| > 0.004$. For *araD* the mutations were introduced by mutagenic primers where the position and type of mutation were randomized with uniform probabilities. For *araC* and *araD*, error-prone PCR was used to generate mutations and the mutational spectrum is biased towards transitions (supplementary table S4, Supplementary Material online). Mutants with stop codons were excluded for both methods.

*araD* is a homotetrameric epimerase that catalyzes the final step in the degradation pathway for L-arabinose to D-xylulose-5-phosphate using a $\text{Zn}^{2+}$ as cofactor. *araC* is a transcription factor that binds L-arabinose and activates transcription of the *araBAD* operon and *araE* in the absence of glucose and represses its own transcription and that of *araBAD* in the absence of arabinose (Schleif 2010). *araC* belongs to the *araC/XylS* family of transcriptional regulators that are widespread among bacteria and are involved in regulation of carbon metabolism, stress response, pathogenesis, and antibiotic resistance (Gallegos et al. 1997). *araE* is an arabinose/proton symporter that transports L-arabinose across the inner membrane (Daruwalla et al. 1981). It belongs to the major facilitator (MFS) superfamily of importers and exporters, which are present in all domains of life and have been implicated in antimicrobial resistance and in human diseases including cancer and type 2 diabetes (Quistgaard et al. 2016).

The *AraCDE* proteins represent three major functional classes, transcription factor, enzyme, and transporter, which are expected to have different biophysical constraints. Although all these proteins need to recognize and interact with a small molecule to perform their function there are

![Fig. 1. Experimental system and methods. (A) Three genes (*araC*, *araD* and *araE*) encoding proteins required for rapid growth on L-arabinose were chosen for random mutagenesis. AraE transports L-arabinose into the cell, AraC binds L-arabinose and activates transcription of *araE*, *araC*, and *araBAD*. AraD catalyzes the final step in the degradation pathway for L-arabinose to D-xylulose-5-phosphate, which then enters the pentose phosphate pathway. (B) The change in ratio between mutant and wild-type populations is determined after growth competition under a serial transfer protocol. The fitness of each mutant relative to the wild-type is calculated as a selection coefficient ($s$) per generation that is 0 for neutral, negative for deleterious and positive for advantageous mutations under growth in minimal medium with arabinose.](https://academic.oup.com/mbe/article-abstract/34/2/408/2528248)
additional constraints in interactions with membrane components for AraE, DNA binding and RNA polymerase interaction for AraC and oligomerization for AraD and AraC. There are also differences in system level properties, where AraC regulates transcription of the entire system, the level of AraE determines the flow of L-arabinose across the membrane and the production of D-xylulose-5-phosphate by AraD is the final biochemical determinant for flow into the pentose phosphate pathway (fig. 1A).

The fitness effects of a complete loss-of-function mutation in each of the three genes were determined by examining growth on arabinose when the respective gene was deleted. Loss of AraD is lethal when grown on L-arabinose due to accumulation of a toxic intermediate (Englesberg et al. 1962) and $s$ for $\Delta araD$ was set to −1. Deletion of AraC results in loss of activation of transcription for $araBAD$ and $araE$ ($s = -0.36$). As AraE is not the only way for the cells to take up L-arabinose, deletion of this gene had less dramatic consequences on fitness ($s = -0.25$).

Synonymous Mutations in $araC$, $araD$, and $araE$ Do Not Reduce Fitness

Synonymous substitutions accounted for 47 of the 156 single base pair substitutions and all of them were classified as “neutral”, which here means they cannot be reliably distinguished from wild-type controls ($|s| < 0.004$, in the shaded area in fig. 2A–C, full details in supplementary table S1, Supplementary Material online). The mean cost for all synonymous substitutions was $s = 2 \times 10^{-5}$ (SD = 0.0014), suggesting that their effects are very close to neutrality on average and the frequency of non-neutral synonymous mutations is likely to be below 6% ($P = 0.05$ hypergeometric test, assuming random sampling and correct classification). In contrast, for the RP genes only two out of 38 synonymous substitutions were “neutral” and the average cost was $s = -0.0096$ (SD = 0.0064). This difference could be explained by higher expression levels of RPs (Bakshi et al. 2012), more optimized codon usage and that a minor reduction in the amount of an RP reduces growth to a significant degree (Knöppel et al. 2016). Firnberg et al. (2014) also reported significant effects of single synonymous mutations on TEM-1 activity when expressed from a tac promoter, suggesting that these effects might vary substantially depending on gene, expression level, and selective conditions.

The $AraCDE$ Proteins Are Less Robust than Ribosomal Proteins, but Have a Higher Fraction of Advantageous Mutations

For the amino acid substitution mutants, the fractions of complete loss-of-function mutants were similar for AraD (7/36) and AraC (9/38) whereas only one of the AraE mutants was close in fitness to the deletion mutant (fig. 2D). Deleterious mutants that still maintained some function made up a similar fraction as the complete loss-of-function mutants for AraD (9/36) and AraC (15/38), whereas they were more common for AraE (20/38) (fig. 2D).

Amino acid substitution mutants that were indistinguishable from wild-type fitness ($|s| < 0.004$) made up 33–56% of all mutations (“neutral” in fig. 2D). This provides an upper limit for the fraction of mutants that are nearly neutral in a population genetics sense with fitness effects smaller than the inverse of the effective population size, which for many bacteria would be in the order of $|s| < 10^{-5}$ to $10^{-9}$ (Berg 1996;
Charlesworth 2009). We also found mutants with significantly higher fitness, one in AraC (M164T) and three in AraE (T338I, V31I, R257H) (two-sided t-tests $P < 10^{-4}$) (fig. 2A and C; supplementary table S2, Supplementary Material online). This suggest that several percent of random mutations might be advantageous when the mutated genes are under strong selection, but the few mutations found here makes estimates of frequency uncertain ($f = 2.6\%$, $f > 0.9\%$, $P = 0.05$, hypergeometric test) assuming that our sampling is random and that mutations are classified correctly as advantageous or not. This frequency is similar to that reported for mutations in TEM-1 ß-lactamase that increased resistance to cefotaxime and increased fitness under selection for function (3.4%) (Schenk et al. 2012). The fitness increases for the mutations in araC and araE were relatively small ($s = 0.0061–0.019$) and would not have been detected using less sensitive fitness assays, but as most mutations are expected from theory to have small effects (Fisher 1930) this again highlights the importance of investigating the DFE close to neutrality. Similarly high rates of advantageous mutations with small effects have previously been reported (Desai et al. 2007; Perfeito et al. 2007), but they are often not detected in experimental evolution studies where they are lost through population bottlenecks and competition with rare high fitness mutants (Sniegowski and Gerrish 2010).

Whereas 95% of random substitutions in the genes of RP S20 and RP L1 were weakly deleterious this is expected to be largely due to constraints on the mRNA level given that synonymous substitutions had effects of similar magnitude (Lind et al. 2010). If the average effect of the synonymous mutations in RPs is subtracted from the nonsynonymous, the average cost is only $s = -0.0037$ (fig. 3A) whereas nonsynonymous substitutions in AraCDE reduced fitness with $s = -0.123$ on average (fig. 3B), generating an approximately 33-fold difference in average robustness between RPs and AraCDE. Nonsynonymous mutations in ribosomal proteins never reduced fitness more than $s = -0.1$, even though deletion of the genes resulted in much larger fitness costs (estimated maximum frequency for all nonsynonymous mutations 3.3% $P = 0.05$, hypergeometric test) (Lind et al. 2010). For the AraCDE protein, these highly deleterious mutations represented 27% of amino acid substitution mutants (95% confidence interval for all nonsynonymous mutations 23–31%, assuming random sampling, and correct classification) (fig. 3A and B).

**Selection to Reduce Costs of Misreading Leads to High Robustness for Ribosomal Proteins**

What could be the evolutionary reason for the approximately 30-fold higher robustness of ribosomal proteins as compared with the AraCDE proteins? One possibility is that genes central to growth, such as RPs, might evolve robustness to reduce the costs of deleterious mutations, but based on the calculation outlined below we argue that this is unlikely. The fitness of a population under the pressure of random mutations [rate $u \approx 10^{-10}$ (Foster et al. 2015)] is given by $e^{-uM}$, where $M$ is the number of positions in the genome where a mutation leads to a reduction in fitness (Kimura and Maruyama 1966). The increase in population fitness if one of the deleterious positions were transformed into a neutral one, would be $\Delta s = e^{-uM} \approx u$. This is an extremely small selective advantage that could be effective only in a population that has a very large effective size, i.e., for $N_e > 10^{10}$ or when the mutation rate is very high. The effective population size for many bacteria is usually estimated to be smaller than the inverse of the mutation rate ($N_e \leq 10^9$) (Berg 1996; Charlesworth 2009), which means that mutation pressure alone would have an extremely small effect on the selection for an increased mutational robustness at each individual site. If we allow the removal of one deleterious position to also reduce the cost of mutations in $i$ other positions, the fitness gain for the population is $\Delta s = iue^{-uM} \approx iu$, which is still very small unless the number of other stabilized sites are exceedingly large. Thus, although the effect of a mutation in an
individual cell can be large as indicated by the DFE, the effect of deleterious mutations on the fitness of the whole population is exceedingly small and selection on the mutation pressure is unlikely to shape the DFE.

In contrast, transcription/translation errors will occur at a much higher rate \( p \approx 10^{-3} \) to \( 10^{-4} \) per amino acid; Drummond and Wilke 2009; Meyerovich et al. 2010; Imashimitsu et al. 2013) in all cells and the effect on the population fitness can be large. We can estimate the effects of misreading errors on fitness in the following simplified model. Consider a protein of size \( n \) amino acids that is present in \( m \) copies in the cell. When all copies of the protein carry the same mistake (like a mutation), the fitness effect is \( s \) as measured in the DFE. Assume that the effects are additive so that when only one copy of the protein carries the error, the effect on growth is \( s/m \). Assume further that the average effect of all possible missense errors in the protein is the same as the average of the DFE, i.e., \( s/m \) per copy, which would be the case if most translation errors are due to single base-pair misreading. The expected number of errors at all codons and in all copies of the protein is \( n \times p \times m \). Thus, the overall average effect on fitness from missense errors in this protein is

\[
\Delta s = n \times p \times s.
\]

For a ribosome with \( n \approx 7500 \) (Nikolay et al. 2001), \( s = -0.0037 \), and assuming \( p = 2 \times 10^{-4} \), this gives \( \Delta s = -0.0056 \). The fitness cost of misreading errors is similar for a protein of average size, with mutational robustness like the AraCDE proteins \( \Delta s = -0.0072 \) assuming \( n \approx 300 \) (Brocchieri and Karlin 2005) and \( s = -0.12 \). However, if the early primitive ribosomes had the same DFE as an average protein under strong selection (exemplified by the AraCDE proteins with \( s = -0.12 \)), then fitness would be greatly reduced with \( \Delta s = -0.18 \). Thus, for large protein complexes (like ribosomes), the selective advantage of shifting the average of the fitness spectrum to smaller values, thereby increasing robustness, would be very large and could explain the high robustness of modern ribosomes.

### Inability of Mutations in Ribosomal Proteins to Disrupt Intermolecular Interactions Provides a Mechanism for Mutational Robustness

What could be the mechanistic basis for the experimentally observed difference in robustness between AraCDE versus RP S20 and L1? A structure-function analysis of the AraCDE proteins using homology models (supplementary fig. S1, Supplementary Materials online) suggests that mutations with large deleterious effects appear to disrupt specific functions (supplementary table S2, Supplementary Material online) rather than being generally destabilizing. For the AraD protein, conditional lethal mutations are in the direct proximity of the active site (L16R, G172C, and H118Y), including Zn\(^{2+}\) binding residues, or in the C-terminal end (Y229N and G230R) that is directed across the active site of the adjacent subunit (Luo et al. 2001; Schleif 2010). Although AraC and AraE lack enzymatic active sites, a similar pattern emerges where highly deleterious mutations in AraE appear mainly in the transport channel (A46V, D39H, L155P, T158R, and Y131C), disrupt salt bridges (P202S), or are close to residues involved in substrate recognition (G273V, S400L, and W405C) (Sun et al. 2012; lancu et al. 2013). In AraC, highly deleterious mutations are near residues implicated in arabinose binding (I57V, T50I, G55D, and R84Q), DNA binding (S207L and R178C), interaction with RNA polymerase (G250E, A191S, and D256V), or dimerization (L160S, L108P, L161F, Y103H, and P39L) (supplementary fig. S1, Supplementary Material online) (Soisson et al. 1997; Schleif 2010).

In contrast to the AraCDE proteins many RPs primarily serve a structural function, which could allow them to be stabilized by numerous interactions with rRNA. Specific functions are difficult to assign to individual ribosomal proteins, but RP S20 and L1 do not interact directly with mRNA or tRNA or form part of the peptidyl transferase center and their roles might be described as mainly structural. S20 is a primary RNA binding protein that initiates proper folding of the 5S domain of the 23S rRNA and is crucial for binding of ribosomal protein S16. S20 is crucial for anchoring rRNA helix 44, which is part of the A- and P sites (Schluenzen et al. 2000) and forms bridging contacts with the 50S subunit (Selmer et al. 2006). Deletion of RP S20 leads to reduced mRNA binding and inhibits association with the 50S subunit (Tobin et al. 2010). RP L1 binds directly to 23S rRNA and is not known to influence the binding of other ribosomal proteins. It is located on a flexible stalk region that is believed to be involved in structural rearrangements that is coupled to movement of decacylated tRNA in and out of the E-site (Ning et al. 2014). Deletion of L1 leads to aberrant ribosome biogenesis and L1 is required for balanced formation of subunits (Kährström et al. 2012). Mutational robustness of RPs might be increased by introduction of specific amino acids at the protein–RNA interface (Jones et al. 2001) so that function may not be significantly disrupted by single amino acid substitutions. In addition, strong selection for proper folding and ribosome assembly of proteins containing amino acid substitutions could potentially lead to the recruitment of molecular chaperones (Fares 2015) and assembly factors (Shajani et al. 2011).

An alternative explanation for the increased robustness of RPs is to evolve a codon usage that avoids single mutations that lead to amino acid substitutions with large effects on protein structure. Our data does not support this hypothesis based on similar distribution of values from substitution matrices for the mutations in RPs compared with the AraCDE proteins (supplementary fig. S2A and B, Supplementary Material online). The predicted effects of mutations on changes in thermodynamic stability between the two classes are also similar (supplementary fig. S2C and D, Supplementary Material online). However, it is possible that ribosomal proteins have evolved robustness by increasing their stability threshold by selection for generally stabilizing mutations (Kather et al. 2008).

### Predicting Mutational Effects of Amino Acid Substitutions on Fitness

Prediction of the DFE from sequence or structural data could play a major part in evolutionary forecasting efforts and complement experimental studies. However, it is not clear to what degree fitness effects are predictable, which methods would...
be most successful and how this depends on the functional class the protein belongs to. One fast and readily available method is to use amino acid sequence conservation in general, in the form of substitution matrices such as PAM and BLOSUM, which were reported to be correlated with fitness for TEM-1 β-lactamase ($r = 0.4$) (Firnberg et al. 2014). For AraCDE, RP S20 and RP L1, amino acid substitution matrices did not show a significant predictive ability (fig. 4A, $r = 0.04–0.39$, $P > 0.01$) with the exception of PAM10 and AraD ($r = 0.50$, $P = 0.0015$) (supplementary fig. S3A and B and table S3, Supplementary Material online). Possibly this is due to the fact that AraD is the only enzyme among the five proteins and that other classes of proteins are underrepresented in the sequences used to construct amino acid substitution matrices.

Methods based on multiple alignments of homologous sequences proved more successful and for AraCDE these have a significant predictive ability (Consurf $r = 0.47–0.58$, $P < 0.004$, Provean $r = 0.47–0.61$, $P < 0.002$), whereas for the ribosomal proteins evolutionary conservation did not correlate significantly with fitness cost ($r < 0.26$, $P > 0.09$) (fig. 4B; supplementary fig. S3C and D and table S3, Supplementary Material online).

An alternative approach is to use structural information from experimental structures or homology models to make computational predictions of changes in thermodynamic protein stability, which has been recognized as an important determinant of the fitness costs of mutations (DePristo et al. 2005; Wylie and Shakhnovich 2011). In contrast to the alignment-based methods, information about active site residues or vital intermolecular interactions is not included, which is expected to reduce correlation with fitness because many complete loss-of-function mutations will not be predicted. The simplest of structure-based parameters is the relative solvent accessibility (RSA) of the mutated amino acid. A low RSA represents a buried amino acid and that mutations in that position are more likely to disrupt protein structure. Given the indication that many mutations classified as neutral were at the surface of the proteins, we calculated the exposed surface area of each amino acid in homology models of both ribosomal proteins and AraCDE and found a significant correlation between fitness cost and RSA (fig. 4C, $r = 0.38–0.47$, $P = 0.035–0.0035$, two sided t-test, supplementary table S3 and fig. S3E, Supplementary Material online).

There are a large variety of publicly accessible programs to estimate the effect of mutation on protein stability [reviewed in Kucukkal et al. (2014)]. A comprehensive evaluation of these methods for predicting fitness effects was not performed because several of the studied proteins did not have experimental structures for S. typhimurium and initial efforts using three different methods generally performed worse than for RSA (fig. 4C; supplementary table S3 and fig. S3F and G, Supplementary Material online). In addition, structure-based methods are expected to provide very low correlations for the ribosomal proteins given that a major part of the fitness constraints is on the mRNA level.

We used several publicly available methods that predict phenotypic effects of mutations by combining evolutionary conservation and structural effects as well as a range of other parameters using neural network classifiers (SNAP2; supplementary fig. S3H, Supplementary Material online), support vector machines (Suspect; supplementary fig. S3I, Supplementary Material online), SNPs&GO (supplementary fig. S3K, Supplementary Material online), or random forest (MutPred, supplementary fig. S3J, Supplementary Material online). Whereas these methods were not primarily designed to predict fitness effects, but rather the relevance of mutations for human disease, they could have substantial predictive ability also for nonhuman proteins. It is however reasonable to expect these methods to be better suited for predicting large effect mutations that are have been shown to be strongly associated with disease, have large effects on protein function or are in highly conserved positions in sequence alignments. For AraD, AraE and RP520 there were minor improvements from the methods using only conservation and SNPs&GO using homology models of the proteins produced the highest correlations (AraD $r = 0.71$, $P < 10^{-4}$; AraE $r = 0.68$, $P < 10^{-4}$; and S20 $r = 0.43$, $P = 0.003$) (fig. 4D; supplementary table S3 and fig. S3K, Supplementary Material online). Rank correlations for these methods are similar suggesting a minor influence of extreme values (supplementary table S3, Supplementary Material online). For AraC none of the methods produced a higher correlation than Consurf ($r = 0.58$, $P = 10^{-4}$, supplementary fig. S3D, Supplementary Material online) that uses only evolutionary conservation and no significant predictions were produced for RPL1 where the best
method was RSA ($r < 0.39, P > 0.01$) (supplementary table S3, Supplementary Material online).

Clearly some of these methods can produce highly significant correlations between score and fitness effect and this is largely due to their ability to identify mutations with very large effects, which does not necessarily mean that they are useful for distinguishing between deleterious and neutral mutations. The most successful methods for classifying mutations in the five proteins as either deleterious or neutral were Provean (72% correct) and SNAP2 (73% correct) and several methods produced higher than 60% correct classifications (supplementary table S3, Supplementary Material online). Up to 89% of the highly deleterious ($s < -0.03$) mutations were predicted as costly whereas the proportions were lower for slightly deleterious ($\approx 70\%$) and about 60% of mutations predicted as neutral had no significant fitness effects (supplementary table S3, Supplementary Material online). It should be noted that the proportion of truly neutral mutations is demonstrated experimentally in mutants with increased error rates (Goldsmith and Tawfik 2009; Bratulic et al. 2015) under stringent selection for function. An additional factor that could work with the misreading hypothesis to drive RPs towards higher robustness than the AraCDE proteins is a difference in the strength of selection. Thus, in natural settings, the arabinose proteins are most likely under strong selection only during short periods when arabinose is limiting for growth (as is the case in the present experiments). Under conditions where arabinose is not limiting, most mutations in the ara genes would be (near-) neutral and their DFE would look robust with a much smaller value of $s$ than that measured here. In contrast, ribosomal proteins are in constant use and would be under selection under all growth conditions. In general, the shape of the DFE and its robustness must depend on the growth conditions under which it is measured.

The frequency of nonsynonymous mutations that were classified as deleterious was 56% for the AraCDE proteins and 73% if also including the ribosomal proteins. This is similar to the fraction strongly deleterious effect in mammalian species of about 70% (Eyre-Walker et al. 2002; Yampolsky et al. 2005). We classified 40% of the amino acid substitutions in AraCDE as "neutral", but the truly neutral mutations with $1 < N_e \times |s| \approx 0.004$ whereas fitness costs $10^4$ times smaller could be subject to natural selection in many bacterial species. In humans, with an effective population size of about $10^4$ (Charlesworth 2009), the fraction of slightly deleterious ($s > -10^{-3}$) and neutral mutations has been estimated to be 44–57% (Bataillon and Bailey 2014), which suggest that the major differences in the fractions of neutral and deleterious substitutions between species is largely an effect of variation in population sizes.

Based on the pattern of highly deleterious mutations in AraCDE we suggest that these strong fitness effects, and the resulting bimodal fitness distributions, are not primarily due to reduced thermodynamic stability. This been suggested as the main cause of highly deleterious mutational effects in TEM-1 beta lactamase and GFP, but it is possible that disruption the buried functional sites in these proteins also contributes to the correlation (Firnberg et al. 2014; Sarkisyan et al. 2016). Instead our hypothesis is that the amino acid molecular functions, suggest that a bimodal distribution is widespread among different classes of proteins and that the fractions of deleterious mutations are similar ($\approx 50\%$). This provides a general expectation for the shape of the DFE for individual genes under strong selection, but the results for the RPs demonstrate that some classes of genes have completely different DFES.

We suggest that selection to reduce the costs of misreading errors is the main evolutionary force behind the high robustness of ribosomal proteins and possibly other proteins in large complexes. This hypothesis is based on well-established population genetics theory and experimental measurements of misreading rates and provides a general expectation for the robustness of proteins across functional classes. The importance of misreading errors and evolution towards increased mutational robustness has previously been demonstrated experimentally in mutants with increased error rates (Goldsmith and Tawfik 2009; Bratulic et al. 2015) under stringent selection for function. An additional factor that could work with the misreading hypothesis to drive RPs towards higher robustness than the AraCDE proteins is a difference in the strength of selection. Thus, in natural settings, the arabinose proteins are most likely under strong selection only during short periods when arabinose is limiting for growth (as is the case in the present experiments). Under conditions where arabinose is not limiting, most mutations in the ara genes would be (near-) neutral and their DFE would look robust with a much smaller value of $s$ than that measured here. In contrast, ribosomal proteins are in constant use and would be under selection under all growth conditions. In general, the shape of the DFE and its robustness must depend on the growth conditions under which it is measured.

The frequency of nonsynonymous mutations that were classified as deleterious was 56% for the AraCDE proteins and 73% if also including the ribosomal proteins. This is similar to the fraction strongly deleterious effect in mammalian species of about 70% (Eyre-Walker et al. 2002; Yampolsky et al. 2005). We classified 40% of the amino acid substitutions in AraCDE as "neutral", but the truly neutral mutations with $1 < N_e \times |s| \approx 0.004$ whereas fitness costs $10^4$ times smaller could be subject to natural selection in many bacterial species. In humans, with an effective population size of about $10^4$ (Charlesworth 2009), the fraction of slightly deleterious ($s > -10^{-3}$) and neutral mutations has been estimated to be 44–57% (Bataillon and Bailey 2014), which suggest that the major differences in the fractions of neutral and deleterious substitutions between species is largely an effect of variation in population sizes.

Based on the pattern of highly deleterious mutations in AraCDE we suggest that these strong fitness effects, and the resulting bimodal fitness distributions, are not primarily due to reduced thermodynamic stability. This been suggested as the main cause of highly deleterious mutational effects in TEM-1 beta lactamase and GFP, but it is possible that disruption the buried functional sites in these proteins also contributes to the correlation (Firnberg et al. 2014; Sarkisyan et al. 2016). Instead our hypothesis is that the amino acid
substitutions disrupt intermolecular interactions and directly reduce function through loss of proper binding affinity to a small molecular substrate (AraC, AraD, and AraE), loss of DNA or protein recognition (AraC) and disruption of oligomerization (AraC and AraD). If this interpretation would prove correct, studies of monomeric enzymes or fluorescent proteins might not be representative for the majority of proteins. The mechanistic reason for the lack of highly deleterious mutations in ribosomal proteins could then be a result of their binding to ribosomal RNA through numerous amino acid residues and this interaction cannot be completely disrupted by single substitutions. Although rates of molecular evolution are largely determined by biophysical constraints on thermodynamic stability and protein abundance (DePristo et al. 2005; Drummond et al. 2005; Serohijos et al. 2013), the shapes of DFEs are, under this hypothesis, mainly determined by the potential for loss of intermolecular interactions. This idea is similar to the concept of functional density, which was introduced by Zuckerkandl (1976) as an hypothesis to explain rates of molecular evolution. However, we suggest that the proportion of molecular sites involved in specific functions as compared with general functions (i.e., functional density) determines the fraction of highly deleterious mutations and forms the basis for bimodal DFEs, but it does not determine evolutionary rate at the molecular level. Our results also show that very highly conserved proteins, like ribosomal proteins, are not necessarily fragile as previously suggested (Assis and Kondrashov 2014), and that fitness constraints on the mRNA level might also reduce evolutionary rates.

The effects of individual nonsynonymous mutations on fitness can be predicted with some success using evolutionary conservation or machine learning methods, achieving a correlation of up to 0.71 and correctly classifying up to 73% of mutations as deleterious or neutral, but with highly varying results for different genes depending on methods used (supplementary table S3, Supplementary Material online). The difficulty of correctly predicting mutations with small effects of $s = -0.03$ to $-0.004$, a range that is not possible to study using less sensitive fitness assays, suggests that a substantial fraction of deleterious mutations are not correctly classified using these methods. This could have relevance for the genetic basis of complex human disease where alleles with small deleterious effects are not easily detected, but could make a significant contribution to the etiology of the disease when combined (Gibson 2011). If the DFEs of different protein classes are conserved within and between species then prediction of mutational effects would be greatly simplified, which could inform evolutionary forecasting by reducing the region of sequence space that must be experimentally explored to estimate the robustness of a gene or functional network.

Materials and Methods

Strains and Media

All strains used are derived from S. enterica subsp. enterica serovar Typhimurium str. LT2 (designated S. typhimurium in the text). Luria-Bertani agar (LA) was used as solid media, supplemented with sucrose 6%, kanamycin 30 mg/l, chloramphenicol 12.5 mg/l, or tetracycline 7.5 mg/l for selection following transformation or maintenance of the pSIMS λ Red plasmid. LA plates without added sodium chloride were used for selection against sacB. Lysogeny broth (LB), SOC, or M9 minimal media supplemented with 0.4% (w/v) L-arabinose were used as liquid media.

Strain Construction

Mutations in araD (STM0101), araC (STM0104), and araE (STM3016) were introduced using λ Red recombineering (Datsenko and Wanner 2000). For araD, mutations were selected by computer randomization with uniform probability for mutations location and mutation type, but excluding stop codons. Mutagenic primers containing the desired mutation 6 bp from the 3’-end and with 40 bp of homology 5’ to the mutation were used to PCR amplify part of the araD open reading frame and an adjacent kanamycin resistance marker using Phusion high fidelity DNA polymerase (Thermo Scientific). For araC and araE we used the Genemorph II random mutagenesis kit (Agilent Technologies) to introduce mutations in the open reading frames using pZE21 plasmids with araC or araE in the multiple cloning site as template (0.5–0.8 μg per 20 μl reaction for 21–25 cycles).

Mutated DNA was electroporated into λ Red competent cells and for araD kanamycin resistance was used to select for recombinants. For araC and araE, we constructed strains with a sacB marker, conferring sucrose sensitivity, inserted into the chromosome replacing the araC and araE open reading frames. This allowed us to create marker-less mutations by introducing the mutated DNA into λ Red competent cells and then select for sucrose resistance and verify inserts with PCR and sequencing. This was the main reason for changing from the more labor-intensive mutagenic primers method used for araD, which had the advantage of introducing a less biased spectrum of mutations (supplementary table S4, Supplementary Materials online). All mutations were moved using P22 transduction into strains with fluorescent protein marker cassettes (amp-PLacO-mTagBFP2/sYFP2, Genbank: KM018302/KM018301 (Gullberg et al. 2014)) inserted at a neutral position (cbiA, STM2035; Knöppel et al. 2014) for the competition experiments. For the araD mutants kanamycin resistance was used to select for successful transductions and for the araC and araE mutants the strains with fluorescent proteins had sacB cassettes inserted into the araC or araE genes which allowed selection for sucrose resistance after P22 transduction. This protocol reduces the impact of secondary mutations/rearrangements introduced during λ Red recombineering and adaptive mutations during laboratory manipulation and creates two independent strains to ascertain that the fitness effects are dependent on the mutation introduced. It should be noted that mutations in araDE causing extreme toxicity leading to lethality when grown on LB agar cannot be detected in our experiments, but there is little reason to expect these kinds of mutations to exist given the low expression level in the absence of arabinose and the molecular functions of the genes.
Independent control strains were isolated from the same populations as the mutants with the only difference being the lack of mutations (i.e., failed mutagenesis) when sequenced. All colonies were selected at random, only dependent on their position on the agar plate, after transformation, transductions, and for competitions. All strains were stored in 10% dimethyl sulfoxide and 90% LB at −80 °C and all cultures were frozen immediately after completion of transformation or transduction protocols.

**Fitness Assay**

Strains were taken from the freezer and incubated overnight at 37 °C on LA plates. Single colonies were used to inoculate 180 µl LB cultures in 96-well round bottom plates (Thermo Scientific #262162) and grown for 6 h with 900 rpm shaking (Grant PHMP-4) at 37 °C. Mutant and control strains containing opposite fluorescent markers were mixed 1:1 (5 µl each) and grown overnight under the same conditions. After 16 h, 1 µl of each culture was transferred to 180 µl M9 arabinose and grown for 24 h (900 rpm, 37 °C). At the same time, each culture was diluted 200-fold in phosphate buffered saline (PBS) for flow cytometry. In order to estimate the ratio of wild-type to mutant for each culture we counted 100,000 cells and determined the fraction of BFP and YFP positive cells using flow cytometry (MACSQuant VYB, Miltenyi Biotec). After 24 h growth the protocol was repeated to obtain the ratio after approximately seven generations of growth. We maintained the cultures for three additional serial transfers of 1.4 µl culture to 180 µl M9 arabinose before measuring final ratios after 28 generations of growth.

The selection coefficients were determined using the regression model $s = \frac{\ln (R(t)/R(0))}{\ln(t)}$, as previously described (Dykhuizen 1990) where $R$ is the ratio of mutant to wild-type. For mutants with very high fitness costs ($s < -0.15$) the ratios in the final count were too low and selection coefficients were calculated using the ratios from the first flow cytometry measurements (after growth in LB) and after 24 h growth in M9 arabinose, representing approximately seven generations. For the majority of mutants the selection coefficients were calculated from the second flow cytometry assay (after 24 h growth in M9 arabinose) and after three serial transfers, approximately 21 generations of growth.

The fitness of each mutant was measured in triplicate for both BFP and YFP markers, where each competition was inoculated with independent single colonies of mutant and control strains. The differences in fitness cost of the fluorescent markers were measured using control competitions included for each competition plate (minimum of six cultures) and used to adjust the selection coefficients. Three (araC, araE) or four (araD) wild-type control strains that were independently constructed on different days, were used to measure the standard variation between genetically identical strains. The average cost of the BFP marker relative to the YFP marker were $s = -0.0018$ (SD = 0.0021, $n = 24$) for araD, $s = -0.0028$ (SD = 0.0020, $n = 17$) for araC, and $s = -0.0015$ (SD = 0.0019, $n = 18$) for araE (Supplementary fig. S5, Supplementary Material online). Based on these results, we decided that a mean selection coefficient between $-0.004$ and 0.004 (approximately ±2 SD) for six cultures is designated as “neutral” in the sense that they cannot be reliably distinguished from the wild-type. After adjustment for the differential cost of the marker, the fitness effects in the BFP and YFP strains were highly correlated ($r = 0.979$, median difference $s_{BFP-s_{YFP}} = 4 \times 10^{-5}$ indicating no systematical differences between strains compared with the wild-type reference).

Adaptive or compensatory mutations can arise during competition experiments and to reduce the influence of such events on estimated fitness we removed a maximum of one extreme outlier if it had a selection coefficient more than two standard deviations from the average of all replicates. For highly deleterious mutations that could clearly be classified as costly, outliers were not removed. In cases where more than one replicate diverged from the others or there was a large difference between YFP and BFP strains, the competition was repeated. If differences between YFP and BFP remained, the strains were independently reconstructed using Red recombination. This was also done for synonymous mutations with non-neutral effects or for any mutations with significantly higher fitness than the wild-type control.

**Analysis of Fitness Effects**

Information about the mutational spectrum is provided in (supplementary table S4, Supplementary Material online). The distribution of mutations over the length of the genes was not significantly biased (Kolmogorov–Smirnov test for uniform distribution: $araC \ P = 0.9684$, $araD \ P = 0.8824$, and $araE \ P = 0.9975$) (supplementary fig. S4, Supplementary Material online). Homology models were made using Phyre2 intensive mode (http://www.sbg.bio.ic.ac.uk/phyre2; last accessed November 7, 2016) (Kelley et al. 2015) and models had ≥95% of amino acids modeled at >90% confidence. Relative surface area was calculated at http://caps.ncbs.res.in/iws/psa.html (last accessed November 7, 2016). We used Consurf (consurf.tau.ac.il; last accessed November 7, 2016) (Ashkenazy et al. 2010), Provean (http://provean.jcvi.org; last accessed November 7, 2016) (Choi and Chan 2015), and SIFT sequence (http://sift.jcvi.org; last accessed November 7, 2016) (Ng and Henikoff 2003) to investigate evolutionary conservation. Effects on protein stability were estimated using I-mutant 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi; last accessed November 7, 2016) (Carpiotti et al. 2005), SDM (http://mordred.bioc.cam.ac.uk/~sdm/sdm.php; last accessed November 7, 2016) (Worth et al. 2007), and Eris (http://troll.med.uned.es/eris/login.php; last accessed November 7, 2016) (Yin et al. 2007). Functional effects of nonsynonymous mutations were predicted with Suspect (http://www.sbg.bio.ic.ac.uk/suspect; last accessed November 7, 2016) (Yates et al. 2014), SNAP2 (https://roslab.org/services/snap2; last accessed November 7, 2016) (Hecht et al. 2015), MutPred (http://mutpred.mutdb.org; last accessed November 7, 2016) (Li et al. 2009), and SNPs&GO (http://snps.biofold.org/snps-and-go/index.html; last accessed November 7, 2016) (Capriotti et al. 2013). Results are available in supplementary...
table S2, Supplementary Material online and correlations and classifications in supplementary table S3, Supplementary Material online. Default settings were used except when indicated in supplementary table S2, Supplementary Material online and classification thresholds for deleterious or neutral mutations were set as recommended by the creators of each method.

**Supplementary Material**

Supplementary tables S1–S4 and figures S1–S5 are available at *Molecular Biology and Evolution* online.

**Author Contributions**

P.A.L. and D.I.A. conceived and initiated this study. P.A.L., O.G.B., and D.I.A. designed the experiments and interpreted results. O.G.B. developed mathematical models. P.A.L. and L.A. carried out experiments. P.A.L., O.G.B., and D.I.A. wrote the manuscript.

**Acknowledgments**

This work was supported by grants from the Swedish Research Council (VR-NT) to D.I.A. The authors declare no competing financial interests.

**References**

Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 38:W529–W533.

Assis R, Kondrashov AS. 2014. Conserved proteins are fragile. *Mol Biol Evol.* 31:419–424.

Bakshi S, Sirajdeen A, Goulain M, Weishaar JC. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol Microbiol.* 85:21–38.

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

Berg OG. 1996. Selection intensity for codon bias and the effective population size of *Escherichia coli*. *Mol Biol Evol.* 13:135–145.

Bratlalic S, Gerber F, Wagner A. 2015. Mistranslation drives the evolution of robustness in TEM-1 beta-lactamase. *Proc Natl Acad Sci U S A.* 112:12758–12763.

Brockieler L, Karlin S. 2005. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* 33:3390–3400.

Caprioni E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. 2013. WS-SNP: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics* 14 Suppl 3:56.

Caprioni E, Fariselli P, Casadio R. 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* 33:W306–W310.

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

Choi Y, Chan AP. 2015. PROVEAN: a server to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 31:2745–2747.

Darzuwalla KR, Paxton AT, Henderson PJ. 1981. Energization of the transport systems for arabinose and comparison with galactose transport in *Escherichia coli*. *Biochem J.* 200:611–627.

Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A.* 97:6640–6645.

de Visser JA, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet.* 15:680–690.

DePristo MA, Weinreich DM, Hartl DL. 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet.* 6:578–587.

Desai MM, Fisher DS, Murray AW. 2007. The speed of evolution and maintenance of variance in asexual populations. *Curr Biol.* 17:385–394.

Dykhuijzen DE. 1990. Experimental studies of natural selection in bacteria. *Annu Rev Ecol Syst.* 21:373–398.

Englesberg E, Anderson RL, Weinberg R, Lee N, Hoffee P, Huttenhauer G, Boyer H. 1962. L-Arabino-sensitively, L-ribulose 5-phosphate 4-epimerase-deficient mutants of *Escherichia coli*. *J Bacteriol.* 84:137–146.

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

Eyre-Walker A, Keightley PD, Smith NG, Gaffney D. 2002. Quantifying the slightly deleterious mutation model of molecular evolution. *Mol Biol Evol.* 19:2142–2149.

Fares MA. 2015. The origins of mutational robustness. *Trends Genet.* 31:373–381.

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

Fisher RA. 1930. The genetical theory of natural selection. Oxford, UK: Clarendon Press.

Foster PL, Lee H, Popodi E, Townes JP, Tang H. 2015. Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. *Proc Natl Acad Sci U S A.* 112:E5990–E5999.

Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL. 1997. Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev.* 61:393–410.

Gibson G. 2011. Rare and common variants: twenty arguments. *Nat Rev Genet.* 13:135–145.

Goldsmith M, Tawfik DS. 2009. Potential role of phenotypic mutations in the evolution of protein expression and stability. *Proc Natl Acad Sci U S A.* 106:6197–6202.

Gullberg E, Albrecht LM, Karlsson C, Sandegren L, Andersson DI. 2014. Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. *mBio* 5:e01918–e01914.

Hecht M, Bromberg Y, Rost B. 2015. Better prediction of functional effects for sequence variants. *BMC Genomics* 16 Suppl 8:S1.

Hietpas RT, Jensen JD, Bolon DN. 2011. Experimental illumination of a fitness landscape. *Proc Natl Acad Sci U S A.* 108:7986–7991.

Iancu CV, Zamoons J, Woon SB, Aleshin A, Choe YJ. 2013. Crystal structure of a glucose/H+/symporter and its mechanism of action. *Proc Natl Acad Sci U S A.* 110:17862–17867.

Imashimizu M, Oshima T, Lubkowska L, Kashlev M. 2013. Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Res.* 41:9590–9110.

Jacquier H, Birge A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot D. 2011. Crystal structure of the beta-lactamase TEM-1. *Nat Chem Biol.* 41:9090–9110.

Jaccquier H, Birge A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot D, Petit E, Pouliain J, Barnaud G, et al. 2013. Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc Natl Acad Sci U S A.* 110:3067–31072.

Jimenez J, Dulvi-Brunet R, Campbell GW, Turk-MacLeod R, Chen IA. 2013. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc Natl Acad Sci U S A.* 110:14984–14989.

Jones S, Doley DT, Luscombe NM, Berman HM, Thornton JM. 2001. Protein–RNA interactions: a structural analysis. *Nucleic Acids Res.* 29:943–954.

Kährström CT, Andersson DI, Sanyal S. 2012. Comprehending a molecular conundrum: functional studies of ribosomal protein mutants.
from Salmonella enterica serovar typhimurium. In: Annous BA, ed. Salmonella—Distribution, Adaptation, Control Measures and Molecular Technologies. Croatia: InTech.

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.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 10:845–858.

Kimura M, Maruyama T. 1966. The mutational load with epistatic gene interactions in fitness. Genetics 54:1337–1351.

Knöppel A, Lind PA, Lustig U, Nasvall J, Andersson DI. 2014. Minor fitness costs in an experimental model of horizontal gene transfer in bacteria. Mol Biol Evol. 31:1220–1227.

Knöppel A, Näsjö M, Andersson DI. 2016. Compensating the fitness costs of synonymous mutations. Mol Biol Evol. 33:1461–1477.

Kucukkal TG, Yang Y, Chapman SC, Cao W, Alexov E. 2014. Computational and experimental approaches to reveal the effects of single nucleotide polymorphisms with respect to disease diagnostics. Int J Mol Sci. 15:9670–9717.

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

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

Luo Y, Samuel J, Mosimann SC, Lee JE, Tanner ME, Strynadka NC. 2007. Adaptive mutations in of translation. Proc Natl Acad Sci U S A. 104:L1–L3.

Meyerovich M, Marnou G, Ben-Yehuda S. 2010. Visualizing high error levels during gene expression in living bacterial cells. Proc Natl Acad Sci U S A. 107:11543–11548.

Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 31:3812–3814.

Nikolay R, van den Bruck D, Achenbach J, Nierhaus KH. 2001. Ribosomal proteins: Role in ribosomal functions. Chichester, UK: John Wiley & Sons, Ltd.

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

Schenk MF, Szendro IG, Krug J, de Visser JA. 2012. Quantifying the adaptive potential of an antibiotic resistance enzyme. PLoS Genet. 8:e1002783.

Schleif R. 2010. AraC protein, regulation of the I-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. FEMS Microb. Rev. 34:779–796.

Schluenzen F, Todoli A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, et al. 2000. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell 102:615–623.

Selmer M, Dunham CM, Murphy FV, Weixlbaumer A, Petry S, Kelley AC, Weir JR, Ramakrishnan V. 2006. Structure of the 70S ribosome complexed with mRNA and tRNA. Science 313:1935–1942.

Serojihos AW, Lee SY, Shakhnovich EL. 2013. Highly abundant proteins favor more stable 3D structures in yeast. Biophys J. 104:L1–L3.

Shajani Z, Sykes MT, Williamson JR. 2011. Assembly of bacterial ribosomes. Annu Rev Biochem. 80:501–526.

Sniegowski PD, Gerrish PJ. 2010. Beneficial mutations and the dynamics of adaptation in asexual populations. Philos Trans R Soc Lond B Biol Sci. 365:1255–1263.

Soisson SM, MacDougall-Shackleton B, Schleif R, Wolberger C. 1997. Structural basis for ligand-regulated oligomerization of AraC. Science 276:421–425.

Suckow J, Markiewicz P, Kleina LG, Miller J, Kisters-Woike B, Muller-Hill B. 1996. Genetic studies of the Lam repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. J Mol Biol. 261:509–523.

Sun L, Zeng X, Yan C, Sun X, Gong X, Rao Y, Yan N. 2012. Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. Nature 490:361–366.

Tobin C, Mandava CS, Ehrenberg M, Andersson DI, Sanyal S. 2010. Ribosomes lacking protein S20 are defective in mRNA binding and subunit association. J Mol Biol. 397:767–776.

Travisano M, Shaw RG. 2013. Lost in the map. Evolution 67:305–314.

Worth CL, Bickerton GR, Schreyer A, Forman JR, Cheng TM, Lee S, Gong S, Burke DF, Blundell TL. 2007. A structural bioinformatics approach to the analysis of nonsynonymous single nucleotide polymorphisms (nsSNPs) and their relation to disease. J Bioinform Comput Biol. 5:1297–1318.

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

Yampolsky LY, Kondrashov FA, Kondrashov AS. 2005. Distribution of the strength of selection against amino acid replacements in human proteins. Hum Mol Genet. 14:3191–3201.

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

Yin S, Ding F, Dokholyan NV. 2007. Eris: an automated estimator of protein stability. Nat Methods. 4:466–467.

Zuckerkerandl E. 1976. Evolutionary processes and evolutionary noise at the molecular level. I. Functional density in proteins. J Mol Evol. 7:167–183.