Adaptive mutations in RNA polymerase and the transcriptional terminator Rho have similar effects on *Escherichia coli* gene expression

Andrea González-González[1,3], Shaun M. Hug[1], Alejandra Rodríguez-Verdugo[2,3], Jagdish Suresh Patel[4], Brandon S. Gaut[1,5]

[1] Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA.

[2] Department of Environmental Systems Sciences, ETH Zürich, Zürich, Switzerland.

[3] Department of Environmental Microbiology, Eawag, Dübendorf, Switzerland

[4] Center for Modeling Complex Interactions, University of Idaho, Moscow, ID, USA.

[5] Address for Correspondence: bgaut@uci.edu; 321 Steinhaus Hall, UC Irvine, 92617

ABSTRACT

Experimental microbial evolution (EME) provides the opportunity to study the mutations that drive adaptation. We recently completed an EME that subjected 115 independent *Escherichia coli* populations to thermal stress (42.2°C) for a year. At the end of the experiment, we identified two major, negatively epistatic 'adaptive pathways'. One pathway had mutations in *rpoB*, the gene encoding the RNA polymerase β subunit, and the other that had mutations in *rho*, which encodes a transcriptional terminator. Here we focused on four putatively beneficial *rho* mutations by introducing them into the ancestral genotype and by measuring their effects on fitness, gene expression, protein structure and termination efficiency. Results were heterogenous among mutations; two conferred adaptive effects at levels similar to four beneficial *rpoB* mutations, without any discernible effect on transcriptional termination. The other two enhanced transcriptional read-through and were not adaptive, suggesting they are beneficial only in the context of epistatic interactions. Fitness had no discernible relationship with the predicted effect of mutations on the free energy of protein folding. Overall, the *rho* mutations had similar effects on gene expression as *rpoB* mutations. Of 1726 genes that were differentially expressed between *rpoB* mutations and the EME ancestor, 83% were also affected by *rho* mutations. Moreover, both *rho* and *rpoB* mutations tended to restore gene expression to that of the unstressed (37.7°C) ancestor, and the extent of restoration correlated quantitatively with fitness. This rare opportunity to compare beneficial mutations within and across adaptive pathways suggests the pathways traverse similar fitness landscapes.

Key words: transcriptional termination, protein structure, relative fitness, experimental evolution, adaptive pathways; epistasis.

INTRODUCTION

Experimental microbial evolution (EME) has long been able to demonstrate adaptation in the lab, but the ability to identify the genetic variants that accompany adaptation has come more recently. It is now routine to sequence hundreds of isolates or populations of yeast, *Escherichia coli* and other microbes. These sequence data provide important insights into the type and number of potentially adaptive mutations, epistatic interactions between mutations, and the diversity of alternative adaptive pathways. However, the description of genetic variants is only a first step. As noted in a recent review (Rosenzweig and Sherlock 2014), "the challenge now is not to identify the mutations, but instead to distinguish the passengers from the drivers" – i.e., to unequivocally separate the mutations that contribute to an adaptive response from those that hitchhike as neutral or slightly deleterious. Once identified, there remain additional difficult tasks: to understand the effect of adaptive mutations on phenotype and physiology and to contrast those effects between alternative adaptive trajectories.

We recently performed an EME that identified alternative genetic pathways to adapt to thermal stress. The experiment evolved 115 populations of *E. coli* independently from the same ancestral clone (REL1206) for 2000 generations under strong thermal stress at 42.2°C (Tenaillon et al. 2012). We studied thermal stress because temperature is a complex environmental variable that governs the rate of a broad range of biological reactions. At the end of the experiment, we isolated one clone from each population and found that fitness had improved ~40%, on average, relative to REL1206. We also sequenced the evolved clones, identifying over 1300 mutations in the process. Using parallelism as a *prima facie* argument for adaptation (Wichman et al. 1999; Fong et al.

2005; Stern 2013), about half of the observed 1300 mutations were adaptive, either because the same mutation occurred in more than one clone or because mutations occurred in the same gene in different clones.

Two of the genes mutated most often in our EME are necessary for cell growth due to their role in transcription. One gene, the *rpoB* gene that encodes the β subunit of RNA polymerase (RNAP), accumulated a total of 87 mutations in 76 of the 115 clones. The second gene encodes the transcription termination factor Rho*;* the *rho* gene accumulated a total of 24 different nonsynonymous mutations across 43 of 115 clones. Mutations in *rpoB* and *rho* were found together less often than expected at random, which is indicative of negative epistatic interactions (Tenaillon et al. 2012).

Broadly speaking, these two genes define two alternative 'adaptive pathways', both of which include additional genes. The *rpoB* pathway tended to include mutations in other RNA polymerase subunits (*rpoA*, *rpoC* and *rpoD*) and in the six *rod* genes that affect cell shape. In contrast, the *rho* pathway included mutations of the cardiolipin synthase (*cls*) gene and the transcription factor gene *iclR.* We have shown that these two adaptive pathways differ in some aspects of their phenotypes, such as chemical sensitivities (Hug and Gaut 2015) and trade-offs for growth at low temperature (Rodriguez-Verdugo et al. 2014). Nonetheless, questions remain regarding the phenotypic and mechanistic effects of mutations in *rho* and *rpoB* as well as the mechanistic cause(s) of negative epistasis. Do beneficial mutations in *rho* and *rpoB* have similar affects on fitness and phenotype? That is, do they represent distinct or largely overlapping paths toward optimizing fitness? And what might their comparison reveal about the molecular mechanisms that contribute to adaptation?

In retrospect, it is not surprising that we observed many mutations in *rpoB*, because mutations in this gene have been found commonly in EMEs (Herring et al. 2006; Charusanti et al. 2010; Conrad et al. 2009), as have mutations in other global regulators of gene expression (GE) (Conrad et al. 2011; Le Gac et al. 2013; Saxer et al. 2014). Among the observed *rpoB* point mutations from our EME, we introduced four as single mutations into the ancestral REL1206 background (Rodriguez-Verdugo et al. 2014). Each of these mutations conferred a substantial (>17%) fitness advantage, arose and fixed rapidly within the experimental populations (Rodriguez-Verdugo et al. 2013), altered the expression of hundreds to thousands of genes (Rodriguez-Verdugo et al. 2015), and tended to restore GE to become more similar to the unstressed 37.0°C state of the ancestor rather than the physiologically stressed state at 42.2°C (Rodriguez-Verdugo et al. 2015). Thus, these four *rpoB* mutations are unambiguously advantageous and highly pleiotropic.

Like *rpoB* mutations, mutations within the *rho* gene have also been found in EMEs (Goodarzi et al. 2009; Haft et al. 2014; Lee and Helmann 2014; Kishimoto et al. 2010; Conrad et al. 2009; Herron and Doebeli 2013; Le Gac et al. 2013; Dillon et al. 2016). Thus far, two *rho* mutations - a phenylalanine to leucine replacement at codon 62 (*F62L*) and a leucine to methionine replacement in codon 270 (*L270M*) – have been studied at the level of GE. Both altered the expression of hundreds of coding and intergenic regions (Freddolino et al. 2012; Haft et al. 2014). The majority of differentially expressed genes (DEGs) in *rho* mutants were up-regulated (Freddolino et al. 2012; Haft et al. 2014), perhaps because the termination properties of the Rho protein had been altered, leading to increased transcriptional read-through (Martinez et al. 1996).

It is worth noting an important difference between the potential functional effects of *rpoB* and *rho* mutations. RNAP mutations have the potential to affect the transcription of every gene within *E. coli* due to its fundamental role in transcription. In contrast, Rho terminates the transcription of ~20% to 50% of genes (Cardinale et al. 2008; Peters et al. 2009). Rho termination proceeds, first, by the protein recognizing a *rho ut*ilization (*rut*) site on an elongating mRNA; second, by binding *rut* sites and translocating mRNA through a central cavity (Skordalakes and Berger 2003); until, third, it contacts RNAP while it pauses, thereby disassociating the elongation complex (Peters et al. 2011; Skordalakes and Berger 2003). Genome-wide studies indicate that *rut* sites are located both intergenically and intragenically (Peters et al. 2009), but *rut* sites remain difficult to predict bioinformatically because they do not have a strong consensus sequence (Hollands et al. 2014; Ciampi 2006). As a consequence, the function(s) of Rho remain incompletely characterized.

For this study, we engineer four nonsynonymous *rho* mutations from our EME into the REL1206 ancestor. We then examine their effects on fitness and GE relative to four beneficial *rpoB* mutants (Rodriguez-Verdugo et al. 2014, 2015). Each of the four *rho* mutations are unambiguously adaptive by the criterion of parallelism in our thermal stress EME (Tenaillon et al. 2012), because they were found 2 or more independent experimental populations. The four mutations also represent different regions of the Rho protein, which is a hexameric ring composed of six monomers (Skordalakes and Berger 2003) (fig. 1). The N-terminal domain (NTD) of each monomer contains both the Primary RNA binding sites (PBS), which recognize and bind to mRNA, and the positively charged N-terminal helix bundle (NHB) motif that may facilitate the

recruitment of RNA to PBSs. Three of our four point mutations are located in this NHB domain. The fourth mutation is located in the central part of the C-terminal domain (CTD), specifically in the Loop-CTD. This region of the CTD contains the Secondary RNA binding sites (SBS) and the ATP hydrolysis motifs, both of which are involved in the translocation of mRNA along Rho (Skordalakes and Berger 2003; D'Heygère et al. 2013).

Our overarching goals are to differentiate "…passengers from drivers", to characterize the effects of mutation on *rho* function, and to begin to understand the phenotypic differences and similarities of alternative adaptive trajectories. With fitness and GE data from a set of eight *rho* and *rpoB* mutations, we focus on four sets of questions. First, do the four *rho* mutations confer similar relative fitness benefits to one another and to four *rpoB* mutations? Second, do the four *rho* mutations affect the expression of similar numbers and sets of genes, and are shifts in GE similar to those of *rpoB* mutations in both magnitude and direction? Third, is there a relationship between protein stability and fitness? To address this question, we use molecular modeling to assess the effects of amino acid replacements on protein stability, which in turn may be closely related to fitness (DePristo et al. 2005). Finally, we have previously hypothesized that both the *rpoB* and *rho* pathways alter the efficiency of transcriptional termination and that alterations to both genes are negatively epistatic because they tend to "over-tune" transcriptional termination (Rodriguez-Verdugo et al, 2014). Do *rho* mutations affect transcription termination in manner consistent with this hypothesis?

RESULTS

Relative fitnesses of *rho* and *rpoB* mutations

We engineered three *rho* mutations into the REL1206 ancestral background and also studied a fourth, previously engineered *rho* mutation (Rodriguez-Verdugo et al. 2014). The *I15N*, *I15F, A43T* and *T231A* substitutions were observed in 15, 2, 3 and 2 independent populations, respectively, of our 42.2°C thermal stress experiment (Tenaillon et al. 2012), and thus all four were beneficial by the criterion of parallelism. Given lines with these mutations in the REL1206 background, we assessed their relative fitness ($\overline{w}_r$) to REL1206 at 42.2˚C (supplementary table S1, Supplementary Material online). Relative fitness data conveyed three observations (table 1). First, two of the mutations (*A43T* and *T231A*) yielded a $\overline{w}_r$ value significantly > 1.0, indicating an adaptive benefit under thermal stress. In contrast, neither the *I15N* nor the *I15F* mutation produced a $\overline{w}_r$ value significantly different from 1.0, and *I15N* yielded an estimate < 1.0 (Rodriguez-Verdugo et al. 2014) suggesting a slightly deleterious mutation (table 1). Second, the $\overline{w}_r$ estimates differed significantly among the four mutations (ANOVA, $p = 2.4 \times 10^{-8}$), with *I15N* and *I15F* forming one group and *A43T* and *T231A* forming a second, significantly different group (Tukey-test; $p < 0.05$).

Finally, we contrasted $\overline{w}_r$ among four *rpoB* mutants (*I572L*, I*572F*, *I572N*, *I966S*) (Rodriguez-Verdugo et al., 2014) and the four *rho* mutants. Each of the *rpoB* mutations had $\overline{w}_r > 1.0$, with estimates ranging from $\overline{w}$ =1.16 to 1.26 at 42.2°C (Rodriguez-Verdugo et al. 2014) (table 1). Across all mutants, comparisons again indicated heterogeneity of $\overline{w}_r$ (ANOVA, $p = 1.91 \times 10^{-08}$; table 1), and Tukey tests grouped *I15N*

and *I15F* against the remaining six mutations ($p < 0.05$ for 10 of 12 comparisons between *I15* mutations and the remaining mutations).

Gene expression data

Differential Gene Expression among rho *mutants:* To gain insight into the phenotypic effects of *rho* mutations, we measured GE at 42.2˚C using replicated RNAseq data (see Materials and Methods). We hypothesized that expression should covary with $\overline{w}_r$, so that *I15N* and *I15F* differ more dramatically in expression from *A43T* and *T231A* than they do from each other. To test this hypothesis, we measured the correlation in GE across all genes for each pair of *rho* mutants (table 2). All pairwise correlations were high ($r^2 > 0.95$), but *A43T* and *T231A* had the highest correlation ($r^2 = 0.992$), while contrasts between an *I15* mutant with either *A43T* or *T231A* had lower correlations (table 2). We also identified differentially expressed genes (DEGs) between *rho* mutants, based on a significance cut-off of $q < 0.001$ (see Material and Methods). Only one gene was expressed differentially between *I15F* and *I15N*, but both *I15* mutations had > 100 DEGs between *A43T* and *T231A* (table 2). These data suggest that *A43T* and *T23A* mutations have quantitatively different effects on GE relative to the two *I15* mutations, mimicking observed differences in fitness among clones.

Differential gene expression between rho mutants and REL1206: Given that single *rho* mutations can confer extensive GE shifts relative to an ancestral background (Freddolino et al. 2012; Haft et al. 2014), we compared GE between the REL1206 ancestor at 42.2˚C and each of the four *rho* mutants. Across the four *rho* mutants, we detected a total of 1140 DEGs compared to the 42.2˚C ancestor, but there were more

DEGs for *T231A* (1028 genes) and *A43T* (656) than for *I15F* (372) and *I15N* (195) (fig. 2*A*).

Among the complete set of 1140 DEGs, 141 (or 12%) were shared among all four *rho* mutants (fig. 2*A*). Gene Ontology (GO) analysis of these 141 genes revealed significant enrichment for down-regulation of maltose transport genes (GO: 0015768, p = 1.04 x10^{-3}) and significant up regulation of genes involved in transcription (GO:0010467, $p = 1.12 \times 10^{-12}$) and translation (GO: 0006412, $p = 9.17 \times 10^{-37}$) (supplementary table S2, Supplementary Material online). Interestingly, the *A43T* and *T231A* mutants shared 330 DEGs that were not identified in either *I15F* or *I15N*. These 330 genes are likely candidates to contribute to $\overline{w}_r$ differences among *rho* mutants, and they were enriched for up-regulation of purine nucleotide biosynthetic processes (GO: 0006164, $p = 1.15 \times 10^{-12}$) and down-regulation for glycerol catabolic processes (GO: 0006164, $p = 1.15 \times 10^{-12}$) (supplementary table S3, Supplementary Material online).

Genes affected by both rho *and rpoB mutations:* To determine whether putatively adaptive mutations in *rpoB* or *rho* have similar effects on GE, we identified DEGs in four *rpoB* mutants -- *I572N*, *I572F*, *I572L* and *I966S* -- using previously published data (Rodriguez-Verdugo et al. 2015). We took the union of the DEG genes for the four *rpoB* mutations and found 1726 DEGs relative to REL1206 at 42.2˚C, and we compared them to the 1140 DEGs identified in *rho* mutants. These two genic sets were remarkably overlapping (fig. 2*B*): of the 1726 DEGs identified with *rpoB* mutations, 957 genes (or 55.4%) were also identified as DEGs in *rho* mutants. This similarity is more impressive when one considers that *rho* mutations altered GE for significantly fewer genes than *rpoB*

mutations (1140 vs. 1726; binomial test: $p < 2.2 \times 10^{-16}$) so that 83% (=957/1140) of DEGs caused by *rho* mutations were also identified as DEGs in *rpoB* mutants (fig. 2*B*).

The DEG overlap between *rho* and *rpoB* mutations left only 183 DEGs unique to the set of *rho* mutations (fig. 2B). Among these 183 DEGs, GO analyses identified up-regulated responses to temperature stimuli (e.g., *clpB*, *dnaK*, *dnaJ*, *degP, grpE*; GO: 0009408, $p = 3.35 \times 10^{-2}$) (supplementary table S4, Supplementary Material online). In contrast, the 769 genes uniquely modified by *rpoB* mutants were enriched for up-regulation of iron ion transport (GO: 0006826, $p = 3.2 \times 10^{-05}$) and flagellum cell motility (GO: 00048870, $p = 7.1 \times 10^{-06}$) genes. Monossacharide catabolic processes (GO: 0005996, $p = 2.0 \times 10^{-04}$) and cell wall biogenesis (GO: 0042546, $p = 2.94 \times 10^{-4}$) were also unique to *rpoB* mutants, but they were not consistently up- or down-regulated (supplementary table S4, Supplementary Material online).

Directional shifts: Although *rpoB* and *rho* mutants affect a highly overlapping gene set, it is not a foregone conclusion that they affect expression in the same direction. Here direction refers to expression shifts in mutants relative to the stressed ancestral state (i.e., REL1206 at 42.2 ˚C) and the unstressed ancestral state (i.e., REL1206 at 37.0˚C) (Carroll and Marx 2013).

We determined directionality in two ways. First, we applied PCA to all expression data for *rho, rpoB* and ancestral replicates, and then plotted the first two principal components, PC1 and PC2. Based on the PCA, the *rpoB* mutants were intermediate on PC1 between the ancestor at 37.0˚C and 42.2 ˚C (fig. 3). This position is consistent with an overall tendency of *rpoB* mutants to restore GE patterns from the stressed (42.2˚C) toward the unstressed (37.0˚C) ancestral state (Rodriguez-Verdugo et

al. 2015). GE data from *rho* mutants were also intermediate between the two ancestral states on PC1, but they were not restored as fully as the *rpoB* mutants (fig. 3). Intriguingly, replicates from the two less fit *rho* mutants (*I15N* and *I15F*) clustered more closely to REL1206 at 42.2˚C. These clustering patterns suggest that fitness is related to restoration in PCA space. To test this relationship quantitatively, we assessed the correlation between $\overline{w}$ and the average PC1 eigenvectors across the eight mutant types (r^2 = 0.4156; p = 0.042).

Second, we categorized expression of single genes into categories. Following previous studies (Carroll and Marx 2013), genes in mutants were classified as: *i*) "restored", if expression in the mutant shifts toward the unstressed ancestral state from the stressed ancestral state; *ii*) "unrestored", if expression remains similar to the REL1206 ancestor in its stressed state; *iii*) "reinforced" if expression in the mutant is exaggerated in the mutant relative to REL1206 at both 37.0˚C and 42.2˚C and, finally, *iv*) "novel" if the two ancestral treatments did not differ in GE but the mutant differed from both. (The Material and Methods section includes more precise quantitative definitions of these four directional categories.) The majority of genes fall into the restored category for all mutants except *rpoB I572L* (table 3). Moreover, the number of restored genes was correlated with $\overline{w}_r$ across the eight mutants (*rho* and *rpoB* mutants) at a borderline level of significance (Spearman's r = 0.67; p = 0.055). Altogether, directional analyses support the idea that increases in fitness are quantitatively linked with the restoration of GE to the unstressed state.

Potential molecular effects of *rho* mutations

Genome-wide analysis: We have shown that *rho* mutations differ in $\overline{w}_r$ and that $\overline{w}_r$ reflects shifts in the magnitude and direction of GE. However, the question of mechanism remains – i.e. how do the mutations in *rho* affect its function as a transcriptional terminator? Previous studies have indicated that *rho* mutations tend to increase GE due to enhanced read-through activity (Freddolino et al. 2012; Haft et al. 2014). We therefore wondered whether our *rho* mutations might have the same effect.

To address this hypothesis, we first focused on genome-wide expression data in genes. As a baseline, we compared expression data between REL1206 at 37.0˚C and 42.2˚C; this comparison provides insight into the physiological effects of high temperature acclimation without accompanying adaptive genetic change. Similar to Rodriguez-Verdugo et al. (2016), we found that ~1700 of 4204 coding regions were differentially expressed ($q < 0.001$) during thermal stress, with a tendency toward lower expression at 42.2 ˚C (972 down-regulated genes vs. 695 up regulated genes; binomial $p = 1.243 \times 10^{-11}$) (fig. 4*A*). Given this baseline, we then assessed the effect of *rho* mutants relative to REL1206 at 42.2˚C. GE was significantly biased toward higher expression in each of the mutants relative to REL1206 at 42.2˚C at $q < 0.001$ (Figure 4*A*; binomial, $p = 1.41 \times 10^{-10}$ (*I15N*), $p = 2.13 \times 10^{-6}$ (*I15F*), $p = 8.6 \times 10^{-12}$ (*A43T*), $p = 2.2 \times 10^{-16}$ (*T231A*)). In effect, thermal stress led to lower GE at 42.2˚C relative to 37.0˚C, and the four mutations reversed that trend.

The picture was, however, different for intergenic regions (IRs). Among 2306 IRs (supplementary table S5, Supplementary Material online) we found more were up-regulated in REL1206 at 42.2˚C relative to 37.0˚C (212 up regulated IR vs. 164 down; regulated; binomial $p = 0.015$) (fig. 4*B*). This up-tick in the expression of IRs could

indicate more transcriptional read-through under thermal stress. This trend was reversed by *rho* mutations, such that differentially expressed IRs tended to have *lower* expression in the mutants relative to REL1206 at 42.2˚C (fig. 4*B*; binomial $p = 0.0016$ (*I15N*), $p = 0.0001$ (*I15F*), $p = 0.04164$ (*A43T*)). This effect was consistent among all four mutants (although not statistically significant in *T231A*), raising the possibility that our *rho* mutations enhance termination efficiency at 42.2˚C.

Rho-terminated regions: The problem with a genome-wide analysis is that the effects are a mix of direct (primary) effects of the *rho* mutation and indirect (secondary) effects on downstream genes. To better examine the primary effects of *rho* mutants, we focused on a set of 183 *Rho*-dependent terminated regions defined by Peters et al. (2009) and previously used to analyze read-through activity of a *rho* mutation (Haft et al. 2014) (supplementary table S6, Supplementary Material online). The size of these regions spanned from 100 to ≈ 4000 bp, with located at the 3' end of coding regions (intergenic) and the other half within genes (intragenic). For each of the 183 regions, we calculated γ, the $\log_2$ ratio of normalized RNAseq counts between each mutant at 42.2˚C and REL1206 at 42.2˚C (see Materials and Methods). Under the null hypothesis that the mutant and ancestor have similar termination properties, we expected a random distribution of γ, such that half of the 183 regions have $\gamma > 0.0$ and half have $\gamma < 0.0$. If mutations enhanced termination, we expected significantly more than 50% of the regions to have $\gamma < 0.0$. In contrast, enhanced read-through should result in >50% of regions having $\gamma > 0.0$ (Haft et al., 2014).

Figure 5 plots γ for each of 183 regions for each of the mutants, and it provides the basis for three conclusions. First, both *I15N* and *I15F* yield evidence of read-through

activity, with 110 and 118 of 183 regions having $\gamma > 0.0$ (binomial, $p < 0.01$). Second, neither *A43T* nor *T231A* yielded evidence of altered termination properties, because they had 91 and 90 regions of $\gamma > 0.0$, neither of which deviates from the null expectation of 91.5 regions ($p \geq 0.88$). Third, these results imply that the four mutations have heterogeneous effects on transcription termination.

Mutations and stability of the Rho protein: In addition to read-through, we evaluated the predicted folding stability of the Rho protein. Our principal goal was to assess whether changes in fitness correlated with the direction and magnitude of shifts in the free energy of protein folding stability, as is expected under some evolutionary models (DePristo et al. 2005; Wylie and Shakhnovich 2011). We hypothesized that $\overline{w}_r$ from specific mutations correlates with the difference of the free energy of folding between ancestral and mutant proteins (ΔΔG) at 42.2° C. To determine the free energy of folding (ΔG), we applied an approach that combines classical Molecular Dynamics (MD) with the FoldX algorithm (see Materials and Methods) and then measured the change in folding stability (ΔΔG) induced by each of the four *rho* mutations. Positive values of ΔΔG suggest a decrease either in the stability of the protein fold or in the ability of the protein to fold relative to the ancestor; negative values indicate increased stability (Guerois et al. 2002; Miller et al. 2016). Three of four *rho* mutations (*I15F*, *A43T* and *I15N*) were destabilizing at 42.2˚C, with ΔΔG values ranging between 0.50 to 1.22 kcal/mol (table 4). In contrast, the *T231A* substitution mildly increased folding stability at 42.2˚C (ΔΔG=-0.06). Although our sample size was small ($n = 4$), there was no clear trend between $\overline{w}_r$ (table 4) and ΔΔG.

Our four engineered mutations were found in >1 of our EME populations, but additional *rho* mutations were found in single populations and were putatively adaptive (Tenaillon et al. 2012). We also estimated ΔΔG for 16 additional *rho* mutations from our EME, for a total of 20 ΔΔG estimates (table 4). Of these, 17 of 20 ΔΔG estimates were >0.00, indicative of destabilizing mutations. Among this larger data set, none of our four engineered mutations were extreme in neither range nor average value (*t*-test, $p = 0.57$). Finally, we also assessed ΔΔG trends in structural features. For example, the *I15N*, *I15F* and *A43T* mutations were within α-helices, so we contrasted their ΔΔG estimates against the estimates of other α-helix mutations; there were no difference in average ΔΔG values (t-test; $p = 0.98$). Similarly, we found no obvious differences for ΔΔG in α-helices vs. β-sheets (t-test; $p = 0.45$) or for defined regions of function (e.g., NHB, and CSD-like RBD; see table 4) vs. non-defined regions ($p = 0.36$). Hence the only consistent effect of mutations was toward destabilizing the wild type Rho protein at high temperature.

DISCUSSION

This study was designed to compare fitness and phenotype, as measured by gene expression, among *rho* mutants and also between *rho* and *rpoB* mutants. These comparisons have relevance for understanding the dynamics of adaptation, because each of the eight mutations were found in more than one population of our EME (Tenaillon et al. 2012) and were thus beneficial under the widely held criterion of parallelism (Christin et al. 2010). In addition, *rho* and *rpoB* represent two distinct, largely (but not completely) exclusive genetic pathways to adaption within our EME. The existence of two distinct pathways provides a rare opportunity to address fundamental questions in evolutionary

biology, including the distinction between passengers and drivers, the phenotypic consequences of alternative adaptive pathways and potential molecular mechanisms of adaptation.

Passengers, Drivers and Positive Epistasis

We have measured the relative fitness of eight mutants -- four each for *rho* and *rpoB* (table 1) -- and also measured GE for each mutant. Of the eight mutations, six have relative fitnesses ($\overline{w}_r$) significantly > 1.00, but two *rho* mutations have $\overline{w}_r \leq 1.0$ and thus confer no apparent advantage under thermal stress (table 1) (Rodriguez-Verdugo et al. 2014). Of the six advantageous mutations, the three *rpoB* codon *572* mutations are unambiguous drivers of adaptive change, both because they have high fitness values and because they often fix rapidly in experimental populations (Rodriguez-Verdugo et al. 2013).

Based on our estimates of $\overline{w}_r$, the two mutations in *rho* codon *I15* are unlikely to be drivers as solo mutations. They thus illustrate the difficulty of discriminating passengers from drivers, because *I15N* was one of the two most common nucleotide substitutions in the entire EME, suggesting it confers a substantial benefit in the proper genetic background. But what is the proper background? Two possibilities stand out: mutations in the *cls* gene or in the *iclR* gene. Knock-downs or knock-outs of the *cls* gene, which encodes a cardiolipin synthase, are present in 14 of 17 the clones with *rho I15F* and *I15N* mutations; mutations within *iclR*, a transcriptional repressor, are in 10 of 17 of clones; and either *cls* or *iclR* was mutated in all 17 clones (Tenaillon et al. 2012). Because *iclR* affects transcription, it is not difficult to surmise functional interactions

with Rho, especially because *rho* mutations are known to interact epistatically with mutations in other genes -- such as *rpsL*, *metJ* and *rpsQ* -- that have a role in transcription or translation (Freddolino et al. 2012; Haft et al. 2014). Both *rpsL* and *rpsQ* are critical to the 30S ribosomal unit (Cabezon et al. 1975), and *metJ* represses the expression of genes involved in methionine biosynthesis (Marincs et al. 2006).

In contrast to the *rho I15* mutations, can we conclude that the *A43T* and *T231A* mutations are drivers? On the basis of $\overline{w}_r$, they appear to be capable of driving adaptive change as solo mutations, but there is a twist: they too, are often found in *cls* and *iclR* backgrounds. All three A43T clones and one of two T231A clones included both *cls* and *iclR* mutations, again implying the possibility of positively epistatic interactions. Potential functional interactions between *cls* and *rho*, if any, are unknown. The *cls* gene has not been identified as *rho*-terminated (Peters et al. 2009) and is not differentially expressed in any one of our four *rho* mutants. One possible - but purely speculative - interaction between *cls* and *rho* is through membrane signaling. In addition to its role in membrane structure, cardiolipin plays a key role in recruiting signaling molecules to the membrane (Romantsov et al. 2007), and *E. coli* can perceive alterations in membrane fluidity and transfers that information to signal-transduction pathways that alter gene expression (Los and Murata 2004; Los et al. 2013). The potential for, and causes of, interactions between *rho* and *cls* require further work, including understanding the fitness effects of solo *cls* and *iclR* mutations under thermal stress, their epistatic fitness effects when coupled with *rho*, and the population dynamics among mutations in these genes over the time course of our EME.

Alternative adaptive pathways, gene expression phenotypes and pleiotropy

Although it remains difficult to distinguish passengers from drivers for *rho* mutations, there are clear similarities between the *rho* and *rpoB* adaptive pathways. One such similarity is that mutations in both genes act in part by restoring GE from the stressed condition (in this case, REL1206 under thermal stress at 42.2˚C) back toward the wild type state (in this case, REL1206 at 37.0˚C). We have detected this restoration both as a shift in PCA space (fig. 4) and as the dominant directional shift among individual genes (table 3). Several previous studies have also shown that adaptation proceeds through phenotypic restoration (Fong et al. 2005; Carroll and Marx 2013; Sandberg et al. 2014; Carroll et al. 2015; Hug and Gaut 2015; Rodriguez-Verdugo et al. 2015), but our study is unique in establishing a quantitative correlation between restoration and $\overline{w}_r$. There is, however, an interesting counter-example. In their study of *Methylobacterium extorquens*, Carroll and Marx (2013) have shown that adaptation was accompanied by the restoration of ancestral GE levels, just as we find here. In their case, growth rates (a measure of absolute fitness) were not correlated with restoration but with reinforcement (i.e., the number of genes that exaggerate the acclimation response to physiological stress). We do not find a correlation between the number of reinforced genes and $\overline{w}_r$ (r^2 = -0.48; p= 0.24).

A second similarity is that the set of genes affected by mutations in *rho* and *rpoB* overlap to a remarkable extent (fig. 2*B*). There are, of course, caveats to this observation. To compare the *rho* and *rpoB* pathways in this way, we have taken the union of DEGs among *rho* and *rpoB* mutants. In reality, different mutations are associated with different DEG sets (fig. 2*A*). It is also important to recognize that the number of DEGs is an

imperfect metric, because it varies with statistical power and can be non-transitive across pairwise comparisons. Nonetheless, the overarching impression is that the alternative adaptive pathways defined by *rho* and *rpoB* may traverse broadly similar phenotypic paths.

The common set of genes affected by the *rho* and the *rpoB* mutations provides insights into the physiological response to thermal stress. Mutations in *rho* and *rpoB* mutations up-regulate functions related to cellular growth, including purine and pyridine biosynthesis, amino acid biosynthesis, ribosome biogenesis, nitrogen assimilation, translation and the regulation of translation. Mutations in *rho* and *rpoB* also down-regulate a common set of genes, particularly those associated with carbohydrate transport and amino acid degradation. The shared patterns between *rho* and *rpoB* mutations are consistent with the view that long-term acclimation to thermal stress leads to the down-regulation of growth-related pathways, potentially resulting in energy conservation and enhanced survival (Rodriguez-Verdugo et al. 2015) and that adaptation adjusts to support growth in the stressful environment (Gunasekera et al. 2008).

Although there is a considerable overlap of genes affected by *rho* and *rpoB* mutations, some functional categories were altered predominantly by one of the two pathways. For example, *rho* mutants are enriched for DEGs that respond to temperature stimuli, such as *clpB*, *dnaK*, *dnaJ*, *degP* and *grpE*. For reasons not fully understood, the *rpoB* mutations do not up-regulate these temperature response genes (Rodriguez-Verdugo et al. 2015). In contrast, *rpoB* but not *rho* mutations enhance the expression of iron transport and flagellum biogenesis. This last observation is particularly interesting, because the up-regulation of flagellar genes is unlikely to have been adaptive in our EME

and because we have shown that these are later down-regulated by subsequent adaptive mutations in some EME populations (Rodriguez-Verdugo et al. 2015). This example of flagellar genes raises an important point: for all of these gross-scale effects on GE, it is difficult to distinguish between directly adaptive shifts vs. pleiotropic shifts that are neutral or perhaps even detrimental.

Mutations, Mechanisms and Negative Epistasis

Genome sequencing uncovers the type and potential function of adaptive mutations in EMEs (Long et al. 2015; Tenaillon et al. 2016), but the gap from 'potential function' to an established molecular mechanism remains a chasm for all but a few exceptions [e.g., (Blount et al. 2012; Quandt et al. 2015; Utrilla et al. 2016)]. With respect to our thermal stress EME, we are still uncertain about the effects of mutations on RNAP and Rho protein structure, the resulting alteration(s) of molecular mechanism, and the cause of negative epistasis between mutations in the two genes. It is known, however, that the speed of RNAP increases with increasing temperature (Ryals et al. 1982). We have therefore hypothesized that adaptive *rpoB* mutations slow RNAP transcription under thermal stress (Rodriguez-Verdugo et al. 2014), leading to increased termination efficiency (Jin et al. 1988).

Termination efficiency may require a balance (or 'kinetic coupling') between Rho and RNAP activity (Jin et al. 1992). Under this model, when RNAP increases in speed, Rho must match the increase to maintain efficient termination. We predicted, then, that adaptive *rho* mutants translocate along RNA more quickly, leading to increased termination efficiency relative to REL1206 at 42.2˚C. Yet, this is not what we found

based on direct assays of Rho mutant function (fig. 5). Instead, the two mutations with improved fitness (*A43T* and *T231A*) had no detectable effect on termination, and the two mutations with lower fitness (*I15F* and *I15N*) exhibited enhanced read-through (i.e, weaker termination). These results suggest either that the relationship between Rho and RNAP is more complex than our predictions encompass or that, for reasons unknown, the previously defined *rho* termination sites are not representative in our experimental conditions. We note that several mutations in *rho* have been shown to increase read-through (Haft et al. 2014; Martinez et al. 1996; Freddolino et al. 2012), but at least one mutant increases termination efficiency (*L3F*). Like our *I15* mutations, *L3F* is located in the positively charged N-terminal helix bundle (NH) (fig. 1) (Mori et al. 1989).

Our read-through results have at least four implications. First, they imply that a shift in Rho termination efficiency – at least as measured here – is not a requirement for adaptation via the *rho* pathway. Second, they suggest the possibility that the lower relative fitnesses of *I15N* and *I15F* could be a function of their decreased termination efficiency (Table 1). Third, they unfortunately provide few insights into structure-function relationships, because the free energy of folding ($\Delta\Delta G$) bears no discernible relationship with read-through properties or with fitness, for that matter, despite many predictions regarding the latter (DePristo et al. 2005; Bershtein et al. 2006; Tokuriki and Tawfik 2009; Wylie and Shakhnovich 2011; Bershtein et al. 2012; Serohijos and Shakhnovich 2014). The only discernible structural pattern was a tendency toward destabilization of the protein, which is not surprising given that ~70% of all possible mutations on globular proteins yield $\Delta\Delta G > 1$ (Guerois et al. 2002; Schymkowitz et al. 2005).

Finally, the results have implication for the nature of negative epistasis between *rho* and *rpoB*. We have postulated that the negative epistasis between *rho* and *rpoB* mutations in our EME results from altered kinetic coupling, because mutations in both genes 'over-tune' termination efficiency (Rodriguez-Verdugo et al. 2014). At present, there is little direct evidence for this model. And yet, there is a remarkable genome-wide result: Even though none of the *rho* mutations produced obvious biases toward increased efficiency at known termination sites, the genome-wide effect of each of our *rho* mutations was to reduce expression in IR regions (fig. 4). The reduction of expression in IRs is potentially consistent with our hypothesis of enhanced termination efficiency in *rho* mutants, but this genome-wide data includes genes directly terminated by *rho* and genes that may be affected indirectly via downstream effects. As a consequence, we are at a loss about the mechanism by which this enhancement of genome-wide termination occurs.

Concluding Remarks

Although a direct link to function remains elusive for our set of *rho* and *rpoB* mutations, the comparison of eight mutations into the REL1206 ancestral background yields at least three fundamental insights into *E. coli* adaptation to thermal stress. First, we have verified that one of the two most common single nucleotide mutations from our experiment (i.e., *rho I15N*) is unlikely to confer adaptive benefits as a solo mutation, illustrating both the difficulty of discriminating drivers from passengers -- even for the most obvious of candidates -- and the potential importance of epistatic interactions. Second, we have shown that fitness gains are correlated with gene expression, particularly the degree to which expression is restored to that of pre-stress physiology.

Third, we have shown that the two alternative pathways affect a largely overlapping set of genes. The mapping of genotypes to phenotypes on fitness landscapes remains a central topic in biology (de Visser and Krug, 2014). Here, we show that mutations in two genes that typify alternative adaptive pathways climb similar – but by no means identical – pathways on the adaptive landscape.

MATERIALS and METHODS

The construction of *rho* mutants

Mutations were introduced into the REL1206 ancestral strain using the pJk611 recombineering plasmid, as previously described (Rodriguez-Verdugo et al. 2014). Briefly, the pJk611 plasmid was introduced into REL1206, electroporating 2 µl of plasmid (80 ng of plasmid in total) into 50 µl of competent cells using an Eppendorf Electroporator 2510 set at 1.8 kV. After electroporation, 1 ml LB was added to the electroporated cells, and they were incubated at 30°C for 2 h under constant shaking (120 rpm). Thereafter 100 µl of cells were plated on LB agar plates containing 100 µg/ml ampicillin to select ampicillin-resistant transformants. The ancestral strain carrying the pJk611 plasmid was grown overnight at room temperature (~20°C) in 25 ml LB with 100 µg/ml ampicillin and 1mM $_{L}$-arabinose (Sigma) until it reached and OD_{600} of 0.6. Electrocompetent cells were made by washing the culture five times with ice-cold water.

To construct each mutant, two oligos of 70 bp were electroporated into cells (supplementary table S7, Supplementary Material online); one oligo introduced the desired single mutation in *rho* and the second oligo incorporated the change that produces an Arabinose positive (Ara+) phenotype. Two µl of each oligo (10 µM) was electroporated into 50 µl of cells. After electroporation 1 ml LB was added, and the cells

were incubated at 37.0˚C for 3 h with shaking, followed by 500 µl spread onto minimal medium agar (MA) plates supplemented with L-arabinose. The plates were incubated 48 h at 42.2ºC, after which 188 single colonies were selected and grown on TA agar plates overnight at 37.0˚C. We subsequently performed colony PCR, followed by Sanger sequencing to amplify the *rho* gene fragment surrounding the putative mutation (supplementary table S7, Supplementary Material online). PCR conditions were 10 min at 94˚C followed by 30 cycles of 20 s at 94˚C, 30 s at 60˚C, and 1 min at 68˚C ending with an extension step of 5 min at 68˚C. For each mutation, we isolated multiple lines with the mutation of interest, such that biological replicates were independently derived.

Relative fitness competitions

We estimated the relative fitness (*w*) of clones by competing them against the ancestral REL1206 strain. Briefly, both competitors were revived in LB broth and grown separately overnight at 37.0ºC under constant shaking (120rpm). After cultures were revived, they were subcultured by 1:10000 dilution in 10 ml DM25 media for 24 hours at 37.0ºC and 120 rpm and inoculated 1:100 into 9.9 ml of fresh DM25 and incubated at 42.2ºC for another day in order to acclimate cells to high temperature. Although the clones had evolved at 42.2°C in DM25, it is common practice in thermal stress studies to allow clones to recover from freezing under less stressful conditions (Bennett and Lenski 1993; Rodriguez-Verdugo et al. 2014).

Once revived, the two competitors were mixed at a 1:1 volumetric ratio and diluted 100-fold into 10 ml of fresh DM25 and incubated one day at the assay temperature. The initial and final densities were estimated by plating the culture onto

tetrazolium-arabinose (TA) agar plates, counting the number of Arabinose+ colonies of the REL1206 ancestor and the number of Arabinose- clones of the mutants and evolved clones. Relative fitness assays were performed on three biological replicates for each mutation and for three technical replicates for each biological replicate; thus each mutation was tested against the ancestor a total of nine times.

ANOVA analyses were based on the *avo* package in R. With ANOVA, we first tested for the effects specific to different biological replicates within a mutation using a model of Fitness~Mutant*Replicate. Given that there were no differences among biological replicates, the model was simplified to Fitness~Mutant.

RNA Harvesting and Sequencing

To harvest cells for RNA sequencing (RNAseq), we revived cells as described above and then cultivated populations at 42.2˚C until the mid-exponential phase of population growth. The mid-exponential phase was determined by electronic counts using an electronic particle counter (Coulter Counter model Multisizer 3 equipped with a 30 µm diameter aperture tube). For each of the single *rho* mutations, at least two biological replicates were grown to mid-exponential for RNA harvesting.

To harvest RNA, 80 ml of culture from the mid-exponential phase was filtered through a 0.2 µm cellulose membrane (Life Science, Germany). Then cells were washed with Qiagen RNA-protect Bacteria Reagent and pelleted for storage at -80˚C prior to RNA extraction. Cell pellets were thawed and incubated with lysozyme (Sigma-Aldrich), for 5 minutes at room temperature ~20˚C. Total RNA was isolated and purified using RNeasy Mini Kit (Qiagen). An on-column DNase treatment was performed for 30 min at

room temperature. RNA quality was assessed by running an Agilent RNA-Nano chip on a bioanalyzer. rRNA was depleted using the Ribo-Zero^{TM} rRNA Removal kit for Gram-Negative Bacteria (Epicentre Biotechnologies, Medion, WI, USA). cDNA libraries were constructed using TruSeq RNA v2 kit (Illumina, San Diego, CA, USA). After amplification, the samples were quantified by qPCR using the KAPA Library Quantification Kit code KK4822 for the Illumina Genome Analyzer platform (Kapa Biosystems, Wilmington, MA, USA). Libraries were sequenced on an Illumina Hiseq 2000 platform to generate 100 bp single-end reads.

RNAseq data for *rpoB* mutations and the REL1206 ancestor were generated previously, using identical protocols (Rodriguez-Verdugo et al. 2015).

RNAseq data and DEG analysis

We gathered three sets of RNAseq data. First, we generated data for four *rho* mutants, with two replicates each. Second, we used the REL1206 data from Rodriguez-Verdugo et al. (2015), representing states of both thermal stress (42.2˚C) and normal laboratory conditions (37.0˚C). Finally, we also used data from *rpoB* single mutants from (Rodriguez-Verdugo et al. 2015). The *rpoB* data can be accessed at the NCBI's Sequence Read Archive (SRA) under the BioProject Number PRJNA291128. The *rho* data generated for this study have BioProject Number PRJNA339971 and SRA accession number SRP083065.

For all RNAseq data, sequence reads were filtered to a quality cut-off of 20. Filtered reads were mapped to the *E. coli* B REL606 reference genome (NCBI: NC_012967.1 GI:254160123) using BWA aligner (Li and Durbin 2010) using default

parameters. Statistics about mapping are provided in supplementary table S8 of the Supplementary Material online. Unique matching reads to the 4204 annotated coding regions were retained for further analyses, as were reads that mapped to 2306 non-coding intergenic regions (IRs) of the REL606 reference (supplementary table S5, Supplementary Material online). IRs were defined as the sequence that extends from the 3' end of a coding region to its neighboring coding region. IRs were analyzed if they exceeded 1 bp in length and had a minimum coverage of at least 1 read in at least one of the ancestors and *rho* mutants replicates. Note, however, that similar trends were obtained when IRs were defined as ≥100 bp in length.

Coding regions and IRs counts were obtained from BAM files using HTSeq-count tool (Anders et al. 2015) from HTSeq, at http://www.huber.embl.de/users/anders/HTSeq/doc/count.html using intersection-nonempty counting mode. Differential gene expression analysis was performed with DESeq version 1.18.0 (Anders and Huber 2010). The FDR was determined using Benjamini-Hochberg adjusted *p*-values. Genes with *q* values <0.001 were considered as differentially expressed.

Gene Ontology enrichment analysis was performed through the Gene Ontology Consortium (http://geneontology.org/page/go-enrichment-analysis) with Bonferroni correction for multiple testing. Principal component analysis (PCA) was applied to RNAseq data using the *plotPCA* function in DESeq version 1.18.0 (Anders and Huber, 2010) with the default parameters (`ntop=500`).

Identification and analysis of rho-terminated genome regions

To evaluate the effects of each *rho* mutation on the transcription termination activity, we evaluated the gene expression of a set of previously defined Rho dependent terminated regions (Peters et al. 2009). Because these Rho-terminated regions were originally defined in *E. coli* K-12, it was necessary to first determine their coordinates in our REL606 genome. To do so, we utilized the start positions and lengths of the regions as defined by Peters et al. (2009) located the closest gene to the start of the Rho-terminated region in *E. coli* K-12, noting the position of the Rho-terminated region relative to the closest gene. We then located this gene in the REL606 genome and used the relative position information from K-12 to assign our Rho-terminated regions in REL606. In the few cases where gene orientation was not preserved across strains, coordinates were determined in an orientation-dependent manner (supplementary table S6, Supplementary Material online). The number of read counts of each region was obtained from BAM files using the HTseq-count tool (Anders et al. 2014) and normalization was performed with DEseq based on a complete gff file that included the putatively Rho-terminated regions (Anders and Huber 2010). Normalized gene counts were averaged among replicates to calculate the ratio γ , the $\log_2$ ratio of mutant to ancestral counts.

Directions of evolutionary change

GE was classified into one of four directions of evolutionary change, as described previously (Carroll and Marx 2013; Rodriguez-Verdugo et al. 2015). The four directions classify GE in a mutant relative to REL1206 at 42.2˚C and to REL1206 at 37.0˚C. Briefly, a gene had *novel* expression if the mutant differed significantly ($q < 0.001$) from

both ancestral treatments (42.2˚C and 37.0˚C), but GE did not differ between the two ancestral treatments (i.e. Anc42≈Anc37 & Anc37≠Mutant & Anc42≠Mutant). *Restoration* occurred when the mutant GE levels fell between those of the two ancestral treatments and when GE differed significantly between the ancestral treatments and also between the mutant and REL1206 at 42.2˚C [i.e., (Anc42>Mut>Anc37 or Anc42<Mut<Anc37) & Anc42 ≠Anc37 & Anc42 ≠Mutant]. *Reinforced* expression occurred when GE differed significantly between the ancestral treatments and between the mutant and REL1206 at 42.2˚C, but GE in the mutant did not fall between the ancestral treatments [i.e., Anc42 ≠Anc 37 & Anc42 ≠Mutant & (Mut>Anc42>Anc37 or Mut<Anc42<Anc37)]. Finally, *unrestored* expression occurred when GE did not differ significantly between the mutant and REL1206 at 42.2˚C, but the two ancestral treatments differed significantly from each other (i.e., Anc42 ≠Anc 37 & Mut≈Anc42).

Estimates of free energy of folding.

We calculated two values: ΔG, the free energy of folding, and ΔΔG, the difference of ΔG between the REL1206 protein and mutant proteins. We focused on folding free energy differences of the Rho hexamer. To calculate ΔΔG, we used an approach that combines classical molecular dynamics with FoldX (Miller et al. 2016; Guerois et al. 2002), which improves the calculation of ΔΔG compared to using FoldX alone (Miller et al. 2016). To perform the calculations, we downloaded the X-ray crystal structure of Rho transcription protein bound to RNA and ADP-BeF3 from the Protein Data Bank (PDB id: 3ICE) (Thomsen and Berger 2009). The file was modified to remove all but the six chains of Rho hexamer. Out of all the six chains in the X-ray crystal

structure, chain C had least missing residues. MODELLER v9.15 software (Sali and Blundell 1993) was then used to rebuild the coordinates of the missing residues for chain C and complete chain C was used as a template to fill gaps in the other chains of the hexamer. The final hexamer structure had 2484 residues.

The software package GROMACS 5.0.7 (Van Der Spoel et al. 2005) was used for 10 ns long Rho hexamer Molecular Dynamics simulations with the AMBER99SB forcefield (Hornak et al. 2006). We followed the standard energy minimization, thermalization and equilibration protocol (Miller et al. 2016) before carrying out the final production run. During the 10 ns production simulation snapshots were saved every 1 ns giving 10 snapshots of Rho hexamer protein for the further analysis.

Fold X was used to analyze the starting structure and all 10 snapshots (total 11 snapshots) captured during molecular dynamics simulations (Guerois et al., 2002). Before, analyzing all 11 snapshots using FoldX, we first extracted chain C from each of the snapshots and later only the monomer structures - i.e. chain C - were used for folding free energy calculations. We started by minimizing monomer structure using the *RepairPDB* command of FoldX to further minimize the potential energy. All single amino acid mutations were then generated using *BuildModel*. Finally, protein folding stabilities were estimated using *Stability* on the monomer structure. For each mutation we then estimated $\Delta\Delta G$ by averaging across all 11 individual snapshots estimates. The molecular graphics package VMD was used to produce fig. 1 (Humphrey et al. 1996).

ACKNOWLEDGMENTS

We thank R. Gaut and P. MacDonald for technical assistance. A.G.G. and A.R.V. were supported by University of California Institute for Mexico and the United States-Consejo Nacional de Ciencia y Tecnología (Mexico) Post-doctoral and Doctoral fellowships respectively. S.H. was supported by the National Institute of Biomedical Imaging and Bioengineering, National Research Service Award EB009418 from University of California, Irvine, Center for Complex Biological Systems. JSP was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number P20GM104420. Computer resources to perform the Molecular dynamic simulations were provided by the Institute for Bioinformatics and Evolutionary Studies Computational Resources Core sponsored by the National Institute of Health (P30 GM103324) at University of Idaho. BSG was supported by a fellowship from the Borchard Foundation, and the work was also supported by the National Science Foundation grant DEB-0748903.

REFERENCES

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106.

Anders S, Pyl PT, Huber W. 2015. HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166-1669.

Bennett AF, Lenski RE. 1993. Evolutionary adaptation to temperature II. Thermal niches of experimental lines of *Escherichia coli*. *Evolution* 47:1-12.

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.

Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489:513-518.

Cabezon T, Faelen M, De Wilde M, Bollen A, Thomas R. 1975. Expression of ribosomal protein genes in *Escherichia coli*. *Mol Gen Genet.* 137:125-129.

Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E. 2008. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* 320:935-938.

Carroll SM, Chubiz LM, Agashe D, Marx CJ. 2015. Parallel and divergent evolutionary solutions for the optimization of an engineered central metabolism in *Methylobacterium extorquens* AM1. *Microorganisms* 3:152-174.

Carroll SM, Marx CJ. 2013. Evolution after introduction of a novel metabolic pathway consistently leads to restoration of wild-type physiology. *PLoS Genet.* 9:e1003427.

Charusanti P, Conrad TM, Knight EM, Venkataraman K, Fong NL, Xie B, Gao Y, Palsson BO. 2010. Genetic basis of growth adaptation of *Escherichia coli* after deletion of *pgi*, a major metabolic gene. *PLoS Genet.* 6:e1001186.

Christin PA, Weinreich DM, Besnard G. 2010. Causes and evolutionary significance of genetic convergence. *Trends Genet.* 26:400-405.

Ciampi MS. 2006. Rho-dependent terminators and transcription termination. *Microbiology* 152:2515-2528.

Conrad TM, Joyce AR, Applebee MK, Barret CL, Xie B, Gao Y, Palsson BO. 2009. Wole-genome resequencing of *Escherihia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biol.* 10:R118.

Conrad TM, Lewis NE, Palsson BO. 2011. Microbial laboratory evolution in the era of genome-scale science. *Mol Syst Biol.* 7:509.

D'Heygère F, Rabhi M, Boudvillain M. 2013. Phyletic distribution and conservation of the bacterial transcription termination factor Rho. *Microbiology* 159:1423-1436.

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

de Visser JAGM and Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet.* 15:480-490.

Dillon MM, Rouillard NP, Van Dam B, Gallet R, Cooper VS. 2016. Diverse phenotypic and genetic responses to short-term selection in evolving *Escherichia coli* populations. *Evolution* 70:586-599.

Fong SS, Joyce AR, Palsson BO. 2005. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res.* 15:1365-1372.

Freddolino PL, Goodarzi H, Tavazoie S. 2012. Fitness landscape transformation through a single amino acid change in the rho terminator. *PLoS Genet.* 8:e1002744.

Goodarzi H, Hottes AK, Tavazoie S. 2009. Global discovery of adaptive mutations. Nat Methods. 6:581-583.

Guerois R, Nielsen JE, Serrano L. 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol.* 320:369-387.

Gunasekera TS, Csonka LN, Paliy O. 2008. Genome-wide transcriptional responses of *Escherichia coli* K-12 to continuous osmotic and heat stresses. *J Bacteriol*. 190:3712-3720.

Haft RJ, Keating DH, Schwaegler T, Schwalbach MS, Vinokur J, Tremaine M, Peters JM, Kotlajich MV, Pohlmann EL, Ong IM, Grass JA, Kiley PJ, Landick R. 2014. Correcting direct effects of ethanol on translation and transcription machinery confers ethanol tolerance in bacteria. *Proc Natl Acad Sci U S A.* 111:E2576-85.

Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BO. 2006. Comparative genome

sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet.* 38:1406-1412.

Herron MD, Doebeli M. 2013. Parallel evolutionary dynamics of adaptive diversification in Eshcerichia coli. *PloS Biol.* 11:e1001490.

Hollands K, Proshkin S, Sklyarova S, Epshtein V, Mironov A, Nudlet E, Groisman EA. 2012. Riboswitch control of Rho-dependent transcription termination. *Proc Natl Acad Sci U S A.* 109:5376-5381.

Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. 2006. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins: Struct Func Bioinf.* 65:712-725.

Hug SM, Gaut BS. 2015. The phenotypic signature of adaptation to thermal stress in *Escherichia coli*. *BMC Evol Biol.* 15:177.

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph.* 14:33-8, 27-8.

Jin DJ, Burgess RR, Richardson JP, Gross CA. 1992. Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. *Proc Natl Acad Sci U S A.* 89:1453-1457.

Jin DJ, Walter WA, Gross CA. 1988. Characterization of the termination phenotypes of rifampicin-resistant mutants. *J Mol Biol.* 202:245-253.

Kishimoto T, Iijima L, Tatsumi M, Ono N, Oyake A, Hashimoto T, Matsuo M, Okubo M, Suzuki S, Mori K. 2010. Transition from positive to neutral in mutation fixation along with continuing rising fitness in thermal adaptive evolution. *PloS Genet.* 6:e1001164.

Le Gac M, Cooper TF, Cruveiller S, Medigue C, Schneider D. 2013. Evolutionary history and genetic parallelism affect correlated responses to evolution. *Mol Ecol.* 22:3292-3303.

Lee YH, Helmann JD. 2014. Mutations in the primary sigma factor sigmaA and termination factor rho that reduce susceptibility to cell wall antibiotics. *J Bacteriol.* 196:3700-3711.

Li H, DurbinR. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.

Long A, Liti G, Luptak A, Tenaillon O. 2015. Elucidating the molecular architecture of adaptive via evolve and resequence experiments. *Nat Rev Gen.* 16:567-582.

Los DA, Mironov KS, Allakhverdiev SI. 2013. Regulatory role of membrane fluidity in gene expression and physiological functions. *Photosynth Res.* 116:489-509.

Los DA, Murata N. 2004. Membrane fluidity and its roles in the perception of environmental signals. *Biochim Biophys Acta.* 1666:142-157.

Marincs F, Manfield IW, Stead JA, McDowall KJ, Stockley PG. 2006. Transcript analysis reveals an extended regulon and the importance of protein-protein co-operativity for the Escherichia coli methionine repressor. *Biochem J.* 396:227-234.

Martinez A, Opperman T, Richardson JP. 1996. Mutational analysis and secondary structure model of the RNP1-like sequence motif of transcription termination factor Rho. *J Mol Biol.* 257:895-908.

Miller CR, Johnson EL, Burke AZ, Martin KP, Miura TA, Wichman HA, Brown CJ, Ytreberg FM. 2016. Initiating a watch list for Ebola virus antibody escape mutations. *PeerJ* 4:e1674.

Mori H, Imai M, Shigesada K. 1989. Mutant rho factors with increased transcription termination activities. II. Identification and functional dissection of amino acid changes. *J Mol Biol.* 210:39-49.

Peters JM, Mooney RA, Kuan PF, Rowland JL, Keles S, Landick R. 2009. Rho directs widespread termination on intragenic and stable RNA transcription. *Proc Natl Acad Sci U S A.* 106:15406-15411.

Peters JM, Vangeloff, AD, Landick R. 2011. Bacterial transcription terminators: the RNA 3'-end chronicles. *J. Mol. Biol.* 412:793-813.

Quandt EM, Gollihar J, Blount ZD, Ellington AD, Georgiou G, Barrick JE. 2015. Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. *eLife* 4:e09696.

Rodriguez-Verdugo A, Carrillo-Cisneros D, Gonzalez-Gonzalez A, Gaut BS, Bennett AF. 2014. Different tradeoffs result from alternate genetic adaptations to a common environment. *Proc Natl Acad Sci U S A.* 111:12121-12126.

Rodriguez-Verdugo A, Gaut BS, Tenaillon O. 2013. Evolution of Escherichia coli rifampicin resistance in an antibiotic-free environment during thermal stress. *BMC Evol Biol.* 13:50.

Rodriguez-Verdugo A, Tenaillon O, Gaut BS. 2015. First-Step Mutations during Adaptation Restore the Expression of Hundreds of Genes. *Mol Biol Evol.*

Romantsov T, Helbig S, Culham DE, Gill C, Stalker L, Wood JM. 2007. Cardiolipin promotes polar localization of osmosensory transporter ProP in *Escherichia coli*. *Mol Microbiol*. 64:1455-1465.

Rosenzweig F, Sherlock G. 2014. Experimental evolution: prospects and challenges. *Genomics*. 104:v-vi.

Ryals J, Little R, Bremer H. 1982. Temperature dependence of RNA synthesis parameters in *Escherichia coli*. *J Bacteriol*. 151:879-887.

Sali A, Blundell TL. 1993. Comparative protein modeling by satisfaction spatial restraints. *J Mol Biol*. 5:779-815.

Sandberg TE, Pedersen M, LaCroix RA, Ebrahim A, Bonde M, Herrgard MJ, Palsson BO, Sommer M, Feist AM. 2014. Evolution of *Escherichia coli* to 42˚C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. *Mol Biol Evol.* 31:2647-2662.

Saxer G, Krepps MD, Merkley ED, Ansong C, Deatherage Kaiser BL, Valovska MT, Ristic N, Yeh PT, Prakash VP, Leiser OP, Nakhleh L, Gibbons HS, Kreuzer HW, Shamoo Y. 2014. Mutations in global regulators lead to metabolic selection during adaptation to complex environments. *PLoS Genet.* 10:e1004872.

Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. 2005. The FoldX web server: an online force field. *Nucleic Acids Res.* 33:W382-8.

Serohijos AW, Shakhnovich EI. 2014. Merging molecular mechanism and evolution: theory and computation at the interface of biophysics and evolutionary population genetics. *Curr Opin Struct Biol*. 26:84-91.

Skordalakes E, Berger JM. 2003. Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* 114:135-146.

Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet.* 14:751-764.

Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, Wu GC, Wielgoss S, Cruveiller S, Médigue C, Schneider D, Lenski RE. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* 536:165-170.

Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. *Science* 335:457-461.

Thomsen ND, Berger JM. 2009. Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* 139:523-534.

Tokuriki N, Tawfik DS. 2009. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol.* 19:596-604.

Utrilla J, O'Brien EJ, Chen K, McCloskey D, Cheung J, Wang H, Armenta-Medina D, Feist AM, Palsson BO. 2016. Global Rebalancing of Cellular Resources by Pleiotropic Point Mutations Illustrates a Multi-scale Mechanism of Adaptive Evolution. *Cell Syst.* 2:260-271.

van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. 2005. GROMACS: Fast, flexible, and free. *J Comput Chem.* 26:1701-1718.

Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ. 1999. Different trajectories of parallel evolution during viral adaptation. *Science* 285:422-424.

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.

Table 1

Relative fitness measured at 42.2°C for four *rho* mutants and four *rpoB* mutants.

Gene[1]	Mutant	Mean	± 95% CI	*p*-value[2]
rho	*I15F*	1.008	0.117	0.837
rho	*I15N*	0.945	0.184	0.396
rho	*A43T*	1.237	0.093	< 0.05
rho	*T231A*	1.266	0.104	< 0.05
rpoB	*I572F*	1.155	0.047	< 0.05
rpoB	*I572L*	1.229	0.099	< 0.05
rpoB	*I572N*	1.184	0.067	< 0.05
rpoB	*I966S*	1.258	0.106	< 0.05

[1] *rho* data from this study, based on nine replicated measures. *rpoB* data include that from (Rodriguez-Verdugo et al. 2014) plus three additional replicates, for a total of six replicates

[2] $p < 0.05$ rejects the null hypothesis that $\overline{w}_r = 1.0$, indicating that the mutation is beneficial ($\overline{w}_r > 1.0$) relative to the REL1206 ancestor.

Table 2

Comparisons of gene expression between single *rho* mutants. The cells above the diagonal report the Pearson correlation coefficient in GE across all genic regions (*n*=4204). The cells below the diagonal indicate the number of genes significantly different via DEseq (*q* <0.001).

	I15N	***I15F***	***A43T***	***T231A***
I15N		0.991	0.962	0.955
I15F	1		0.969	0.964
A43T	149	190		0.992
T231A	419	552	16	

Table 3

The direction of GE for genes within *rho* and *rpoB* mutants.

	Restored[1]	**Unrestored**	**Reinforced**	**Novel**	**Total**[3]
rho I15F	762	886	19	4	1671
rho I15N	802	857	8	16	1683
rho A43T	1115	544	7	32	1698
rho T231A	1250	405	12	46	1713
rpoB I572F [2]	1092	644	1	59	1796
rpoB I572L [2]	563	1174	0	6	1743
rpoB I572N [2]	1165	569	3	53	1790
rpoB I966S [2]	1308	428	1	39	1776

[1] The number of genes was counted as restored, unrestored, reinforced and novel using the rules elaborated in the Material and Methods.

[2] Data are from (Rodriguez-Verdugo et al. 2015).

[3] The total tallies the number of genes that differ at $q < 0.001$ from either REL1206 at 37.0˚C or REL1206 at 42.2˚C or both. The total differs significantly between *rho* and *rpoB* mutations (t-test; $p = 0.0016$)

Table 4

Estimated ΔΔG values for mutations in *rho*

Mutation	Location[1]	ΔΔG at 42.2°C (kcal/mol) ± S.E.[2]
W.T.	--	0
I15N	α2 NHB *	1.22 ± 0.25
I15F	α2 NHB *	0.50 ± 0.35
R30C	Loop NHB *	0.17 ± 0.19
A43T	α3 NHB *	1.17 ± 0.39
A74T	Loop CSD-like RBD *	0.47 ± 0.24
E56A	β1 CSD-like RBD *	1.43 ± 0.34
K100T	β4 CSD-like RBD *	0.82 ± 0.82
N126I	α5 **	0.91 ± 0.22
A127T	α5 **	1.22 ± 0.68
N129D	α5 **	0.05 ± 0.16
Q189H	α7 **	-0.06 ± 0.48
V206A	β7 **	4.13 ± 0.26
I209F	β7 **	1.41 ± 0.91
V223G	α8 **	3.87 ± 0.21
A229S	β8 **	0.08 ± 0.36
T231A	Loop -CTD **	-0.06 ± 0.45
R238C	α9 **	1.38 ± 0.27
Q241L	α9 **	0.28 ± 0.26
A243V	α9 **	-0.59 ± 0.77
E244G	α9 **	2.26 ± 0.27
I247V	α9 **	1.58 ± 0.10
D359A	α13 **	0.38 ± 0.46
I386V	α16 **	0.80 ± 0.20
H388Y	α16 **	-0.38 ± 0.25

[1] Location of mutation in the Rho protein. NHB=N-terminal helix bundle; CSD-like RBD=cold shock domain like RNA-binding domain. * Mutations located in the NTD= N-terminal domain and **CTD= C-terminal domain.

[2] Errors are calculated with 95% confidence interval.

FIGURE LEGENDS

FIG. 1

A schematic of the Rho protein, showing its hexameric structure and the location of the single-mutations (red) produced for this experiment. Mutations in *I15* and *A43* are located in the N-terminal domain (NTD) (gray); the mutation in *A231* in the C-terminal domain (CTD) (green).

FIG. 2

(A) A Venn diagram of the number of genes that exhibit significant GE differentiation between *rho* mutants. *(B)* A Venn diagram of the number of genes that differ in expression between the four *rho* mutants and REL1206 at 42.2˚C and all four *rpoB* mutants and REL1206 at 42.2˚C. The *x*-axis represents restoration along PC1, with smaller values indicating greater restorative effects.

FIG. 3

A PCA plot of the RNAseq data, including replicated data from REL1206 at 37.0˚C (light blue dots), REL1206 at 42.2˚C (dark blue dots); replicated data from each of the four *rho* mutations (light and dark green dots); and replicated data from each of the four *rpoB* mutations (pink dots). The *x*-axis represents restoration along PC1, where smaller values reflect greater restorative effects.

FIG. 4

Genome-wide expression patterns of genic and intergenic regions at 42.2°C. *(A)* Histograms showing the number of coding regions that are up or down regulated compared to REL1206. Up or down regulation is relative to the first of the listed pair. For example, the first pair of histograms show that REL1206 at 42.2˚C has ~700 and ~1000 genes that are expressed at higher and lower levels, respectively, than REL1206 at 37.0˚C. The bars with three, two and one asterisks indicate $p < 0.0001$, $p < 0.001$ and $p < 0.01$ respectively, a dot indicates $p < 0.05$. *(B)* Similar histograms as in *A*, but based on intergenic regions (IRs).

FIG. 5

Analysis of gene expression in 183 Rho-terminated regions. Each graph represents one mutant, with histograms representing the $\log_2$ expression of mutant and Ancestor at 42.2˚C (i.e., γ), with the 183 regions ordered from highest to lowest γ value. The *p*-value represents a test of the random expectation that half of the regions have $\gamma > 0.00$.