Mutations that improve the efficiency of a weak-link enzyme are rare compared to adaptive mutations elsewhere in the genome.

Andrew B. Morgenthaler¹,², Wallis R. Kinney¹,², Christopher C. Ebmeier¹, Corinne M. Walsh³,⁴, Daniel J. Snyder⁴, Vaughn S. Cooper⁴, William M. Old¹, Shelley D. Copley¹,²

¹Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder, Boulder, United States
²Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, United States
³Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, United States
⁴Department of Microbiology and Molecular Genetics, and Microbial Genomic Sequencing Center, University of Pittsburgh, Pittsburgh, United States

Abstract

New enzymes often evolve by amplification and divergence of genes encoding enzymes with a weak ability to provide a new function. Experimental studies to date have followed the evolutionary trajectory of an amplified gene, but have not addressed other mutations in the genome when fitness is limited by an evolving gene. We have adapted Escherichia coli in which an enzyme’s weak secondary activity has been recruited to serve an essential function. While the gene encoding the “weak-link” enzyme amplified in all eight populations, mutations improving the new activity occurred in only one. This beneficial allele quickly swept the amplified array,
displacing the parental allele. Most adaptive mutations, however, occurred elsewhere in the
genome. We have identified the mechanisms by which three of the classes of mutations increase
fitness. These mutations may be detrimental once a new enzyme has evolved, and require
reversion or compensation, leading to permanent changes in the genome.

Introduction

The expansion of huge superfamilies of enzymes, transcriptional regulators, transporters,
and signaling molecules from single ancestral genes has been a dominant process in the
evolution of life (Bergthorsson et al., 2007; Chothia et al., 2003; Glasner et al., 2006; A. L.
Hughes, 1994; Ohno, 1970; Todd et al., 2001). The emergence of new protein family members
has enabled organisms to access new nutrients, sense new stimuli, and respond to changing
conditions with ever more sophistication (Bouquin & Khila, 2017; Conant & Wolfe, 2008; Nei &
Rooney, 2005; Reams & Neidle, 2004; Starr et al., 2017; Storz, 2016).

The Innovation-Amplification-Divergence (IAD) model (Figure 1) posits that evolution
of new genes by duplication and divergence begins when a physiologically irrelevant side
activity of a protein, known as a promiscuous activity (Copley, 2003, 2017), becomes important
for fitness due to a mutation or environmental change (Bergthorsson et al., 2007; Francino, 2005;
A. L. Hughes, 1994; Näsvall et al., 2012). Often this newly useful activity is inefficient, making
the enzyme the “weak-link” in metabolism. Gene duplication/amplification provides a ready
mechanism to improve fitness by increasing the abundance of a weak-link enzyme. If further
mutations lead to evolution of an enzyme capable of efficiently carrying out the newly needed
function, selective pressure to maintain a high copy number will be removed, allowing extra
copies to be lost and leaving behind two paralogs (or just one gene encoding a new enzyme if the original function is no longer needed).

While the IAD model provides a satisfying theoretical framework for the process of gene duplication and divergence, our understanding of the process is far from perfect. Although the signatures of gene duplication and divergence are obvious in extant genomes, we have little information about the genome contexts and environments in which new enzymes arose.

Laboratory evolution offers the possibility of tracking this process in real time. In a landmark study, Näsvall et al. used laboratory evolution to demonstrate that a gene encoding an enzyme with two inefficient activities required for synthesis of histidine and tryptophan amplified and diverged to alleles encoding two specialists within 2000 generations (Näsvall et al., 2012). However, this study followed only mutations in the diverging gene. In general, when an organism is exposed to a novel selection pressure that requires evolution of a new enzyme, any mutation – either in the gene encoding the weak-link enzyme or elsewhere in the genome – that improves fitness will provide a selective advantage.

We have generated a model system in *E. coli* for exploring the relative importance of mutations in a gene encoding a weak-link enzyme and elsewhere in the genome. ProA (*γ*-glutamyl phosphate reductase, Figure 2) is essential for proline synthesis in *E. coli*. ArgC (*N*-acetylglutamyl phosphate reductase) catalyzes a similar reaction in the arginine synthesis pathway, although the two enzymes are not homologous (Goto et al., 2003; Ludovice et al., 1992; Page et al., 2003). ProA can reduce *N*-acetylglutamyl phosphate, but its activity is too inefficient to support growth of a ΔargC strain of *E. coli* in glucose. However, a point mutation that changes Glu383 to Ala allows slow growth of the ΔargC strain in glucose. Enzymatic assays show that E383A ProA (ProA*) has severely reduced activity with *γ*-glutamyl semialdehyde
(GSA), but substantially improved activity with \(N\)-acetylglutamyl semialdehyde (NAGSA) (Khanal et al., 2015; McLoughlin & Copley, 2008). (It is necessary to assay kinetic parameters in the reverse direction because the substrates for the forward reaction are too unstable to prepare and purify.) Glu383 is in the active site of the enzyme; the change to Ala may create extra room in the active site to accommodate the larger substrate for ArgC, but at a cost to the ability to bind and orient the native substrate. The poor efficiency of the weak-link ProA* creates strong selective pressure for improvement of both proline and arginine synthesis during growth of \(\Delta\text{argC} \ E. \ coli\) on glucose as a sole carbon source.

We adapted eight replicate populations of \(\Delta\text{argC} \ proA^* \ E. \ coli\) in minimal medium supplemented with glucose and proline for up to 1000 generations to identify mechanisms by which the impairment in arginine synthesis could be alleviated. Our expectation that amplification of \(proA^*\) would be beneficial was borne out in all populations. However, whole-genome sequencing of the adapted populations showed that a mutation in \(proA^*\) occurred in only one population. Indeed, most of the adaptive mutations occurred outside of \(proA^*\). We have identified the mechanisms by which three common classes of mutations allow a relatively rapid increase in fitness while the more difficult process of improving the weak-link enzyme progresses. However, they increase fitness at a cost to presumably well-evolved functions.

Our results demonstrate that mutations elsewhere in the genome play an important role during the process of gene amplification and divergence when the inefficient activity of a weak-link enzyme limits fitness. Thus, the process of evolution of a new enzyme by gene duplication and divergence is inextricably intertwined with mutations elsewhere in the genome that improve fitness by other mechanisms.
Figure 1. The Innovation-Amplification-Divergence (IAD) model of gene evolution. A physiologically irrelevant side activity B of an enzyme may become physiologically relevant due to a mutation or environmental change. Gene amplification increases the abundance of the weak-link enzyme. Acquisition of mutations can improve the efficiency of the newly important activity B. Once sufficient B activity is achieved, selection is relaxed and extra gene copies are lost, leaving behind two paralogs.

Figure 2. Reactions catalyzed by ProA (γ-glutamyl phosphate reductase) in proline synthesis and ArgC (N-acetylglutamyl phosphate reductase) in arginine synthesis. E383A ProA (ProA*) can catalyze both reactions, albeit poorly.
Results

Growth rate of ΔargC proA* E. coli increased 3-fold within a few hundred generations of adaptation in M9/glucose/proline

We generated a progenitor strain for laboratory evolution by replacing argC with the kan^r antibiotic resistance gene, modifying proA to encode ProA*, and introducing a mutation in the -10 region of the promoter of the proBA operon that was previously shown to occur frequently during adaptation of the ΔargC strain (Kershner et al., 2016). We also introduced yfp downstream of proA* and deleted several genes (fimAICDFGH and csgBAC, which are required for the formation of fimbriae and curli, respectively (Barnhart & Chapman, 2006; Proft & Baker, 2009)) to minimize the occurrence of biofilms. We evolved eight parallel lineages of this strain (AM187, Table 1) in M9 minimal medium supplemented with 0.2% glucose, 0.4 mM proline, and 20 µg/mL kanamycin in a turbidostat to identify mutations that improve synthesis of arginine. We used a turbidostat rather than a serial transfer protocol because turbidostats can maintain cultures in exponential phase and thereby avoid selection for mutations that simply decrease lag phase or improve survival in stationary phase. Turbidostats also avoid population bottlenecks during serial passaging that can result in loss of genetic diversity.

Growth rate in each culture tube was averaged over each 24-hour period and was used to calculate the number of generations each day. Each culture was maintained until a biofilm formed (33-57 days, corresponding to 470-1000 generations). While it is possible to restart cultures from individual clones after biofilm formation, this practice introduces a severe population bottleneck. Thus, we decided to stop the adaptation for each population when a biofilm formed. For this reason, every population was adapted for a different number of generations.
Table 1. Strains used in this work.

| strain  | genotype                                                                 | notes                                                                 |
|---------|---------------------------------------------------------------------------|----------------------------------------------------------------------|
| AM008   | *Escherichia coli* BW25113; argC::kan′                                 | Keio strain (Baba et al., 2006)                                      |
| AM187   | AM008 + -45 C→T [M2 promoter mutation from (Kershner et al., 2016)] upstream of proB + A1148→C (changes Glu383 to Ala) in proA + yfp construct downstream of proBA consisting of (in order from 5’ to 3’) BBa_B0015 terminator, P3 promoter, synthetic RBS, yfp (see Materials and Methods); ΔfimAICDFGH; ΔcsgBAC | parental strain for adaptation, GenBank accession number CP037857 |
| AM209   | *Escherichia coli* BL21(DE3); argC::kan′; proA::cat                    | expression and purification of wild-type and mutant ProA*             |
| AM239   | AM187 + 58 bp deletion upstream of argB (pos. 4145856-4145913)             |                                                                      |
| AM241   | AM187 + C4145901→G (24 bp upstream of argB start codon)                  |                                                                      |
| AM242   | AM187 + C4145903→A (22 bp upstream of argB start codon)                  |                                                                      |
| AM243   | AM187 + C4145903→T (22 bp upstream of argB start codon)                  |                                                                      |
| AM244   | AM187 + C4145907→A (18 bp upstream of argB start codon)                  |                                                                      |
| AM245   | AM187 + 38 bp duplication upstream of argB (pos. 4145912-4145949)          |                                                                      |
| AM267   | *Escherichia coli* BL21; carAB::kan′                                   | expression and purification of wild-type and mutant carbamoyl phosphate synthetases |
| AM279   | AM187 + C1169→T in ygcB (changes Ala390 to Val in Cas3)                  |                                                                      |
| AM320   | AM187 + T1116→G in proA* (changes Phe372 to Leu)                         |                                                                      |
| AM327   | AM187 + 82 bp deletion upstream of pyrE (pos. 3808881-3808962)            |                                                                      |
| AM329   | AM187 + 82 bp deletion upstream of pyrE (pos. 3808881-3808962); C1169→T ygcB (changes Ala390 to Val in Cas3) |                                                                      |

a Genome positions refer to the sequence of strain AM187 (GenBank accession number CP037857), which was modified from the *E. coli* BW25113 sequence (GenBank accession number CP009273 (Grenier et al., 2014)) based on the mutations that had been introduced.
Over the course of the experiment, growth rate increased 2.5-3.5-fold for all eight populations (Figure 3). This improvement corresponds to a change in doubling time from ~3.3 hours to ~1 hour. As expected, the rate at which growth rate increased slowed over the course of adaptation. Occasional dips in growth rate occurred during the adaptation (Figure 3). These dips are artifacts arising from temporary aberrations in selective conditions due to turbidostat malfunctions that prevented introduction of fresh medium, causing the cultures to enter stationary phase. Occasionally cultures were saved as frozen stocks until the turbidostat was fixed (see Materials and Methods). Restarting cultures from frozen stocks may have caused a temporary drop in growth rate.

**Figure 3.** Growth rate increases ~3-fold during adaptation of ΔargC M2-proA* E. coli in M9 minimal medium containing 0.2% glucose, 0.4 mM proline and 20 µg/mL kanamycin.
The copy number and extent of amplification around proA* varied among replicate populations.

We monitored proA* copy number during the adaptation experiment using qPCR of population genomic DNA (Figure 4A). proA* was present in at least 6 copies by generation 300 in all eight populations. Six of the populations maintained 6-9 copies for the remainder of the adaptation. In contrast, proA* copy number in population 2 increased to as many as 20 copies. In population 3, proA* copy number dropped to three by generation 400.

We identified the boundaries of the amplified regions in all eight populations by sequencing population genomic DNA (Figure 4B, Supplementary File 1). The amplified region in population 2 was unusually small, spanning only 4.9 kb and resulting in co-amplification of only two other genes besides proBA*. Population 2 also appeared to have a second region of amplification of 18.5 kb. (Whether these two distinct amplification regions coexisted in the same clone or as two separate clades within the population could not be determined from population genome sequencing.) In contrast, the amplified regions in the other seven populations ranged from 41.1 to 163.8 kb, encompassing between 55 and 177 genes. We attribute the variation in proA* copy number to these differences in the size of the amplified region on the genome. The population with the smallest amplified region (4.9 kb, population 2) carries fewer multicopy genes and thus should incur a lower fitness cost, allowing proA* to reach a higher copy number (Adler et al., 2014; Kugelberg et al., 2006; Pettersson et al., 2009; Reams et al., 2010).
A mutation in \textit{proA*} led to deamplification in population 3

The decrease in \textit{proA*} copy number in population 3 was particularly noteworthy since it might have been an indication that a mutation had improved the neo-ArgC activity of ProA*, resulting in a decreased need for multiple copies. In fact, a mutation in \textit{proA*} that changes Phe372 to Leu (Figure 5A) was observed in population 3. E383A F372L ProA will be designated ProA** hereafter. Introduction of this mutation into the parental strain (which carried \textit{proA*}) increased growth rate by 69\% (Figure 5B), confirming that the mutation is adaptive and not a random hitchhiker. In contrast, no mutations in \textit{proA*} were identified in any of the other populations.

The neo-ArgC and native ProA activities of wild-type, ProA*, and ProA** were assayed (in the reverse direction) with NAGSA and GSA, respectively (Table 2). The $\frac{k_{\text{cat}}}{K_M}$NAGSA for
ProA** is 3.6-fold higher than that of ProA* and nearly 80-fold higher than that for ProA. In contrast, there is no difference between $k_{cat}/K_{M,GSA}$ for ProA* and ProA**.

To determine when the mutation that changes Phe372 to Leu in ProA* occurred, we performed sequencing of population genomic DNA at generations 270, 440, 630 and at the end of the adaptation (Figure 5C, 130x, 122x, 70x and 81x sequencing depth, respectively). proA** was present in 9% of the sequencing reads of the population by generation 270. By the time deamplification of proA* had occurred at generation 440, the frequency of proA** had risen to 21% of sequencing reads. By the end of the adaptation, proA** was fixed in the population, yet 3 copies remained in the genome, suggesting that ProA** does not have sufficient neo-ArgC activity to be present at a single copy in the genome.

The fact that a mutation that improved the neo-ArgC activity of ProA* occurred in only one population was surprising considering that ProA* is the weak-link enzyme limiting growth rate. Because the growth rates of all 8 populations improved substantially (Figure 3), mutations outside of the proBA* operon must also be contributing to fitness.

### Table 2. Kinetic parameters for GSA and NAGSA dehydrogenase activities of ProA, ProA*, and ProA**.

|                      | GSA activity (ProA) | NAGSA activity (neo-ArgC) |
|----------------------|---------------------|--------------------------|
|                      | $k_{cat}$ (s$^{-1}$) | $K_M$ (mM) | $k_{cat}/K_{M,GSA}$ (M$^{-1}$ s$^{-1}$) | $k_{cat}$ (s$^{-1}$) | $K_M$ (mM) | $k_{cat}/K_{M,NAGSA}$ (M$^{-1}$ s$^{-1}$) |
| WT                   | 16 ± 0.3            | 0.22 ± 0.01             | 72000 ± 2000             | 0.0083 ± 0.0009 | 0.30 ± 0.09 | 28 ± 9 |
| ProA* (E383A)        | 0.0076 ± 0.0008     | 0.20 ± 0.04             | 37 ± 8                  | 0.046 ± 0.002  | 0.076 ± 0.009 | 610 ± 74 |
| ProA** (E383A F372L) | 0.023 ± 0.005       | 0.42 ± 0.14             | 55 ± 22                 | 0.21 ± 0.01   | 0.095 ± 0.011 | 2200 ± 260 |

$^a$ Values reported were calculated from a nonlinear least squares regression of three replicates at each substrate concentration ± standard error.
**Figure 5.** proA* acquired a beneficial mutation in population 3. (A) Crystal structure of *Thermotoga maritima* ProA (PDB 1O20) (Page et al., 2003). Yellow, catalytic cysteine; green, equivalent of *E. coli* ProA Glu383; red, equivalent of *E. coli* ProA Phe372; magenta, NADPH-binding domain; blue, catalytic domain; gray, oligomerization domain. (B) Change in growth rate when the mutation changing Phe372 to Leu is introduced into the genome of the parental strain (AM187). Error bars represent one standard error from the mean, N = 4. (C) Growth rate (right axis, dotted lines) and proA* copy number (left axis, solid lines) for population 3. Vertical dotted lines indicate when population genomic DNA was sequenced. The frequency of the proA** allele at each time point is noted above the plot.
Mutations outside of proA* improved fitness

Population genome sequencing at the end of the experiment and at several intermediate timepoints from populations with longer adaptations (Figure S1) revealed that the final populations contained 13-178 mutations present at frequencies ≥5%, 3-5 mutations at frequencies ≥30%, and 1-4 mutations (not including amplification of proA*) that were fixed (100%). We found several mutations in the same genes in different populations, suggesting that these mutations confer a fitness advantage (Supplementary File 1).

The first mutation to appear in all populations was either an 82 bp deletion in the rph pseudogene directly upstream of pyrE or a C→T mutation in the intergenic region between rph and pyrE. PyrE is required for de novo synthesis of pyrimidine nucleotides. Both of these mutations have arisen in other E. coli evolution experiments, and have been shown to restore a known PyrE deficiency in the BW25113 E. coli strain (Blank et al., 2014; Bonekamp et al., 1984; Conrad et al., 2009; Jensen, 1993; Knöppel et al., 2018). Thus, these mutations are general adaptations to growth in minimal medium and do not pertain to the selective pressures caused by the weak-link enzyme ProA*.

A second early mutation in four of the populations occurred in ygcB. This mutation changes Ala390 to Val in Cas3, a nuclease/helicase in the Type I CRISPR/Cas system in E. coli (Howard et al., 2011). We introduced this mutation into the genome of the parent AM187 and compared the growth rates of the mutant and AM187 (Figure S2). Surprisingly, we saw no significant change in growth rate. Since this mutation appeared about the same time as the mutations upstream of pyrE, we wondered whether the ygcB mutation might only improve growth rate in the context of restored pyrE expression. Thus, we also tested the growth rate of a strain with the Cas3 mutation and the 82 bp deletion upstream of pyrE. Again, we saw no
significant change in relative growth rate (Figure S2). Thus, the $ygcB$ mutation is most likely a neutral hitchhiker. The most likely explanation for its prevalence is that it was present in a clade of the parental population that later rose to a high frequency when an additional beneficial mutation was acquired by one of its members.

Mutations upstream of $argB$ increase ArgB abundance

All eight final populations contained mutations in the intergenic region upstream of $argB$ and downstream of $kan^r$. These mutations were fixed in two populations, and present at frequencies of 9-82% in the other populations (Figure 6A).

We reintroduced six of the mutations upstream of $argB$ into the parental strain AM187. The mutations increased growth rate by 36-61% (Figure 6B). Levels of mRNAs for $argB$ and $argH$, which is immediately downstream of $argB$, were little affected by the mutations (Figure 6C). However, levels of ArgB protein measured by label-free high resolution Orbitrap mass spectrometry increased 2.6-8.2-fold (Figure 6D). In contrast, ArgH levels increased only modestly. These data suggest that the mutations upstream of $argB$ increase translational efficiency of $argB$ mRNA.

Increased translational efficiency can be due to increased accessibility of the region surrounding the Shine-Dalgarno site and the start codon, which might be achieved in three ways: 1) a decrease in the stability of the lowest-free-energy secondary structure in this region (Bentele et al., 2013; Espah Borujeni et al., 2014; Goodman et al., 2013); 2) a change in the ensemble of mRNA structures such that the Shine-Dalgarno sequence is more likely to be single-stranded in multiple accessible structures (Espah Borujeni et al., 2014; Salis et al., 2009); or 3) a decrease in the folding rate of the mRNA such that a ribosome can bind to the unfolded mRNA emerging
behind a preceding ribosome before the mRNA folds and obscures the Shine-Dalgarno sequence
(a process called ribosomal drafting, (Espah Borujeni & Salis, 2016)). The \textit{argB} mRNA, like
16\% of $\gamma$-proteobacterial mRNAs (Scharff et al., 2011), lacks a canonical Shine-Dalgarno
sequence, but the ribosome is expected to bind to a region encompassing the start codon and at
least the upstream 8-10 nucleotides. We calculated the minimum free energy secondary
structures of the 140-nucleotide RNA sequences encompassing the intergenic region affected by
the various mutations through 33 bp downstream of the \textit{argB} start codon (Figure S3). Direct
comparisons of the predicted free energies of these structures are not particularly informative,
especially for mutants with large deletions or duplications. However, for 5 of the 8 of the mutant
structures, the probability that the 5'-UTR from about 20 nt upstream of the start codon is
sequestered in a stem-loop in the lowest free-energy structure is decreased relative to the parent
(Figure S3B); the increased accessibility of this region should increase translation efficiency.
However, for three cases (-94 A$\rightarrow$G, -22 C$\rightarrow$A, and -18 C$\rightarrow$A), this region is equally or more
likely to be sequestered in a stem-loop. The thermodynamic stability of this region is clearly not
the only factor responsible for the effects of the mutations upstream of \textit{argB}.

We also considered the possibility that mutations upstream of \textit{argB} might increase
expression by increasing ribosome drafting. Figure S4 shows the predicted folding times of RNA
sequences encompassing 30 nucleotides downstream to 30 nucleotides upstream of the \textit{argB} start
codon (totaling 63 nucleotides) for each mutant except the -94 A$\rightarrow$G mutant. Folding is
predicted to be significantly slower for three of the seven mutant RNAs, which should increase
the efficiency of translation. For four of the mutants, however, folding rate is similar to or even
faster than that of the parental mRNA. For three of these mutants (the 58 bp deletion, -22 C$\rightarrow$T,
and the 38 bp duplication), the secondary structure prediction shown in Figure S3 suggests that
the translation initiation region is less likely to be sequestered in a hairpin. Thus, the effects of 6
of the 8 mutations upstream of argB can be explained by a decrease in secondary structure
stability around the ribosome binding site or a decrease in the folding rate of the mRNA in this
region, or both. The effects of the -94 A→G and -22 C→A mutations, however, cannot be
explained by either of these mechanisms.
Figure 6. Several adaptive mutations occurred upstream of \textit{argB}. (A) Locations of adaptive mutations (red) found upstream of \textit{argB} (yellow) and \textit{argH} (purple). \textit{argC} was replaced with \textit{kan} in the parental strain AM187, giving this operon two promoters, one native to the operon (\textit{P}_{\text{argCBH}}), and the other introduced with the \textit{kan} gene (\textit{P}_{\text{kan}}). The table shows the percentages of each adapted population that contained a given \textit{argB} mutation at the final time point. Six of the \textit{argB} mutations were introduced into the genome of the ancestral strain and changes in growth rate (B), gene expression (C), and protein abundance (D) were determined.
Mutations in *carB* either increase activity or impact allosteric regulation

We found eight different mutations in *carB* in six of the evolved populations: four missense mutations, three deletions (≥12 bp), and one 21 bp duplication (Figure 7A). CarB, the large subunit of carbamoyl phosphate synthetase (CPS), forms a complex with CarA to catalyze production of carbamoyl phosphate from glutamine, bicarbonate, and two molecules of ATP (Eq. [1]) (Rubino et al., 1986, 1987).

\[\text{Gln} + \text{HCO}_3^- + 2 \text{ATP} \rightarrow \text{carbamoyl phosphate} + \text{Glu} + 2 \text{ADP} + \text{P}_i\]

Synthesis of carbamoyl phosphate involves four reactions that take place in three separate active sites connected by a molecular tunnel of ~100 Å in length (Thoden et al., 2002). CarA catalyzes hydrolysis of glutamine to glutamate and ammonia (Eq. [2]), while CarB phosphorylates bicarbonate to form carboxyphosphate in its first active site (Eq. [3]). Ammonia from the CarA active site is channeled to CarB, where it reacts with carboxyphosphate to form carbamate (Eq. [4]). Carbamate migrates to a second active site within CarB where it reacts with ATP to form carbamoyl phosphate and ADP (Eq. [5]).

\[\text{Gln} + \text{H}_2\text{O} \rightarrow \text{NH}_3 + \text{Glu}\]

\[\text{HCO}_3^- + \text{ATP} \rightarrow \text{carboxyphosphate} + \text{ADP}\]

\[\text{NH}_3 + \text{carboxyphosphate} \rightarrow \text{carbamate} + \text{P}_i\]

\[\text{carbamate} + \text{ATP} \rightarrow \text{carbamoyl phosphate} + \text{ADP}\]

Carbamoyl phosphate feeds into both the pyrimidine and arginine synthesis pathways and its production is regulated in response to intermediates or products of both pathways (Figure 7B), as well as by IMP (a purine) (Pierrat & Raushel, 2002). It is inhibited by UMP (a pyrimidine) and moderately activated by IMP. UMP and IMP compete to bind the same region of CarB (Eroglu & Powers-Lee, 2002). The allosteric effects of UMP and IMP are dominated, however, by activation by ornithine. Ornithine, an intermediate in arginine synthesis that reacts with
carbamoyl phosphate, binds to and activates CarB even when UMP is bound (Figure 7C) (Braxton et al., 1999; Eroglu & Powers-Lee, 2002).

Seven of the eight mutations found in carB affect residues in the allosteric domain of CarB. The lone mutation outside of the allosteric domain changes Gly369, which is immediately adjacent to the allosteric region in the 3D structure, to Val (Figure 7C).

The kinetic parameters for CPS activity [determined as the glutamine- and bicarbonate-dependent ATPase activity (Eq. [1])] of all eight CPS mutants are shown in Table 3. All mutations decreased $k_{cat}/K_{m,ATP}$ by 34-63%, with the exception of the mutation that changes Lys966 to Glu, which nearly doubles $k_{cat}/K_{m,ATP}$. None of the mutations affected the enzyme’s ability to couple ATP hydrolysis with carbamoyl phosphate production (Figure S5).

Table 3. Kinetic parameters$^a$ for the glutamine-dependent ATPase reaction of wild-type and variant carbamoyl phosphate synthetases.

| Enzyme          | $K_{M,ATP}$ (mM) | $k_{cat}$ (s$^{-1}$) | $k_{cat}/K_{M,ATP}$ (M$^{-1}$ s$^{-1}$) | UMP $K_d$ (µM) | UMP $a$ | ornithine $K_d$ (µM) | ornithine $a$ |
|-----------------|------------------|---------------------|----------------------------------------|----------------|---------|----------------------|--------------|
| WT              | 1.05 ± 0.08      | 13.5 ± 0.3          | 12.9 (± 1.0) × 10$^3$                  | 0.81 ± 0.04    | 0.27 ± 0.04          | 130 ± 7     | 3.28 ± 0.03          |
| G369V           | 3.31 ± 0.25      | 21.5 ± 0.7          | 6.51 (± 0.54) × 10$^3$                 | 1.53 ± 0.29    | 0.48 ± 0.02          | 372 ± 20    | 11.8 ± 0.13          |
| L960P           | 1.12 ± 0.05      | 9.10 ± 0.15         | 8.12 (± 0.41) × 10$^3$                 | na             | na                  | na         | na                   |
| L964Q           | 1.25 ± 0.08      | 8.04 ± 0.17         | 6.41 (± 0.42) × 10$^3$                 | na             | na                  | na         | na                   |
| K966E           | 0.97 ± 0.06      | 20.6 ± 0.4          | 21.2 (± 1.4) × 10$^3$                  | 0.61 ± 0.04    | 0.21 ± 0.01          | 181 ± 34    | 3.23 ± 0.08          |
| ∆12 bp (at nt 2906)$^a$ | 1.09 ± 0.06 | 4.80 ± 0.09        | 4.39 (± 0.25) × 10$^3$                 | na             | na                  | na         | na                   |
| ∆132 bp (at nt 2986) | 1.16 ± 0.06   | 6.47 ± 0.11        | 5.57 (± 0.31) × 10$^3$                 | na             | na                  | na         | na                   |
| ∆12 bp (at nt 3108) | 1.30 ± 0.10   | 5.86 ± 0.16        | 4.51 (± 0.37) × 10$^3$                 | na             | na                  | na         | na                   |
| 21 bp dup.      | 1.40 ± 0.12      | 9.70 ± 0.30         | 6.94 (± 0.64) × 10$^3$                 | 597 ± 133      | 0.51 ± 0.03          | na         | na                   |

$^a$ Values reported were calculated from a nonlinear least squares regression of data for three replicates at each substrate concentration ± standard error. Values for $K_d$ and $a$ for UMP and ornithine were determined by fitting the data to the following equation: $v/v_0 = (aL + K_d)/(L + K_d)$, where $L$ is the ligand concentration, $v$ is the initial reaction rate.
rate, $v_0$ is the initial reaction rate in the absence of ligand, $a$ is $v/v_0$ at infinite $L$, and $K_d$ is the apparent dissociation constant. No $K_d$ or $a$ values are given (indicated by $na$) when the inhibition by the allosteric ligand was too weak to measure.

$b$ Nucleotide (nt) numbers refer to the position of deletions or duplications in $carB$.

We also measured the effect of mutations on UMP inhibition and ornithine activation of CPS (Table 3, Figure 7D-E). Regulation of the K966E variant, the enzyme for which $k_{cat}/K_{m,ATP}$ was nearly doubled, was minimally affected. Five of the variants showed a complete loss of allosteric regulation by both UMP and ornithine. The variant with the 21 bp duplication retained modest inhibition by UMP (~50% reduction of turnover at high UMP concentrations), but only at very high concentrations of UMP; the apparent $K_{d,UMP}$ was increased by 740-fold. Similarly, G369V CPS retained partial inhibition by UMP (~50% reduction of turnover at high UMP concentrations). The apparent $K_{d,UMP}$ of the G369V enzyme was only doubled, but this variant showed a 3.5-fold increase in activation at high ornithine concentrations.
**Figure A: Adapted Population**

| CarB Mutation | $k_{cat}/K_{M}$ (L-Orn activation) | L-Orn activation |
|---------------|-----------------------------------|------------------|
| G369V        | ↓                                 | ↑                |
| L960P        | ↓                                 | X                |
| L964Q        | ↓                                 | X                |
| K966E        | ↑                                 | n.c.             |
| Δ12 bp (2066-2077 nt) | ↓ | X | X |
| Δ132 bp (2986-3117 nt) | ↓ | X | X |
| 21 bp dup. (2150 nt) | ↓ | ↓ | X |

Arabidopsis thaliana indicates where in the 3122 nt carB gene deletions or duplications occurred.

| and $↓$ indicate that a kinetic parameter was either increased or decreased, respectively, in the variant enzyme relative to the wild-type enzyme.

n.c. indicates no change in the kinetic parameter relative to the wild-type enzyme.

**Figure B: CarA and CarB Pathway**

- Gln + HCO$_3^-$ → CarA → CarB → ornithine → carbamoyl phosphate → arginine → UMP

**Figure C: Molecular Model**

- G369V
- L960P
- L964Q
- K966E

**Figure D: Graph**

- $V/V_0$ vs. log[UMP] (mM)

**Figure E: Graph**

- $V/V_0$ vs. log[ornithine] (mM)
Recruitment of promiscuous enzymes to serve new functions followed by mutations that improve the promiscuous activity has been a dominant force in the diversification of metabolic networks (Copley, 2017; Glasner et al., 2006; Khersonsky & Tawfik, 2010; O’Brien & Herschlag, 1999; Rauwerdink et al., 2016). New enzymes may be important for fitness or even survival when an organism is exposed to a novel toxin or source of carbon or energy, or when synthesis of a natural product can enable manipulation of competing organisms. This process also contributes to non-orthologous gene replacement, which can occur when a gene is lost during a time in which it is not required, but its function later becomes important again and is replaced by recruitment of a non-orthologous promiscuous enzyme (Albalat & Cañestro, 2016; Ferla et al., 2017; Juárez-Vázquez et al., 2017; Olson, 1999).

We have modeled a situation in which a new enzyme is required by deleting argC, which is essential for synthesis of arginine in E. coli. Previous work showed that the most readily available source of neo-ArgC activity that enables ∆argC E. coli to grow on glucose as a sole carbon source is a promiscuous activity of ProA. However, a point mutation that changes Glu383 to Ala is required to elevate the promiscuous activity to a physiologically useful level. This
mutation substantially damages the native function of the enzyme, creating an inefficient bifunctional enzyme whose poor catalytic abilities limit growth rate on glucose. It is important to note that the decrease in the efficiency of the native reaction may be a critical factor in the recruitment of ProA to serve as a neo-ArgC because it will diminish inhibition of the newly important reaction by the native substrate (Khanal et al., 2015; McLoughlin & Copley, 2008).

When we previously carried out short-term adaptation of ΔargC proA* E. coli in glucose, we observed four mutations that increased fitness by increasing expression of proA*: a G→A substitution in the -35 promoter region (M1) and a C→T substitution in the -10 promoter region (M2) upstream of the proBA* operon, a synonymous mutation in proB that strengthens a cryptic promoter upstream of proA* (M3), and amplification of a region of the genome that includes the proBA* operon (Kershner et al., 2016). The point mutations occurred independently, never together. Amplification occurred even after acquisition of M1, M2, or M3. The amplification of proA* suggests that it has the potential to diverge to restore efficient ProA function and evolve an efficient neo-ArgC.

We chose to carry out longer-term evolution of a ΔargC proA* strain in glucose in the presence of proline to specifically address the evolution of an efficient neo-ArgC. We introduced the M2 promoter mutation in the parent strain to ensure that all populations had the same promoter mutation during adaptation. After 470-1000 generations of adaptation, growth rate had increased ~3-fold in all eight replicate cultures. We have focused on five types of genetic changes that clearly increase fitness: (1) mutations upstream of pyrE; (2) amplification of a variable region of the genome surrounding the proBA* operon; (3) a mutation in proA* that changes Phe372 to Leu; (4) mutations upstream of argB; and (5) mutations in carB. (Each of the final populations contains additional mutations that may also contribute to fitness, but these
mutations were typically found in low abundance and/or in only one population.) The pyrE mutations have previously been shown to be a general adaptation of *E. coli* BW25113 to growth in minimal medium (Blank et al., 2014; Conrad et al., 2009; Jensen, 1993; Knöppel et al., 2018). The other four types of mutations are specific adaptations to the bottleneck in arginine synthesis caused by substitution of the weak-link enzyme ProA* for ArgC. Interestingly, only two of these – gene amplification and the mutation in *proA* – directly involved the weak-link enzyme ProA*.

Surprisingly, we saw evolution of *proA* towards a more efficient neo-ArgC in only one population (Table 2). In this population, *proA* copy number dropped from ~7 to ~3 within 100 generations (Figure 5C). This pattern is consistent with the IAD model; copy number is expected to decrease as mutations increase the efficiency of the weak-link activity. However, the fact that copy number did not return to one implies that the neo-ArgC activity of ProA** is not sufficient to justify a single copy of the gene.

Because ~3 copies of *proA** remained in the population (Figure 5C) and the progenitor *proA* was not detectable, all copies in the amplified array have clearly acquired the mutation that changes Phe372 to Leu – i.e. the more beneficial *proA** allele has “swept” the amplified array. This observation has important implications for the IAD model. In the original conception of the IAD model, it was proposed that amplification of a gene increases the opportunity for different beneficial mutations to occur in different alleles, and then for recombination to shuffle these mutations (Bergthorsson et al., 2007; Francino, 2005). Both phenomena would increase the rate at which sequence space can be searched and thereby increase the rate at which a new enzyme evolves. In order for this to occur, however, it would be necessary for individual alleles to acquire different beneficial mutations before recombination occurred. This scenario is inconsistent with the relative frequencies of point mutations and recombination between large
homologous regions in an amplified array (Anderson & Roth, 1981; Reams et al., 2010). Point
mutations occur at a frequency between $10^{-9}$ and $10^{-10}$ per nucleotide per cell division depending
on the genomic location (Jee et al., 2016), and thus between $10^{-6}$ and $10^{-7}$ per gene per cell
division for a gene the size of proA. If 10 copies of an evolving gene were present, then the
frequency of mutation in a single allele would be between $10^{-5}$ and $10^{-6}$ per cell division.
Homologous recombination after an initial duplication event is orders of magnitude more
frequent, occurring in ~1 of every 100 cell divisions (Reams et al., 2010). Thus, homologous
recombination between replicating chromosomes in a cell could result in a selective allelic sweep
long before a second beneficial mutation occurs in a different allele in the amplified array. Figure
8 depicts how this could occur after acquisition of an advantageous mutation in one allele in an
amplified array. This is indeed the result that we observed; heterozygosity among proA* alleles
was lost within 500 generations (Figure 5C). More recent papers depict selective amplification of
beneficial alleles before acquisition of additional mutations (Andersson et al., 2015; Näsvall et
al., 2012); our results support this revision of the original IAD model. It is possible that alleles
encoding enzymes that are diverging toward two specialists might recombine to explore
combinations of mutations. However, such recombination might not accelerate evolution, as
mutations that lead toward one specialist enzyme would likely be incompatible with those that
lead toward the other specialist enzyme.
While growth rate improved substantially in all populations, a beneficial mutation in *proA* arose in only one, suggesting that either mutations that improve the neo-ArgC activity are uncommon, or that their fitness effects are smaller than those caused by mutations elsewhere in the genome that also improve arginine synthesis. We identified two primary mechanisms that apparently improve arginine synthesis without affecting the efficiency of the weak-link enzyme ProA* itself.

ArgB (N-acetylglutamate kinase) catalyzes the second step in arginine synthesis by phosphorylating N-acetylglutamate to form NAGP, the native substrate for ArgC and a secondary substrate for ProA* (Figure 2). We identified eight mutations upstream of *argB*; the six we tested improved growth rate by 36-61% and increased the abundance of ArgB by 2.6-8.2-fold. Notably, ArgB levels were increased even though the levels of *argB* mRNA were unchanged (Figure 6). The increase in protein levels without a concomitant increase in mRNA

**Figure 8.** Homologous recombination of an amplified *proA* array with one *proA* allele can rapidly lead to a daughter cell with only *proA* alleles. Each arrow represents one cell duplication event and shows the genotypes of the two resulting daughter cells. The less-fit daughter cell from each recombination event is grayed out.
levels suggests that these mutations impact the efficiency of translation. Secondary structure
around the translation initiation site plays a key role because this region must be unfolded in
order to bind to the small subunit of the ribosome (Hall et al., 1982; Scharff et al., 2011). Indeed,
a study of the predicted secondary structures of 5000 genes from bacteria, mitochondria and
plastids, many of which lack canonical Shine-Dalgarno sequences (as does argB), showed that
secondary structure around the start codon is markedly less stable than up- or down-stream
regions (Bentele et al., 2013; Espah Borujeni et al., 2014; Goodman et al., 2013; Scharff et al.,
2011). Our computational studies of the effect of mutations on the predicted lowest free energy
secondary structures of the region surrounding the start codon of argB suggest that the
thermodynamic stability of this region plays a role in the beneficial effects of most of the observe
mutations (Figure S3). In addition, some of the mutations slow the predicted rate of mRNA
folding around the start codon, which would increase the probability of ribosomal drafting
(Figure S4). Both effects would lead to an increase in ArgB abundance. An increase in the level
of ArgB should increase the concentration of the NAGP substrate for the weak-link ProA*,
thereby pushing material through this bottleneck in the arginine synthesis pathway.

The adaptive mutations in carB increase catalytic turnover, decrease inhibition by UMP,
or increase activation by ornithine of CPS. All of these effects should increase the level of CPS
activity in the cell and consequently the level of carbamoyl phosphate. Why would this be
advantageous? Carbamoyl phosphate reacts with ornithine, which will be in short supply due to
the upstream ProA* bottleneck, to generate citrulline (Legrain & Stalon, 1976) (Figure 9). If
ornithine transcarbamoylase is not saturated with respect to carbamoyl phosphate, then
increasing carbamoyl phosphate levels should increase citrulline production and thereby increase
flux into the lower part of the arginine synthesis pathway. Although we do not know whether
ornithine transcarbamoylase is saturated with respect to carbamoyl phosphate in vivo (the $K_M$ for carbamoyl phosphate is 360 µM (Baur et al., 1990), but the intracellular concentration of carbamoyl phosphate is unknown), the occurrence of so many mutations that increase carbamoyl phosphate synthase activity supports the notion that they lead to an increase in carbamoyl phosphate that potentiates flux through the arginine synthesis pathway.

The majority of adaptive mutations we observed in carB cause loss of the exquisite allosteric regulation that controls flux through this important step in pyrimidine and arginine synthesis. This tight regulation likely evolved due to the energetically costly reaction catalyzed by CPS, which consumes two ATP molecules. While a constitutively active CPS is beneficial in the short term to improve arginine synthesis, it will likely be detrimental once arginine production no longer limits growth. We term mutations such as those observed in carB “expedient” mutations because they provide a quick fix when cells are under strong selective pressure, but at a cost to a previously well-evolved function. The damage caused by expedient mutations may be repairable later by reversion, compensatory mutations or horizontal gene transfer. Interestingly, the latter two repair processes may contribute to sequence divergence between organisms that has typically been attributed to neutral drift, but rather may be due to scars left from previous selective processes.

A particularly striking conclusion from this work is that most of the mutations that improved fitness under these strong selective conditions did not impact the gene encoding the weak-link enzyme, but rather adjusted fluxes in the metabolic network to compensate for the bottleneck in metabolism by other mechanisms. Not surprisingly, the process of evolution of a new enzyme by gene duplication and divergence does not take place in isolation, but is inextricably intertwined with mutations in the rest of the genome. The ultimate winner in a
Figure 9. Adaptive mutations are predicted to increase flux through the arginine synthesis pathway. The pathway bottleneck and weak-link enzyme ProA* is shown in red. Steps in the arginine synthesis pathway that are affected by adaptive mutations are highlighted in bold type.
Supplementary Figures

**Figure S1.** Growth rate (right axis, dotted lines) and *proA* copy number (left axis, solid lines) for each adapted population. Vertical dotted lines indicate when population genomic DNA was sequenced.
Figure S2. Effects of the A390V Cas3 and rph-pyrE mutations on growth rate of the parental strain AM187. (All p values are >0.1 by a one-tailed Student’s t-test).
Figure S3. *argB* mRNA secondary structure comparison. (A) The entire intergenic region between *kan* and *argB* plus 33 nucleotides of the *argB* coding region was used for RNA structure prediction. The start codon for *argB* is boxed in each structure. Arrows in the AM187 structure indicate where an adaptive point mutation occurred. Arrows in mutant structures indicate which nucleotide was changed. In the 38 bp duplication mutant structure, the brackets contain the 38 inserted nucleotides. (B) Same as in (A), but showing only the region surrounding the *argB* start codon.
Figure S4. Folding time of a region encompassing 30 nucleotides upstream of the argB start codon plus 33 nucleotides of the argB coding region based on Kinfold simulations (Wolfinger et al., 2004). Error bars represent one standard error from the mean, N = 500.

Figure S5. Ratio of ADP to carbamoyl phosphate (CP) produced by wild-type and mutant carbamoyl phosphate synthetases.
Materials and Methods

Materials

Common chemicals were purchased from Sigma-Aldrich (St. Louis, MO) and Fisher Scientific (Fair Lawn, NJ).

NAGSA was synthesized enzymatically from N-acetylcysteine using N-acetylornithine aminotransferase (ArgD) in a 25 mL reaction as described previously by Khanal et al. 2015 and stored at -70 °C. NAGSA concentrations were determined using the o-aminobenzaldehyde assay as described previously (Albrecht et al., 1962; Mezl & Knox, 1976).

GSA was synthesized enzymatically from L-ornithine using N-acetylcysteine aminotransferase (ArgD) as described previously by Khanal et al. 2015 and stored at -70 °C. GSA concentrations were determined using the o-aminobenzaldehyde assay as described previously (Albrecht et al., 1962; Mezl & Knox, 1976).

Plasmids and primers used in this study are listed in Tables S1 and S2, respectively.

Strains and culture conditions

Strains used in this study are listed in Table 1. *E. coli* cultures were routinely grown in LB medium at 37 °C with 20 µg/mL kanamycin, 100 µg/mL ampicillin, 20 µg/mL chloramphenicol, or 10 µg/mL tetracycline, as required. Adaptation of strain AM187 was performed at 37 °C in M9 minimal medium containing 0.2% glucose, 0.4 mM proline, and 20 µg/mL kanamycin (Adaptation Medium).
Strain construction

The parental strain for the adaptation experiment (AM187) was constructed from the Keio collection argC::kan<sup>+</sup> *E. coli* BW25113 strain (Baba et al., 2006). The fimAICDFGH and *csgBAC* operons were deleted (to slow biofilm formation), and the M2 *proBA* promoter mutation (Kershner et al., 2016) and the point mutation in *proA* that changes Glu383 to Ala (McLoughlin & Copley, 2008) were inserted into the genome using the scarless genome editing technique described in Kim et al. 2014. We initially hoped to measure *proA* copy number during adaptation using fluorescence, although ultimately qPCR proved to be a better approach. Thus, we inserted *yfp* downstream of *proA* under control of the P3 promoter (Mutalik et al. 2013) and with a synthetically designed ribosome binding site (Espah Borujeni et al., 2014; Salis et al., 2009). A double transcription terminator (BioBrick Part: BBa_B0015) was inserted immediately downstream of *proBA* to prevent read-through transcription of *yfp*. We also inserted a NotI cut site immediately downstream of *proA* to enable cloning of individual *proA* alleles after amplification if necessary. A Fis binding site located 32 bp downstream of *proA* was preserved because it might impact *proA* transcription. The NotI-2xTerm-*yfp* cassette was inserted downstream of *proA* using the scarless genome editing technique described in Kim et al. 2014. The genome of the resulting strain AM187 was sequenced to confirm that there were no unintended mutations and deposited to NCBI GenBank under accession number CP037857.

Most mutations observed during the adaptation experiment were introduced into the parental AM187 strain using the scarless genome editing protocol described in Kim et al. 2014. This protocol is preferable to Cas9 genome editing for introduction of point mutations and small indels because it does not require introduction of synonymous PAM mutations that have the potential to affect RNA structure. The 58 bp deletion upstream of *argB* and the 82 bp deletion in
were introduced using Cas9-induced DNA cleavage and λ Red recombinase-mediated homology-directed repair with a linear DNA fragment. Sequences of the protospacers and mutations cassettes used for Cas9 genome editing procedures are listed in Tables S3 and S4. Details of the genome editing procedures are provided in the supplementary material.

Laboratory adaptation

Adaptation of strain AM187 in Adaptation Medium was carried out in eight replicate tubes in a custom turbidostat constructed as described by Takahashi et al. 2015. To start the adaptation, strain AM187 was grown to exponential phase (OD$_{600}$ = 0.7) in LB/kanamycin at 37 °C. Cells were centrifuged at 4,000 x g for 10 min at room temperature and resuspended in an equal volume of PBS. The suspended cells were washed twice more with PBS and resuspended in PBS. This suspension was used to inoculate all eight turbidostat chambers to give an initial OD$_{600}$ of 0.01 in 14 mL of Adaptation Medium. The turbidostat was set to maintain an OD$_{650}$ of 0.4 by diluting individual cultures with an appropriate amount of fresh medium every 60 seconds.

A 3 mL portion of each population was collected every 2-3 days; 800 µl was used to make a 10% glycerol stock, which was then stored at -70 °C. The remaining sample was pelleted for purification of genomic DNA using the Invitrogen PureLink Genomic DNA Mini Kit according to the manufacturer’s protocol. The turbidostat occasionally malfunctioned, requiring the adaptation to be paused. When this occurred, the populations were subjected to centrifugation at 4,000 x g for 10 min at room temperature and the pelleted cells were resuspended in 1.6 mL of Adaptation Medium. Half of the resuspension was used to make a -70 °C 10% glycerol stock, and the other half to purify
genomic DNA. When the turbidostat was fixed, the frozen stock was thawed and the cells were collected by centrifugation at 16,000 x g for 1 min at room temperature. The pelleted cells were resuspended in 1 mL of PBS, washed, and resuspended in 500 µL of Adaptation Medium. The entire resuspension was used to inoculate the appropriate chamber of the turbidostat. Sometimes the adaptation had to be restarted from a frozen stock of a normal sample (as opposed to the entire population as just described), resulting in a more significant population bottleneck. In this case, the entire frozen stock was thawed and only 700 µL washed as described above to be used for the inoculation. The remaining 300 µL of the glycerol stock were re-stored at -70 °C in case the frozen stock was needed for downstream analysis.

520

**Calculation of growth rate and generations during adaptation**

The turbidostat takes an OD$_{650}$ reading every ~3 sec and dilutes the cultures every 60 seconds. Thus, readings between dilutions can be used to calculate an average growth rate each day based on the following equation:

$$\bar{\mu} = \frac{\sum_{i=1}^{n} \ln\left(\frac{N_{t_{2,i}}}{N_{t_{1,i}}}\right)}{t_{0,i} - t_{1,i}}$$

where $\bar{\mu}$ is the average growth rate in hr$^{-1}$, $n$ is the number of independent growth rate calculations within a given 24 hr period, $N_{t_2}$ is the OD$_{650}$ reading right before the dilution, $N_{t_1}$ is the OD$_{650}$ reading right after the dilution, and $t_1$ and $t_0$ are the times at which the OD$_{650}$ was measured. The number of generations per day ($g$) was then calculated from $\bar{\mu}$ using Eq. [7].

$$g = \frac{24 \cdot h}{\ln(2)/\bar{\mu}}$$
Measurement of *proA* copy number

The copy number of *proA* was determined by qPCR of purified population genomic DNA. *gyrB* and *icd*, which remained at a single copy in the genome throughout the adaptation experiment, were used as internal reference genes. Details of the experimental procedure and data analysis are provided in the supplementary material.

Whole-genome sequencing

Libraries were prepared from purified population genomic DNA using a modified Illumina Nextera protocol and multiplexed onto a single run on an Illumina NextSeq500 to produce 151-bp paired-end reads (Baym et al., 2015), giving a 60-130-fold coverage of the strain AM187 genome. Reads were trimmed using BBtools v35.82 (DOE Joint Genome Institute) and mapped using *breseq* v0.32.1 using the polymorphism (mixed population) option (Deatherage & Barrick, 2014).

Growth rate measurements

Growth rates for each strain were calculated from growth curves measured in quadruplicate. Overnight cultures were grown in LB/kanamycin at 37 °C from glycerol stocks. Ten µL of each overnight culture was used to inoculate 4 mL of LB/kanamycin and the cultures were allowed to grow to mid-exponential phase (OD$_{600}$ 0.3-0.6) at 37 °C with shaking. The cultures were subjected to centrifugation at 4,000 x g for 10 min at room temperature and the pellets resuspended in an equal volume of PBS. The pellets were washed once more in PBS. The cells were diluted to an OD of 0.001 in Adaptation Medium and a 100 µL aliquot was loaded into each well of a 96-well plate. The plates were incubated in a Varioskan (Thermo Scientific)
plate reader at 37 °C with shaking every 5 minutes for 1 minute. The absorbance at 600 nm was measured every 20 minutes for up to 500 hours. The baseline absorbance for each well (the average over several smoothed data points before growth) was subtracted from each point of the growth curve. Growth parameters (maximum specific growth, $\mu_{\text{max}}$; lag time, $\lambda$; maximum growth, $A_{\text{max}}$) were estimated by non-linear regression using the modified Gompertz equation (Zwietering et al., 1990). Non-linear least-squares regression was performed in Excel using the Solver feature.

**Measurement of argB and argH gene expression by RT-qPCR**

Overnight cultures were grown in LB/kanamycin at 37 °C from glycerol stocks. Ten µL of each overnight culture was used to inoculate 4 mL of LB/kanamycin and the cultures were grown to mid-exponential phase ($OD_{600}$ 0.3-0.6) at 37 °C with shaking. Exponential phase cultures were centrifuged at 4,000 x $g$ for 10 min and pellets resuspended in equal volume PBS. Pellets were washed once more in PBS. The cells were diluted to an OD of 0.001 in 4 mL of Adaptation Medium and grown to an $OD_{600}$ of 0.2-0.3. Four 2-mL aliquots of culture were thoroughly mixed with 4 mL of RNAprotect Bacteria Reagent (Qiagen) and incubated at room temperature for 5 min before centrifugation at 4,000 x $g$ for 12 min at room temperature. Pellets were frozen in liquid N$_2$ and stored at -70 °C. Procedures for purification of RNA, reverse transcription, qPCR, and data analysis are described in the supplementary material.

**Measurement of ArgB and ArgH levels by mass spectrometry**

Freezer stocks were used to streak each strain on LB/kanamycin. Four parallel 2-mL aliquots of LB/kanamycin were inoculated with individual colonies and the cultures were grown
to mid-exponential phase at 37 °C with shaking. One mL of each culture was subjected to centrifugation at 16,000 x g for 1 min at room temperature. The cell pellets were resuspended in 1 mL PBS and washed twice more in PBS before resuspension and dilution to an OD of 0.001 in 5 mL of Adaptation Medium. Cultures were grown to an OD<sub>600</sub> of 0.2-0.3 at 37 °C with shaking and then chilled on ice for 10 min before pelleting by centrifugation at 4,000 x g at 4 °C. Cell pellets were frozen in liquid N<sub>2</sub> and stored at -70 °C.

Frozen cell pellets were thawed and lysed in 60 µL 50 mM Tris-HCl, pH 8.5, containing 4% (w/v) SDS, 10 mM TCEP and 40 mM chloroacetamide in a Bioruptor Pico sonication device (Diagenode) using 10 cycles of 30 seconds on, 30 seconds off, followed by boiling for 10 min, and then another 10 cycles in the Bioruptor. The lysates were subjected to centrifugation at 15,000 x g for 10 minutes at 20 °C and protein concentrations in the supernatant were determined by tryptophan fluorescence (Wiśniewski & Gaugaz, 2015). Ten µL of each sample (3-6 µg protein/µL) was digested using the SP3 method (C. S. Hughes et al., 2014). Carboxylate-functionalized speedbeads (GE Life Sciences) were added to the lysates. Addition of acetonitrile to 80% (v/v) caused the proteins to bind to the beads. The beads were washed twice with 70% (v/v) ethanol and once with 100% acetonitrile. Protein was digested and eluted from the beads with 15 µL of 50 mM Tris buffer, pH 8.5 with 1 µg endoproteinase Lys-C (Wako) for 2 hours with shaking at 600 rpm at 37 °C in a thermomixer (Eppendorf). One µg of trypsin (Pierce) was then added to the solution and incubated at 37 °C overnight with shaking at 600 rpm. Beads were collected by centrifugation and then placed on a magnet to more reliably remove the elution buffer containing the digested peptides. The peptides were then desalted using an Oasis HLB cartridge (Waters) according to the manufacturer’s instructions and dried in a speedvac.
Samples were suspended in 12 µL of 3% (v/v) acetonitrile/0.1% (v/v) trifluoroacetic acid and 500 ng of peptides were directly injected onto a C18 1.7 µm, 130 Å, 75 µm X 250 mm M-class column (Waters), using a Waters M-class UPLC. Peptides were eluted at 300 nL/minute using a gradient from 3% to 20% acetonitrile over 100 minutes into an Orbitrap Fusion mass spectrometer (Thermo Scientific). Precursor mass spectra (MS1) were acquired at a resolution of 120,000 from 380-1500 m/z with an AGC target of 2.0 x 10^5 and a maximum injection time of 50 ms. Dynamic exclusion was set for 20 seconds with a mass tolerance of +/- 10 ppm.

Precursor peptide ion isolation width for MS2 fragment scans was 1.6 Da using the quadrupole, and the most intense ions were sequenced using Top Speed with a 3-second cycle time. All MS2 sequencing was performed using higher energy collision dissociation (HCD) at 35% collision energy and scanning in the linear ion trap. An AGC target of 1.0 x 10^4 and 35-second maximum injection time was used. Rawfiles were searched against the Uniprot *Escherichia coli* database using Maxquant version 1.6.1.0 with cysteine carbamidomethylation as a fixed modification.

Methionine oxidation and protein N-terminal acetylation were searched as variable modifications. All peptides and proteins were thresholded at a 1% false discovery rate (FDR).

**Enzyme overexpression plasmids**

*proA* and *proA* were amplified from the genomes of *E. coli* BW25113 and AM187, respectively, and inserted into a linearized pET-46 vector backbone by Gibson assembly (Gibson et al., 2009). A sequence encoding a 6xHis-tag followed by a 2xVal-linker was incorporated at the N-terminus of each protein. The *proA** expression plasmid was constructed using the Q5 Site-Directed Mutagenesis Kit (NEB).
argC was cloned into a pTrcHisB vector backbone. The final plasmid included an N-terminal 6xHis-tag followed by a Gly-Met-Ala-Ser linker and with Met1 of ArgC removed.

The argD and argI expression plasmids from the ASKA collection (Kitagawa et al., 2005) were used for expression of N-acetylornithine aminotransferase and ornithine transcarbamoylase, respectively. In each case, the expression plasmid included a sequence encoding an N-terminal 6xHis-tag followed by a Thr-Asp-Ala-Leu-Arg-Ala linker.

Wild-type carAB was amplified from the genome of AM187 and inserted into a linearized pCA24N vector backbone by Gibson assembly (Gibson et al., 2009). The final construct included an N-terminal 6xHis-tag on CarA followed by a Thr-Asp-Ala-Leu-Arg-Ala linker. CPS mutant plasmids were constructed using the Q5 Site-Directed Mutagenesis Kit (NEB).

The correct sequences for all constructs were confirmed by Sanger sequencing.

Protein purification

Wild-type and variant ProAs were expressed in strain AM209 (BL21(DE3) argC::kan′ proA::cat) to enable purification in the absence of wild-type ProA and ArgC. Carbamoyl phosphate synthetase (CPS) consists of a stable complex between CarA and CarB. Thus, in order to purify CPS, both carA and carB were co-expressed on the same plasmid with a His-tag on CarA. Wild-type and variant CPSs were expressed in strain AM267 (BL21 carAB::kan′) to enable purification in the absence of wild-type CPS.

Enzymes were expressed and purified using the following protocol with minor variations between the proteins. A small scraping from the glycerol stock of each expression strain was used to inoculate LB plus appropriate antibiotics. The cultures were grown overnight with
shaking at 37 °C. Overnight cultures were diluted 1:100 into 500 mL-2 L of LB plus appropriate antibiotic and grown with shaking at 37 °C. Once the OD_{600} reached 0.5-0.9, IPTG was added to a final concentration of 0.5 mM. Growth was continued at 30 °C for 5 hrs while shaking. Cells were harvested by centrifugation at 5,000 x g for 20 min at 4 °C. Cell pellets were stored at -70 °C until protein purification.

Frozen cell pellets were resuspended in 5x the cell pellet weight of ice-cold 20 mM sodium phosphate, pH 7.4, containing 300 mM NaCl and 10 mM imidazole. Fifty µL of protease inhibitor cocktail (Sigma-Aldrich, P8849) was added for each gram of cell pellet. Lysozyme was added to a final concentration of 0.2 mg/mL and the cells were lysed by probe sonication (20 sec of sonication followed by 30 sec on ice, repeated 3 times). Cell debris was removed by centrifugation at 18,000 x g for 20 min at 4 °C. The soluble fraction was then loaded onto 1 mL or 3 mL HisPur Ni-NTA Spin Columns (Thermo Scientific) and His-tagged protein was purified according to the manufacturer’s protocol. If the soluble fraction was larger than the column volume, the initial binding step was repeated before washing to accommodate all the soluble fraction. Bound protein was eluted with one column volume of 20 mM sodium phosphate, pH 7.4, containing 300 mM NaCl and increasing amounts of imidazole (100 mM, 250 mM, and finally 500 mM). Two separate elutions were performed with 500 mM imidazole. Fractions containing the protein of interest were pooled and dialyzed overnight against 6-12 L of exchange buffer at 4 °C. (ProA and ArgC were dialyzed against 20 mM potassium phosphate, 20 mM DTT, pH 7.5. N-acetylornithine aminotransferase was dialyzed against 20 mM potassium phosphate, pH 7.5. CPS was dialyzed against 100 mM potassium phosphate, pH 7.6. Ornithine transcarbamoylase was dialyzed against 20 mM Tris-acetate, pH 7.5.) Protein purity was assessed by SDS-PAGE and concentration measured using the Qubit protein assay kit with a
Qubit 3.0 fluorometer (Invitrogen). Purified protein was stored at 4 °C for short-term storage, and frozen in liquid nitrogen and stored at -70 °C for long-term storage.

**GSA and NAGSA dehydrogenase assays**

The native and neo-ArgC activities of ProA were assayed in the reverse direction (dehydrogenase reaction) because the lability of the forward substrates γ-glutamyl phosphate and N-acetylglutamyl phosphate make them difficult to purify. The change in the dehydrogenase activity due to a mutation is proportional to the change in the reductase activity according to the Haldane relationship (Haldane, 1930; McLoughlin & Copley, 2008).

Assaying ProA’s dehydrogenase activity using γ-glutamyl semialdehyde (GSA) and N-acetylglutamyl semialdehyde (NAGSA) as substrates is complicated by the equilibrium of GSA and NAGSA with their hydrated forms, as well as GSA’s intramolecular cyclization to form pyrroline-5-carboxylate (P5C) (Bearne & Wolfenden, 1995; Mezl & Knox, 1976). In order to measure the concentration of the free aldehyde form of these substrates, we mixed 15 μM ProA or ArgC with 2 mM “GSA” (including the hydrate and P5C) or 2 mM “NAGSA” (including the hydrate), respectively, in a solution containing 100 mM potassium phosphate, pH 7.6, and 1 mM NADP+ and measured the burst in NADPH production (Khanal et al., 2015). The concentrations of GSA+P5C+hydrate or NAGSA+hydrate were determined using the o-aminobenzaldehyde assay (Albrecht et al., 1962; Mezl & Knox, 1976). The absorbance at 340 nm due to formation of NADPH exhibited a burst followed by a linear phase when measured for 60 seconds. We assume that the burst corresponds to reduction of the free aldehyde form of GSA or NAGSA and the rate of the linear phase is determined by the conversion of the hydrate (and P5C in the case of GSA).
to the free aldehyde. We calculated the magnitude of the burst by fitting either all of the data or the linear portion of the data to one of the following equations.

\[ f(x) = mx + b \]
\[ f(x) = mx + b(1 - e^{-x}) \]

where \( x \) is time in seconds, \( m \) is the slope of the linear phase, and \( b \) is the magnitude of the burst and thus proportional to the starting concentration of the free aldehyde form of the substrate. In the case of the linear fit, only the linear portion of the \( A_{340} \) data was used. Eq. [8] was used to calculated NAGSA free aldehyde concentration because the exponential equation did not fit the data well. Eq. [9] was used to calculated GSA free aldehyde concentration. We repeated the assay three times and averaged the magnitude of the burst to calculate free aldehyde concentrations for solutions of GSA and NAGSA (under these buffer and temperature conditions) of 4.5% and 4.2% of the total concentration of free aldehyde + hydrate (+ P5C for GSA), respectively.

GSA and NAGSA dehydrogenase activities were measured by monitoring the appearance of NADPH at 340 nm in reaction mixtures containing 100 mM potassium phosphate, pH 7.6, 1 mM NADP\(^+\), varying concentrations of NAGSA or GSA, and catalytic amounts of ProA, ProA*, and ProA**. All kinetic measurements were done at 25 °C. Values for \( K_M \) refer to the concentration of the free aldehyde form of the substrate.

Assays for carbamoyl phosphate synthetase activity and allostERIC regulation

Kinetic assays for carbamoyl phosphate synthetase (CPS) were carried out with minor modifications of the methods described in Pierrat and Raushel 2002. The rate of ATP hydrolysis was measured at 37 °C by coupling production of ADP to oxidation of NADH using pyruvate kinase, which converts ADP and PEP to ATP and pyruvate, and lactate dehydrogenase, which
reduces pyruvate to lactate. Loss of NADH was monitored at 340 nm. Reaction mixtures consisted of 50 mM HEPES, pH 7.5, containing 10 mM MgCl$_2$, 100 mM KCl, 20 mM potassium bicarbonate, 10 mM L-glutamine, 1 mM PEP, 0.2 mM NADH, saturating amounts of pyruvate kinase and lactate dehydrogenase (Sigma-Aldrich, P0294), and varying amounts of ATP (0.01 to 8 mM). Reactions were initiated by adding 0.2 µM CPS to a final concentration of 0.2 µM. The effects of UMP and ornithine were measured under the same reaction conditions but with a fixed ATP concentration of 0.2 mM and varying concentrations of either UMP or ornithine.

Carbamoyl phosphate production was measured with minor modifications of previously described procedures (Snodgrass & Parry, 1969; Stapleton et al., 1996). Formation of carbamoyl phosphate by CPS was coupled with formation of citrulline by ornithine transcarbamoylase; citrulline forms a yellow complex ($\varepsilon_{464} = 37800 \text{ M}^{-1} \text{ cm}^{-1}$ (Snodgrass & Parry, 1969)) when mixed with diacetyl monoxime and antipyrine. Reaction mixtures consisted of 50 mM HEPES, pH 7.5, 10 mM MgCl$_2$, 100 mM KCl, 20 mM potassium bicarbonate, 10 mM L-glutamine, 4 mM ATP, 10 mM L-ornithine, and 0.7 µM ornithine transcarbamoylase. Reactions (0.25 mL) were initiated by adding CPS at a final concentration of 0.2 µM. After incubation for 2.5 min at 37°C, reactions were quenched by addition of 1 mL of a solution consisting of 25% concentrated H$_2$SO$_4$, 25% H$_3$PO$_4$ (85%), 0.25% (w/v) ferric ammonium sulfate, and 0.37% (w/v) antipyrine, followed by addition of 0.5 mL of 0.4% (w/v) diacetyl monoxime/7.5% (w/v) NaCl. The quenched reaction mixtures were placed in a boiling water bath for 15 min before measurement of OD$_{464}$. Control reactions contained all components except CPS.
RNA structure prediction

RNA secondary structures for argB mRNAs were predicted using CLC Main Workbench 8.1. The entire intergenic region between kan' and argB plus the first 33 nucleotides of argB were included in the structure prediction. The first 33 nucleotides were included because an mRNA-bound ribosome prevents another ribosome from binding to the mRNA until it has moved past the first 33 nucleotides (Steitz, 1969). Thus, at least the first 33 nucleotides are available for folding with the upstream region when a mRNA is being translated.

Calculation of RNA folding times

Folding times for a 63-nucleotide region (30 nt downstream and 30 nt upstream of the argB start codon) surrounding the start codon of argB mRNAs were calculated using the Kinfold program (v1.3) from the ViennaRNA v2.4.11 package (Wolfinger et al., 2004). Kinfold utilizes a Monte Carlo algorithm to calculate the folding time of each RNA sequence to the lowest free energy structure. We simulated 500 folding trajectories for each structure.

Acknowledgements

This work was supported by the NASA Astrobiology Institute (NNA15BB04A). We thank Craig Joy (University of Colorado Boulder, Physics Department) and Chris Takahashi (University of Washington) for help building the turbidostat.

Funding

| Funder                           | Grant reference number | Author                  |
|----------------------------------|------------------------|-------------------------|
| National Aeronautics and Space Administration | NNA15BB04A            | Shelley D. Copley       |
The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

**Additional files**

**Supplementary file 1.** Mutations found during the adaptation experiment. Sheet 1: mutations that were present at frequencies greater than 30% or that appeared in different populations. Sheet 2: genome positions of the amplified regions surrounding proA*. Sheet 3: other amplified or deleted genomic regions in the adapted populations.

**Supplementary file 2.** Supplementary Materials and Methods.

**Data availability**

The genome sequence of *E. coli* strain AM187 used in this study has been deposited to NCBI GenBank under accession number CP037857.

**Supplementary file 2: Supplementary Materials and Methods**

**Table S1.** Plasmids used in this study.

| Plasmid | Use/Description | Source/Ref |
|---------|-----------------|------------|
| pSLTS   | scarless genome editing/arabinose-inducible λ Red recombinase, anhydrotetracycline-inducible meganuclease I-SceI, temperature-sensitive origin of replication, ampicillin resistance | (Kim et al., 2014) |
| pSIM5   | introduction of linear DNA fragments into the genome/heat-inducible λ Red recombinase, temperature sensitive origin of replication, chloramphenicol resistance | (Datta et al., 2006) |
pSIM27 introduction of linear DNA fragments into the genome/heat-inducible λ Red recombinase, temperature sensitive origin of replication, tetracycline resistance (Datta et al., 2006)

pAM053 Cas9-mediated genome editing/constitutively-expressed cas9, heat-inducible λ Red recombinase, temperature-sensitive origin of replication, chloramphenicol resistance This study

pET-46 backbone for IPTG-inducible expression of wild-type and mutant ProAs/T7 promoter upstream of 6xHis-tag, ampicillin resistance Novagen

pCA24N backbone for IPTG-inducible expression of wild-type and mutant carbamoyl phosphate synthetases/chloramphenicol resistance (Kitagawa et al., 2005)

pCA24N-argI IPTG-inducible expression of ornithine transcarbamoylase/chloramphenicol resistance (Kitagawa et al., 2005)

pCA24N-argD IPTG-inducible expression of N-acetylornithine aminotransferase/chloramphenicol resistance (Kitagawa et al., 2005)

pAM068 expresses the Cas9 guide RNA that targets the upstream region of argB for introduction of the 58 bp deletion upstream of argB/temperature-sensitive origin of replication, ampicillin resistance This study

pAM100 expresses the Cas9 guide RNA that targets rph for introduction of the 82 bp deletion upstream of pyrE/temperature sensitive origin of replication, ampicillin resistance This study

Table S2. Primers used in this study.

| name     | sequence (5'→3')                  | description            |
|----------|-----------------------------------|------------------------|
| 1357     | CGTGCAGGCGATGGATAA                 | Forward primer for proA qPCR |
| 1358     | CTGTTCACGGCAGAGTATTT              | Reverse primer for proA qPCR |
| 1359     | CGTAGATCTGAGTGAATTT               | Forward primer for gyrB qPCR |
| 1360     | CGTTGGGTGGTCCGTTAGTA              | Reverse primer for gyrB qPCR |
| 1361     | CCCGTGGCTGAAAGTAAAA              | Forward primer for icd qPCR |
| 1362     | CAGGTTCATACAGGGCGATAAC            | Reverse primer for icd qPCR |
| hcaT_F   | CGTGGTGGCGAGGTAGTATTTC            | Forward primer for hcaT RT-qPCR |
| hcaT_R   | CGCCGAGATCAACAGCATATCG            | Reverse primer for hcaT RT-qPCR |
| cysG_F   | GGGCGCAAGAGGCTGGAAC              | Forward primer for cysG RT-qPCR |
| cysG_R   | GAATACCCGAATAACGAGGCAAG          | Reverse primer for cysG RT-qPCR |
| argB_F   | GCCGGGAACGCGCAATAAAAA            | Forward primer for argB RT-qPCR |
| argB_R   | CCGTACCGAGAAACAAAACC             | Reverse primer for argB RT-qPCR |
| argH_F   | TGGTGAAAACCGGCAAAACAAAC          | Forward primer for argH RT-qPCR |
| argH-R   | GCGCCAGCATCTCAACATAG             | Reverse primer for argH RT-qPCR |
| Primer | Sequence | Description |
|--------|----------|-------------|
| rAM209 | GACATAGCGTTGGCTACCCG | Forward primer for amplifying linear mutation fragment for the 58 bp deletion upstream of argB |
| rAM192 | CAGCCCTTTTACATCAGCTCATCC | Reverse primer for amplifying linear mutation fragment for the 58 bp deletion upstream of argB |
| rAM307 | AGCGGTTTGCGATCTGGAAT | Forward primer for amplifying linear mutation fragment for the 82 bp deletion in rph upstream of pyrE |
| rAM308 | ATGGTTTCATGCCTTCGCTC | Reverse primer for amplifying linear mutation fragment for the 82 bp deletion in rph upstream of pyrE |

**Table S3.** Protospacers used for Cas9-mediated scarless genome editing.

| Genome Modification | gRNA Plasmid | Protospacer (5'→3') |
|---------------------|--------------|---------------------|
| 58 bp deletion upstream of argB | pAM068 | GAGACTGCGTTTCATTGAGGC |
| 82 bp deletion in rph upstream of pyrE | pAM100 | GCTACTCATCTTTGTTGCTC |

**Table S4.** Mutation cassettes used for Cas9-mediated scarless genome editing. Brackets represent the primer annealing regions for amplifying the mutation cassettes.

| Genome Modification | Sequence (5'→3') |
|---------------------|-----------------|
| 58 bp deletion upstream of argB | [GACATAGCGTTGGCTACCCG]TGATATTGCTGAAGAGCTTGGCGGC GAATGGGCTGACCGCCTTCTCTCGTGCTTTTGAGGTATCGCCCGCTCCCGA TTCGACGCGATCGCTTCTCTCTTCTTCTTTGTGACAGGTCTTCTCTTATA AAGGGGAATTTAAAGATTCCTAATTCCCAAATGGGACGCGATCTGCTGAGT AATGACAGGGAACGCTTTGCTTGATATCGCACTGTGAATTTC GTGACGTCAGCAGCTCGTCTGCTATTGTGACAGGCGCGCTTTGCGT TTAAAGGCAGGTAGCTGATGAAAGGGCTG |
| 82 bp deletion in rph upstream of pyrE | [AGCGGTTTGCGATCTGGAAT]ACGTGAAGACTCTTCGCCGAGAGAC CGACATGAACGCAGCTGCTGACCGCAGAGACGCGGCGCATCTGAGT GCAAGGGAGCCAGAAGAGGCGCCTGCGAAAATCGATTTTTAAAGGCC GACTGATGAGTCGCTTCTTTTTGTGCAAGGAAGATAGATAGAG[GA GCGAAGGCGCATGAAACCAG] |

**Genome editing**

Strain AM187 was first transformed with a helper plasmid (pAM053, Table S1) encoding cas9 under the control a weak constitutive promoter (pro1 from Davis, Rubin, and Sauer 2011),
λ Red recombinase genes (*exo, gam, and bet*) under the control of a heat-inducible promoter, and a temperature-sensitive origin of replication (Datta et al., 2006). These cells were grown to an OD$_{600}$ of 0.2-0.4 at 30 °C and then incubated at 42 °C with shaking for 15 min to induce expression of the λ Red recombinase genes. The cells were immediately subjected to electroporation with 100 ng of either pAM068 or pAM100, which encode guide RNAs targeting 20-nucleotide sequences upstream of *argB* or in *rph*, respectively (Table S3), and 450 ng of a linear homology repair template that will introduce the desired deletion into the genome (Table S4). (Linear homology repair templates were amplified from genomic DNA of clones isolated during the adaptation experiment that contained the desired deletions and and the PCR fragments were gel-purified. Primers used to generate the linear DNA mutation fragments are listed in Table .) The cells were allowed to recover at 30 °C for 2-3 hours before being spread onto LB/ampicillin plates. Surviving colonies contained the guide RNA plasmid and the desired deletion, as confirmed by Sanger sequencing. pAM053 and the guide RNA plasmids, both of which have temperature-sensitive origins of replication, were cured by growth of individual colonies at 37 °C.

Strain AM209 was constructed from *E. coli* BL21(DE3) for expression of wild-type and mutant ProAs. We deleted *argC* and *proA* to ensure that any activity measured during *in vitro* assays was not due to trace amounts of ArgC or wild-type ProA. To accomplish these deletions, we amplified and gel-purified DNA fragments containing antibiotic resistance genes (kanamycin and chloramphenicol for deletion of *argC* and *proA*, respectively) flanked by 200-400 bp of sequences homologous to the upstream and downstream regions of either *argC* or *proA*. *E. coli* BL21(DE3) cells containing pSIM27 (Datta et al., 2006) – a vector containing heat-inducible λ Red recombinase genes – were grown in LB/tetracycline to an OD of 0.2-0.4 and then incubated
in a 42 °C shaking water bath for 15 min to induce expression of λ Red recombinase genes. The cells were then immediately subjected to electroporation with 100 ng of the appropriate linear DNA mutation cassette. Successful transformants were selected on either LB/kanamycin or LB/chloramphenicol plates.

Strain AM267 was constructed from *E. coli* BL21(DE3) for expression of wild-type and mutant carbamoyl phosphate synthetases (CPS). We deleted *carAB* to ensure that any activity measured during *in vitro* assays was not due to trace amounts of wild-type CPS. To accomplish the deletion, we amplified and gel-purified a DNA fragment containing the kanamycin resistance gene flanked by 40 bp of sequence homologous to the upstream and downstream regions of either *argC* or *proA*. *E. coli* BL21 cells containing pSIM5 (Datta et al., 2006) – a vector carrying heat-inducible λ Red recombinase genes – were grown in LB/chloramphenicol to an OD of 0.2-0.4 and then incubated in a 42 °C shaking water bath for 15 min to induce expression of λ Red recombinase genes. The cells were then immediately subjected to electroporation with 100 ng of the appropriate linear DNA mutation cassette. Successful transformants were selected on either LB/kanamycin plates.

**Measurement of proA* copy number**

The primer sets used for each gene are listed in Table S2. PowerSYBR Green PCR master mix (Thermo Scientific) was used according to the manufacturer’s protocol. A standard curve using variable amounts of AM187 genomic DNA was run on every plate to calculate efficiencies for each primer set. Primer efficiencies were calculated with the following equation:

\[
E_x = 10^{-\left(\frac{1}{m}\right)}
\]
where $E$ is the efficiency of primer set $x$, and $m$ is the slope of the plot of $C_t$ vs. starting quantity for the standard curve. $proA^*$ copy number was then calculated with the following equation (Hellemans et al., 2007):

$$n = \frac{E_{proA}^\Delta C_{t,proA}}{\sqrt{E_{gyrB}^-\Delta C_{t,gyrB} \times E_{icd}^-\Delta C_{t,icd}}}$$

where $n$ is the $proA^*$ copy number, and $\Delta C_{t,x}$ is the difference in $C_t$’s measured during amplification of AM187 and sample genomic DNA with primer set $x$.

**Measurement of argB and argH gene expression by RT-qPCR**

RNA was purified using the Invitrogen PureLink RNA Mini Kit according to the manufacturer’s protocol. The cell lysate produced during the PureLink protocol was homogenized using the QIAshredder column (Qiagen) prior to RNA purification. After RNA purification, each sample was treated with TURBO DNase (Invitrogen) according to the manufacturer’s protocol. Reverse transcription (RT) was performed with 250-600 ng of RNA using SuperScript IV VILO (Invitrogen) master mix according to the manufacturer’s protocol. qPCR of cDNA was performed to measure the fold-change in expression of $argB$ and $argH$ in mutant strains compared to that in AM187. $hcaT$ and $cysG$ were used as reference genes (Zhou et al., 2011). The primer sets used for each gene are listed in Table S2. A standard curve using variable amounts of *E. coli* BW25113 genomic DNA was run to calculate the primer efficiencies for each primer set. Fold-changes in expression of $argB$ and $argH$ were calculated as described above for calculations of $proA^*$ copy number.
References

Adler, M., Anjum, M., Berg, O. G., Andersson, D. I., & Sandegren, L. (2014). High fitness costs and instability of gene duplications reduce rates of evolution of new genes by duplication-divergence mechanisms. *Molecular Biology and Evolution, 31*(6), 1526–1535. https://doi.org/10.1093/molbev/msu111

Albalat, R., & Cañestro, C. (2016). Evolution by gene loss. *Nature Reviews Genetics, 17*(7), 379–391. https://doi.org/10.1038/nrg.2016.39

Albrecht, A. M., Scher, W. I., & Vogel, H. J. (1962). Determination of aliphatic aldehydes by apectrophotometry. *Analytical Chemistry, 34*(3), 398–400. https://doi.org/10.1021/ac60183a028

Anderson, P., & Roth, J. (1981). Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (rrn) cistrons. *Proceedings of the National Academy of Sciences, 78*(5), 3113–3117. https://doi.org/10.1073/pnas.78.5.3113

Andersson, D. I., Jerlström-Hultqvist, J., & Näsvall, J. (2015). Evolution of New Functions De Novo and from Preexisting Genes. *Cold Spring Harbor Perspectives in Biology, 7*(6). https://doi.org/10.1101/cshperspect.a017996

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., … Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology, 2*, 2006.0008. https://doi.org/10.1038/msb4100050

Barnhart, M. M., & Chapman, M. R. (2006). Curli biogenesis and function. *Annual Review of Microbiology, 60*(1), 131–147. https://doi.org/10.1146/annurev.micro.60.080805.142106

Baur, H., Tricot, C., Stalon, V., & Haas, D. (1990). Converting catabolic ornithine

54
carbamoyltransferase to an anabolic enzyme. *Journal of Biological Chemistry*, 265(25), 14728–14731.

Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R. K. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS ONE*, 10(5), 1–15. https://doi.org/10.1371/journal.pone.0128036

Bearne, S. L., & Wolfenden, R. (1995). Glutamate gamma-semialdehyde as a natural transition state analogue inhibitor of *Escherichia coli* glucosamine-6-phosphate synthase. *Biochemistry*, 34(36), 11515–11520.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., & Blüthgen, N. (2013). Efficient translation initiation dictates codon usage at gene start. *Molecular Systems Biology*, 9(675), 1–10. https://doi.org/10.1038/msb.2013.32

Bergthorsson, U., Andersson, D. I., & Roth, J. R. (2007). Ohno’s dilemma: evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences of the United States of America*, 104(43), 17004–17009. https://doi.org/10.1073/pnas.0707158104

Blank, D., Wolf, L., Ackermann, M., & Silander, O. K. (2014). The predictability of molecular evolution during functional innovation. *Proceedings of the National Academy of Sciences*, 111(8), 3044–3049. https://doi.org/10.1073/pnas.1318797111

Bonekamp, F., Clemmesen, K., Karlström, O., & Jensen, K. F. (1984). Mechanism of UTP-modulated attenuation at the pyrE gene of *Escherichia coli*: an example of operon polarity control through the coupling of translation to transcription. *The EMBO Journal*, 3(12), 2857–2861. https://doi.org/10.1002/j.1460-2075.1984.tb02220.x

Bouquin, A. Le, & Khila, A. (2017). *Taxon-restricted genes at the origin of a novel trait allowing access to a new environment*. 390(October), 386–390.
Braxton, B. L., Mullins, L. S., Raushel, F. M., & Reinhart, G. D. (1999). Allosteric dominance in carbamoyl phosphate synthetase. *Biochemistry, 38*(5), 1394–1401. https://doi.org/10.1021/bi982097w

Chothia, C., Gough, J., Vogel, C., & Teichmann, S. A. (2003). Evolution of the protein repertoire. *Science, 300*(5626), 1701–1703. https://doi.org/10.1126/science.1085371

Conant, G. C., & Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews. Genetics, 9*(12), 938–950. https://doi.org/10.1038/nrg2482

Conrad, T. M., Joyce, A. R., Applebee, M. K., Barrett, C. L., Xie, B., Gao, Y., & Palsson, B. T. (2009). Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biology, 10*(10), 1–12. https://doi.org/10.1186/gb-2009-10-10-r118

Copley, S. D. (2003). Enzymes with extra talents: Moonlighting functions and catalytic promiscuity. *Current Opinion in Chemical Biology, 7*(2), 265–272. https://doi.org/10.1016/S1367-5931(03)00032-2

Copley, S. D. (2017). Shining a light on enzyme promiscuity. *Current Opinion in Structural Biology, 47*, 167–175. https://doi.org/10.1016/j.sbi.2017.11.001

Datta, S., Costantino, N., & Court, D. L. (2006). A set of recombineering plasmids for gram-negative bacteria. *Gene, 379*(1–2), 109–115. https://doi.org/10.1016/j.gene.2006.04.018

Davis, J. H., Rubin, A. J., & Sauer, R. T. (2011). Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research, 39*(3), 1131–1141. https://doi.org/10.1093/nar/gkq810

Deatherage, D. E., & Barrick, J. E. (2014). Identification of mutations in laboratory-evolved...
microbes from next-generation sequencing data using breseq. In L. Sun & W. Shou (Eds.), *Engineering and Analyzing Multicellular Systems: Methods and Protocols* (pp. 165–188). https://doi.org/10.1007/978-1-4939-0554-6_12

Eroglu, B., & Powers-Lee, S. G. (2002). Unmasking a functional allosteric domain in an allosterically nonresponsive carbamoyl-phosphate synthetase. *Journal of Biological Chemistry*, 277(47), 45466–45472. https://doi.org/10.1074/jbc.M208185200

Espah Borujeni, A., Channarasappa, A. S., & Salis, H. M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research*, 42(4), 2646–2659. https://doi.org/10.1093/nar/gkt1139

Espah Borujeni, A., & Salis, H. M. (2016). Translation initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism. *Journal of the American Chemical Society*, 138(22), 7016–7023. https://doi.org/10.1021/jacs.6b01453

Ferla, M. P., Brewster, J. L., Hall, K. R., Evans, G. B., & Patrick, W. M. (2017). Primordial-like enzymes from bacteria with reduced genomes. *Molecular Microbiology*, 105(4), 508–524. https://doi.org/10.1111/mmi.13737

Francino, M. P. (2005). An adaptive radiation model for the origin of new gene functions. *Nature Genetics*, 37(6), 573–577. https://doi.org/10.1038/ng1579

Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), 343–345. https://doi.org/10.1038/nmeth.1318

Glasner, M., Gerlt, J., & Babbitt, P. (2006). Evolution of enzyme superfamilies. *Current Opinion in Chemical Biology*, 10(5), 492–497. https://doi.org/10.1016/j.cbpa.2006.08.012
Goodman, D. B., Church, G. M., & Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. *Science, 342*(6157), 475–479. https://doi.org/10.1126/science.1241934

Goto, M., Agari, Y., Omi, R., Miyahara, I., & Hirotsu, K. (2003). Expression, purification and preliminary X-ray characterization of N-acetyl-γ-glutamyl-phosphate reductase from *Thermus thermophilus* HB8. *Acta Crystallographica - Section D Biological Crystallography, 59*(2), 356–358. https://doi.org/10.1107/S0907444902020802

Grenier, F., Matteau, D., Baby, V., & Rodrigue, S. (2014). Complete genome sequence of *Escherichia coli* BW25113. *Genome Announcements, 2*(5), 90005. https://doi.org/10.1128/genomeA.01038-14

Haldane, J. B. S. (1930). *Enzymes*. London, New York, Longmans, Green.

Hall, M. N., Gabay, J., Débarbouillé, M., & Schwartz, M. (1982). A role for mRNA secondary structure in the control of translation initiation. *Nature, 295*(5850), 616–618. https://doi.org/10.1038/295616a0

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., & Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology, 8*(2), R19. https://doi.org/10.1186/gb-2007-8-2-r19

Howard, J. A. L., Delmas, S., Ivančić-Baće, I., & Bolt, E. L. (2011). Helicase dissociation and annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein. *The Biochemical Journal, 439*(1), 85–95. https://doi.org/10.1042/BJ20110901

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 256*(1346), 119–
Hughes, C. S., Foehr, S., Garfield, D. A., Furlong, E. E., Steinmetz, L. M., & Krijgsveld, J. (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular Systems Biology, 10*(10), 757–757. https://doi.org/10.15252/msb.20145625

Jee, J., Rasouly, A., Shamovsky, I., Akivis, Y., Steinman, S. R., Mishra, B., & Nudler, E. (2016). Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. *Nature, 534*(7609), 693–696. https://doi.org/10.1038/nature18313

Jensen, K. F. (1993). The *Escherichia coli* K-12 “wild types” W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. *Journal of Bacteriology, 175*(11), 3401–3407. https://doi.org/10.1128/jb.175.11.3401-3407.1993

Juárez-Vázquez, A. L., Edirisinghe, J. N., Verduzco-Castro, E. A., Michalska, K., Wu, C., Noda-García, L., … Barona-Gómez, F. (2017). Evolution of substrate specificity in a retained enzyme driven by gene loss. *eLife, 6*, 1–21. https://doi.org/10.7554/eLife.22679

Kershner, J. P., Yu McLoughlin, S., Kim, J., Morgenthaler, A., Ebmeier, C. C., Old, W. M., & Copley, S. D. (2016). A synonymous mutation upstream of the gene encoding a weak-link enzyme causes an ultrasensitive response in growth rate. *Journal of Bacteriology, 198*(20), 2853–2863. https://doi.org/10.1128/JB.00262-16

Khanal, A., Yu McLoughlin, S., Kershner, J. P., & Copley, S. D. (2015). Differential effects of a mutation on the normal and promiscuous activities of orthologs: Implications for natural and directed evolution. *Molecular Biology and Evolution, 32*(1), 100–108. https://doi.org/10.1093/molbev/msu271

Khersonsky, O., & Tawfik, D. S. (2010). Enzyme promiscuity: A mechanistic and evolutionary
perspective. Annual Review of Biochemistry, 79(1), 471–505.

https://doi.org/10.1146/annurev-biochem-030409-143718

Kim, J., Webb, A. M., Kershner, J. P., Blaskowski, S., & Copley, S. D. (2014). A versatile and highly efficient method for scarless genome editing in Escherichia coli and Salmonella enterica. BMC Biotechnology, 14(1), 84. https://doi.org/10.1186/1472-6750-14-84

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., & Mori, H. (2005). Complete set of ORF clones of Escherichia coli ASKA library (A complete set of E. coli K-12 ORF archive): unique resources for biological research. DNA Research, 12(5), 291–299. https://doi.org/10.1093/dnares/dsi012

Knöppel, A., Knopp, M., Albrecht, L. M., Lundin, E., Lustig, U., Näsvall, J., & Andersson, D. I. (2018). Genetic adaptation to growth under laboratory conditions in Escherichia coli and Salmonella enterica. Frontiers in Microbiology, 9(April), 756. https://doi.org/10.3389/fmicb.2018.00756

Kugelberg, E., Kofoid, E., Reams, A. B., Andersson, D. I., & Roth, J. R. (2006). Multiple pathways of selected gene amplification during adaptive mutation. Proceedings of the National Academy of Sciences of the United States of America, 103(46), 17319–17324. https://doi.org/10.1073/pnas.0608309103

Legrain, C., & Stalon, V. (1976). Ornithine carbamoyltransferase from Escherichia coli W. Purification, structure and steady-state kinetic analysis. European Journal of Biochemistry, 63(1), 289–301. https://doi.org/10.1111/j.1432-1033.1976.tb10230.x

Ludovice, M., Martin, J. F., Carrachas, P., & Liras, P. (1992). Characterization of the Streptomyces clavuligerus argC gene encoding N-acetylglutamyl-phosphate reductase: expression in Streptomyces lividans and effect on clavulanic acid production. Journal of
McLoughlin, S. Y., & Copley, S. D. (2008). A compromise required by gene sharing enables survival: Implications for evolution of new enzyme activities. *Proceedings of the National Academy of Sciences of the United States of America, 105*(36), 13497–13502. https://doi.org/10.1073/pnas.0804804105

Mezl, V. A., & Knox, W. E. (1976). Properties and analysis of a stable derivative of pyrroline-5-carboxylic acid for use in metabolic studies. *Analytical Biochemistry, 74*(2), 430–440. https://doi.org/10.1016/0003-2697(76)90223-2

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., … Endy, D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods, 10*(4), 354–360. https://doi.org/10.1038/nmeth.2404

Näsvall, J., Sun, L., Roth, J. R., & Andersson, D. I. (2012). Real-time evolution of new genes by innovation, amplification, and divergence. *Science, 338*(6105), 384–387. https://doi.org/10.1126/science.1226521

Nei, M., & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics, 39*(1), 121–152. https://doi.org/10.1146/annurev.genet.39.073003.112240

O’Brien, P. J., & Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chemistry & Biology, 6*(4), R91–R105. https://doi.org/10.1016/S1074-5521(99)80033-7

Ohno, S. (1970). *Evolution by Gene Duplication*. https://doi.org/10.1007/978-3-642-86659-3

Olson, M. V. (1999). When less is more: Gene loss as an engine of evolutionary change. *The
Page, R., Nelson, M. S., von Delft, F., Elsliger, M.-A., Canaves, J. M., Brinen, L. S., … Wilson, I. A. (2003). Crystal structure of γ-glutamyl phosphate reductase (TM0293) from *Thermotoga maritima* at 2.0 Å resolution. *Proteins: Structure, Function, and Bioinformatics, 54*(1), 157–161. https://doi.org/10.1002/prot.10562

Pettersson, M. E., Sun, S., Andersson, D. I., & Berg, O. G. (2009). Evolution of new gene functions: simulation and analysis of the amplification model. *Genetica, 135*(3), 309–324. https://doi.org/10.1007/s10709-008-9289-z

Pierrat, O. A., & Raushel, F. M. (2002). A functional analysis of the allosteric nucleotide monophosphate binding site of carbamoyl phosphate synthetase. *Archives of Biochemistry and Biophysics, 400*(1), 34–42. https://doi.org/10.1006/abbi.2002.2767

Proft, T., & Baker, E. N. (2009). Pili in Gram-negative and Gram-positive bacteria — structure, assembly and their role in disease. *Cellular and Molecular Life Sciences, 66*(4), 613–635. https://doi.org/10.1007/s00018-008-8477-4

Rauwerdink, A., Lunzer, M., Devamani, T., Jones, B., Mooney, J., Zhang, Z.-J., … Dean, A. M. (2016). Evolution of a catalytic mechanism. *Molecular Biology and Evolution, 33*(4), 971–979. https://doi.org/10.1093/molbev/msv338

Reams, A. B., Kofoed, E., Savageau, M., & Roth, J. R. (2010). Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics, 184*(4), 1077–1094. https://doi.org/10.1534/genetics.109.111963

Reams, A. B., & Neidle, E. L. (2004). Selection for gene clustering by tandem duplication. *Annual Review of Microbiology, 58*(1), 119–142. https://doi.org/10.1146/annurev.micro.58.030603.123806
Rubino, S. D., Nyunoya, H., & Lusty, C. J. (1986). Catalytic domains of carbamyl phosphate synthetase. Glutamine-hydrolyzing site of *Escherichia coli* carbamyl phosphate synthetase. *The Journal of Biological Chemistry*, 261(24), 11320–11327.

Rubino, S. D., Nyunoya, H., & Lusty, C. J. (1987). In vivo synthesis of carbamyl phosphate from NH₃ by the large subunit of *Escherichia coli* carbamyl phosphate synthetase. *Journal of Biological Chemistry*, 262(9), 4382–4386.

Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 27(10), 946–950. https://doi.org/10.1038/nbt.1568

Scharff, L. B., Childs, L., Walther, D., & Bock, R. (2011). Local absence of secondary structure permits translation of mRNAs that lack ribosome-binding sites. *PLoS Genetics*, 7(6), e1002155. https://doi.org/10.1371/journal.pgen.1002155

Snodgrass, P. J., & Parry, D. J. (1969). The kinetics of serum ornithine carbamoyltransferase. *Journal of Laboratory and Clinical Medicine*, 73(6), 940–950.

Stapleton, M. A., Javid-Majd, F., Harmon, M. F., Hanks, B. A., Grahmann, J. L., Mullins, L. S., & Raushel, F. M. (1996). Role of conserved residues within the carboxy phosphate domain of carbamoyl phosphate synthetase. *Biochemistry*, 35(45), 14352–14361. https://doi.org/10.1021/bi961183y

Starr, T. N., Picton, L. K., & Thornton, J. W. (2017). Alternative evolutionary histories in the
sequence space of an ancient protein. *Nature, 549*(7672), 409–413.

https://doi.org/10.1038/nature23902

Steitz, J. A. (1969). Polypeptide chain initiation: Nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature, 224*(5223), 957–964.

https://doi.org/10.1038/224957a0

Storz, J. F. (2016). Gene duplication and evolutionary innovations in hemoglobin-oxygen transport. *Physiology, 31*(3), 223–232. https://doi.org/10.1152/physiol.00060.2015

Takahashi, C. N., Miller, A. W., Ekness, F., Dunham, M. J., & Klavins, E. (2015). A low cost, customizable turbidostat for use in synthetic circuit characterization. *ACS Synthetic Biology, 4*(1), 32–38. https://doi.org/10.1021/sb500165g

Thoden, J. B., Huang, X., Raushel, F. M., & Holden, H. M. (2002). Carbamoyl-phosphate synthetase: Creation of an escape route for ammonia. *Journal of Biological Chemistry, 277*(42), 39722–39727. https://doi.org/10.1074/jbc.M206915200

Thoden, J. B., Raushel, F. M., Wesenberg, G., & Holden, H. M. (1999). The binding of inosine monophosphate to *Escherichia coli* carbamoyl phosphate synthetase. *Journal of Biological Chemistry, 274*(32), 22502–22507. https://doi.org/10.1074/jbc.274.32.22502

Todd, A. E., Orengo, C. A., & Thornton, J. M. (2001). Evolution of function in protein superfamilies, from a structural perspective. *Journal of Molecular Biology, 307*(4), 1113–1143. https://doi.org/10.1006/jmbi.2001.4513

Wiśniewski, J. R., & Gaugaz, F. Z. (2015). Fast and sensitive total protein and peptide assays for proteomic analysis. *Analytical Chemistry, 87*(8), 4110–4116.

https://doi.org/10.1021/ac504689z

Wolfinger, M. T., Svrcek-Seiler, W. A., Flamm, C., Hofacker, I. L., & Stadler, P. F. (2004).
Efficient computation of RNA folding dynamics. *Journal of Physics A: Mathematical and General*, 37(17), 4731–4741. https://doi.org/10.1088/0305-4470/37/17/005

Zhou, K., Zhou, L., Lim, Q., Zou, R., Stephanopoulos, G., & Too, H.-P. (2011). Novel reference genes for quantifying transcriptional responses of *Escherichia coli* to protein overexpression by quantitative PCR. *BMC Molecular Biology*, 12(1), 18. https://doi.org/10.1186/1471-2199-12-18

Zwietering, M. H., Jongenburger, I., Rombouts, F. M., & van ’t Riet, K. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, 56(6), 1875–1881.