UC Davis
UC Davis Previously Published Works

Title

Evolution of a Genome-Encoded Bias in Amino Acid Biosynthetic Pathways Is a Potential Indicator of Amino Acid Dynamics in the Environment

Permalink

https://escholarship.org/uc/item/85q356nd

Journal

Molecular Biology and Evolution, 31(11)

ISSN

0737-4038

Authors

Fasani, Rick A
Savageau, Michael A

Publication Date

2014-11-01

DOI

10.1093/molbev/msu225

Peer reviewed

Evolution of a Genome-Encoded Bias in Amino Acid Biosynthetic Pathways Is a Potential Indicator of Amino Acid Dynamics in the Environment

Rick A. Fasani[1] and Michael A. Savageau*[,1]
[1]Department of Biomedical Engineering and Microbiology Graduate Group, University of California, Davis
*Corresponding author: E-mail: masavageau@ucdavis.edu.
Associate editor: Helen Piontkivska

Abstract

Overcoming the stress of starvation is one of an organism's most challenging phenotypic responses. Those organisms that frequently survive the challenge, by virtue of their fitness, will have evolved genomes that are shaped by their specific environments. Understanding this genotype–environment–phenotype relationship at a deep level will require quantitative predictive models of the complex molecular systems that link these aspects of an organism's existence. Here, we treat one of the most fundamental molecular systems, protein synthesis, and the amino acid biosynthetic pathways involved in the stringent response to starvation. These systems face an inherent logical dilemma: Building an amino acid biosynthetic pathway to synthesize its product—the cognate amino acid of the pathway—may require that very amino acid when it is no longer available. To study this potential "catch-22," we have created a generic model of amino acid biosynthesis in response to sudden starvation. Our mathematical analysis and computational results indicate that there are two distinctly different outcomes: Partial recovery to a new steady state, or full system failure. Moreover, the cell's fate is dictated by the cognate bias, the number of cognate amino acids in the corresponding biosynthetic pathway relative to the average number of that amino acid in the proteome. We test these implications by analyzing the proteomes of over 1,800 sequenced microbes, which reveals statistically significant evidence of low cognate bias, a genetic trait that would avoid the biosynthetic quandary. Furthermore, these results suggest that the pattern of cognate bias, which is readily derived by genome sequencing, may provide evolutionary clues to an organism's natural environment.

Key words: cognate bias, genomes, proteomes, environmental stress, stringent response, cell fate.

Introduction

Predicting or measuring the natural microenvironment of an organism is a complex and challenging task (Savageau 1983; Ward et al. 1998; Xu 2006). In contrast, sequencing an organism's genome has become routine, and the scientific community continues to sequence organisms from varied environments at an increasing pace, yet the vast majority of those organisms cannot be cultured by current methods, in part because their natural environment is unknown (Curtis and Sloan 2005; Xu 2006). Genomic data are clearly outpacing environmental data, but the sequences themselves may provide information about the environment from which they were taken. Recent statistical analyses of overall amino acid composition across organisms indicate that the environment is a major evolutionary influence (Chen et al. 2013; Moura et al. 2013). More specifically, the cognate bias hypothesis (Alves and Savageau 2005) suggests that nutritional stress places evolutionary pressure on the composition of the enzymes in the amino acid biosynthetic pathways. (The "cognate amino acid" refers to the amino acid produced by the corresponding biosynthetic pathway, and the "cognate bias" refers to the number of cognate amino acids in the enzymes of the corresponding biosynthetic pathway relative to their number in the proteome of the organism.)

The bacterial response to nutritional stress, the well-known stringent response, has been studied for over five decades. In 1961, it was shown that amino acid starvation inhibited the accumulation of stable RNA, and the locus responsible was christened RC, or the RNA Control gene (Stent and Brenner 1961), later renamed *relA* in reference to the relaxed response of the mutant phenotype (Friesen et al. 1974). Today, it is clear that the stringent response is a general reaction to stress and starvation that is conserved across species (Draper 1996; Harris et al. 1998; van der Biezen et al. 2000; Chatterji and Ojha 2001), and is characterized by increased levels of guanosine tetraphosphate (ppGpp) (Cashel 1969; Wendrich et al. 2002), which has at least 75 known effects in *Escherichia coli*, including decreased rRNA and tRNA transcription, decreased growth rate, and increased expression of the biosynthetic enzymes for many amino acids (Draper 1996; Magnusson et al. 2005; Potrykus and Cashel 2008; Dalebroux and Swanson 2012). However, the stringent response may not be enough to protect the cell from the shock of starvation. Part of the response is the upregulation of amino acid biosynthetic pathways, but the situation creates a potential catch-22. The missing amino acid could hold up the construction of the enzymes needed to create more of their cognate amino acid, a stalemate from

Open Access

Article

which the cell might not recover. A logical evolutionary defense would be to remove the vulnerability—to bias the biosynthetic enzymes against the use of their cognate amino acid.

Our first hint that organisms might evolve such a molecular mechanism came in the early days of protein sequencing when tryptophan synthetase, an enzyme in the tryptophan biosynthetic pathway, was sequenced and the alpha subunit was found to contain no tryptophan (Yanofsky 1988). Hardly conclusive, it took many years and a new technology to begin testing the hypothesis more generally. Additional support for what we termed the cognate bias hypothesis was obtained from genomic data for three well-studied bacteria: Two enteric bacteria, E. coli and Salmonella enterica (serovar Typhimurium), and the soil-dwelling bacteria Bacillus subtilis (Alves and Savageau 2005). The results suggested a bias toward fewer cognate amino acids in certain amino acid biosynthetic pathways and a profile of bias across amino acids that differed between the two groups, suggesting a possible correlation with the organisms' ecological niches. The cognate bias hypothesis was recently tested and confirmed using a few organisms from the other domains of life: Methanococcus jannaschii, Saccharomyces cerevisiae, and Homo sapiens (Perlstein et al. 2007; Meiler et al. 2012). However, the number of organisms examined to date is too few to provide for a meaningful statistical test.

Complete genome sequences are now available for more than 1,800 microorganisms. Thus it is an opportune time to go beyond correlations and comprehensively examine the cognate bias hypothesis, with an analysis of the underlying molecular mechanism using a kinetic model of protein synthesis and amino acid starvation, in order to provide a stronger molecular link between the genomic evidence of amino acid composition and the environmental dynamics of amino acid availability.

In particular, we are interested in the environment's effect on the proteome, in terms of three classes of amino acids: 1) For amino acids the organism is never required to create, it could dispense with the biosynthetic pathway entirely (e.g., as occurs with obligate intracellular parasites that receive the amino acid from their host, and with humans that receive their essential amino acids from their diet); 2) for amino acids the organism is always required to create, it could dispense with the regulation and synthesize the amino acids constitutively, without regard for cognate bias; and 3) for all other environments, regulation would be advantageous, and a compensating cognate bias would likely exist for amino acids that experience the most frequent and extreme fluctuations in the organism's natural environment. In order to study the last case more rigorously, we have created a generic model of amino acid biosynthesis and regulation in response to sudden starvation. The results indicate that there are two distinctly different outcomes—partial recovery or full failure—that are dictated by the cognate bias, the number of cognate amino acids in the corresponding biosynthetic pathway relative to the average number of that amino acid in the proteome. Furthermore, we mine the abundant genomic data that are currently available and reveal statistical evidence

of cognate bias. The results describe how the natural environment of an organism—or more precisely, the stresses and strains to which the organism is exposed—may leave a genetic fingerprint.

Results

Model of Translation during Starvation

Before describing a larger model of amino acid biosynthesis and regulation, we present a model of translation that accounts for the effect of starvation, or more specifically for a decreased supply of the cognate amino acid of interest. Mathematical models of the translational process have been created, but they are too detailed to be tractable within our larger model of regulation (Gromadski and Rodnina 2004; Elf and Ehrenberg 2005; Gilchrist and Wagner 2006; Fluitt et al. 2007; Shah and Gilchrist 2010; Brackley et al. 2011, 2012). Here, we consider a somewhat simpler approach. Translation proceeds through three well-known steps (Alberts et al. 2002). First, the ribosome, which is attached to the mRNA and a growing peptide chain, exposes an empty A-site. Second, a charged, aminoacyl-tRNA (aa-tRNA) fills the empty A-site. In fact, the aa-tRNA may leave the A-site, returning the system to the first step. Finally, the ribosome incorporates the amino acid into the peptide chain, discharges the uncharged tRNA, and advances. The sequence of steps is reminiscent of an enzymatic reaction and has in fact been modeled as such in the past by Elf and Ehrenberg (2005). However, they consider the incorporation of an amino acid in isolation—a single instance of those three steps—and use the result to represent the average rate of amino acid consumption in a larger model of global protein synthesis. Here, we consider the sequential incorporation of every amino acid in the protein as a longer sequence of enzymatic reactions, and consequently the combined impact of starvation at multiple steps in the synthesis of a single protein. When modeling translation as an enzymatic reaction, the ribosome–mRNA complex represents the catalyst. The transition between the first and second steps is treated as a reversible reaction between a complex with an empty A-site and an intermediate complex with an aa-tRNA in the A-site. The transition between the second and third steps—the addition of the amino acid to the peptide chain and the advancement of the ribosome—is treated as the essentially irreversible step in the enzymatic reaction. Figure 1 depicts the transitions between states. We assume that the concentration of aa-tRNA is coupled to the concentration of the free amino acid pool, so that when the supply of amino acid decreases, the concentration of aa-tRNA decreases proportionally, and the rate of its incorporation into the A-site decreases. In the limiting case where none of the particular amino acid is available, the ribosome remains stalled with an empty A-site. We assume that the longer the ribosome is stalled, the more likely it is to prematurely terminate translation and dissociate from the mRNA, which is consistent with models of nonsense errors produced by frameshift, ribosomal drop-off, or release factors (Jørgensen and Kurland 1990; Gilchrist and Wagner 2006; Shah and Gilchrist 2010)

and models of ribosome rescue by transfer-messenger RNA (Roche and Sauer 1999; Moore and Sauer 2007; Keiler 2008).

The complete translation of a protein with N amino acids can be modeled as a linked chain of N enzymatic reactions, as shown in figure 2. The entire process begins with initiation, when the ribosome binds to the mRNA, and ends with the release of the completed protein, which requires the presence of release factors. If all of the amino acid concentrations are high, the rate of abortion becomes vanishingly small, and the entire process can be viewed as an unbranched pathway, implying that at steady state, what goes in must come out. Assuming the concentrations of ribosome and all ancillary factors are constant or saturating, then the steady-state rate of protein production, without starvation and abortion, is simply $v_{out} = k_{in}M$, or proportional to the concentration of the specific mRNA. Nearly all models of biosynthetic gene circuits treat the rate of protein synthesis as proportional to the concentration of mRNA, and there is good evidence in bacterial literature to support this (Guet et al. 2008). On the other hand, if the supply of a particular amino acid is significantly decreased, then the ribosome will stall at each point that requires that amino acid (Pedersen 1984; Varenne et al. 1984; 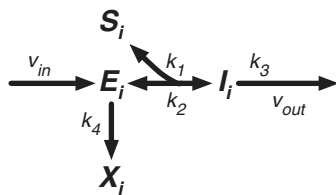Sørensen et al. 1989), increasing the rate of abortion and slowing the overall rate of protein production. The exact rate of protein production can be determined by an inductive argument. In the simplest case, where the protein only requires a single amino acid that is in short supply, the process has a single branch point at step i, as shown in figure 1. It can be shown that at steady state, v_{out} at step i is a function of v_{in} and the limiting amino acid concentration, or

$$v_{out} = v_{in}\left(\frac{S_i}{S_i + K_m}\right), \qquad (1)$$

where

$$K_m = \frac{k_4}{k_3}\left(\frac{k_2 + k_3}{k_1}\right). \qquad (2)$$

More intuitively, K_m is the amino acid concentration for half-maximal velocity through step i. v_{in} is a result of the proceeding steps, which in this case form an unbranched pathway, and is therefore $k_{in}M$ at steady state. The remaining steps also form an unbranched pathway, meaning that the overall rate of protein production is equal to the velocity through step i.

For a protein that requires more than one of the limiting amino acid, v_{out} at the first occurrence is identical to the case with one. v_{out} at the second occurrence is equal to the output of the proceeding steps multiplied by the same attenuating factor. As such, for a protein that requires n of the limiting amino acid, the overall rate of protein production is

$$v_{out} = k_{in}M\left(\frac{S}{S + K_m}\right)^n, \qquad (3)$$

where S is the concentration of the limiting amino acid. The implication of equation (3) is dramatic. Considering the average protein in *E. coli* contains approximately 300 amino acids, and assuming the 20 amino acids are equally represented, n is on the order of 15, making translation ultrasensitive to the limiting amino acid concentration. If the amino acid concentration drops below some critical threshold, K_m, then the rate of translation will practically halt.

The submodel above captures the essential features needed in our larger model of amino acid biosynthesis and regulation in the next section. One could conceivably add more detail, and attempt to account for such factors as tRNA abundance, specific constants for the binding of

Fig. 1. Model of a single branch point in translation. At step i, the advancement of the ribosome is modeled by a single enzymatic reaction. The required aa-tRNA S_i enters the A-site of the ribosome–mRNA–peptide complex E_i, to create a new intermediate complex I_i. The process is reversed if the aa-tRNA leaves the A-site. Otherwise, the amino acid is irreversibly linked to the growing peptide chain, advancing the complex to E_{i+1}. In the absence of aa-tRNA, the complex can dissociate from the mRNA, creating the aborted complex X_i. k_1, rate constant of the appropriate aa-tRNA entering the A-site; k_2, rate constant of the appropriate aa-tRNA leaving the A-site; k_3, rate constant of amino acid i being incorporated into the growing peptide chain; k_4, rate constant of the ribosome aborting and dissociating from the mRNA; v_{in}, rate of ribosomes advancing to the current position; v_{out}, rate of ribosomes proceeding to the next position.
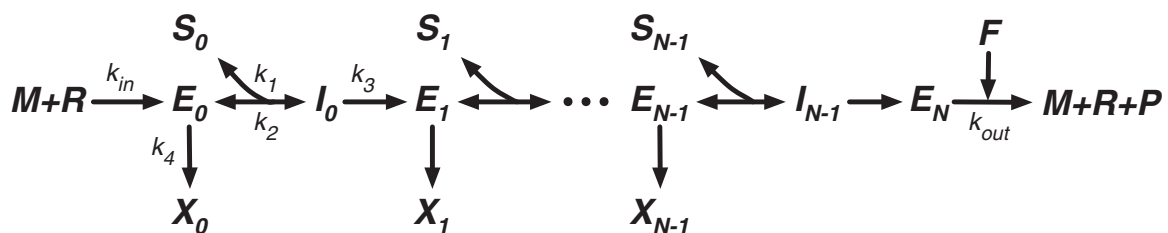

Fig. 2. Model of translation as an enzymatic process. The translation of a full protein is modeled by a series of enzymatic reactions, each as shown in figure 1. Free ribosome R binds to mRNA M and initiates translation. At each step i, the required aa-tRNA S_i is added to the ribosome–mRNA–peptide complex E_i that already has a peptide chain of length i. The final complex E_N is released by release factor F to create free ribosome R, mRNA M, and protein P. k_{in}, rate of translational initiation; k_{out}, rate of protein release. I_i, X_i, and k_1–k_4 are described in figure 1.

amino acid to various tRNAs and the binding of tRNA anticoding sites to codons in mRNA, salt and pH concentrations, or any other physical–chemcial aspect of the intracellular milieu that differs according to cell type (e.g., high pH in extreme halophilic archaea vs. low pH in halophobic bacteria). However, such detailed models would become so complex that analysis would be difficult if no precluded, most values of the parameters would not be available for most organisms, and even if these obstacles were overcome in a particular case, the results would not generalize to other systems.

Model of Amino Acid Biosynthesis and Regulation

Mathematical models of amino acid biosynthetic systems have been developed in the past, in many cases for specific systems, such as Trp biosynthesis in *E. coli* (Xiu et al. 2002; Alves and Savageau 2005; Elf and Ehrenberg 2005). Many of these models tend to be complex with idiosyncratic features that do not readily generalize to other systems, as discussed above for our submodel. For instance, the pathways may have different numbers of enzymes with very different kinetic properties. Rather, we require relatively simple models that nevertheless retain the essential generic character of amino acid biosynthetic systems, can be readily analyzed to make testable predictions, and can be used to elucidate general design principles. For example, models of inducible and repressible pathways, very similar to the one developed below, were used to make predictions regarding the coupling of expression in elementary gene circuits; the resulting predictions were confirmed experimentally in over 50 specific cases and the predicted coupling rules are now established as a general design principle (Hlavacek and Savageau 1996, 1997; Wall et al. 2003, 2004). Figure 3 depicts our model of amino acid biosynthesis, one that includes the transcription and translation of enzymes in the biosynthetic pathway, as well as the synthesis of the cognate amino acid. Feedback repression of the biosynthetic enzymes, which is a prominent control mechanism in bacteria (Neidhardt et al. 1996), is also included, as is the ability to import amino acid from the external environment. As was shown in the previous section, the rate of translation of the biosynthetic enzymes depends on the concentration of the free cognate amino acid, which is also depleted by cellular demand.

The model is mathematically described by a conventional system of ordinary differential equations (ODEs):

$$\frac{dX_1}{dt} = \alpha_1 X_4 \left(\frac{\sigma X_3^{g_{13}} + K_{13}}{X_3^{g_{13}} + K_{13}} \right) - \beta_1 X_1, \tag{4}$$

$$\frac{dX_2}{dt} = \alpha_2 X_1 X_5 \left(\frac{X_3}{X_3 + K_{23}} \right)^n - \beta_2 X_2, \tag{5}$$

$$\frac{dX_3}{dt} = \alpha_{31} X_2 X_6 + \alpha_{32} X_7 - \beta_3 X_8 \left(\frac{X_3}{X_3 + K_{23}} \right)^m. \tag{6}$$

X_1 represents the concentration of mRNA that encodes some critical enzyme of the biosynthetic pathway. Transcription is dependent on the cognate amino acid concentration, X_3, and
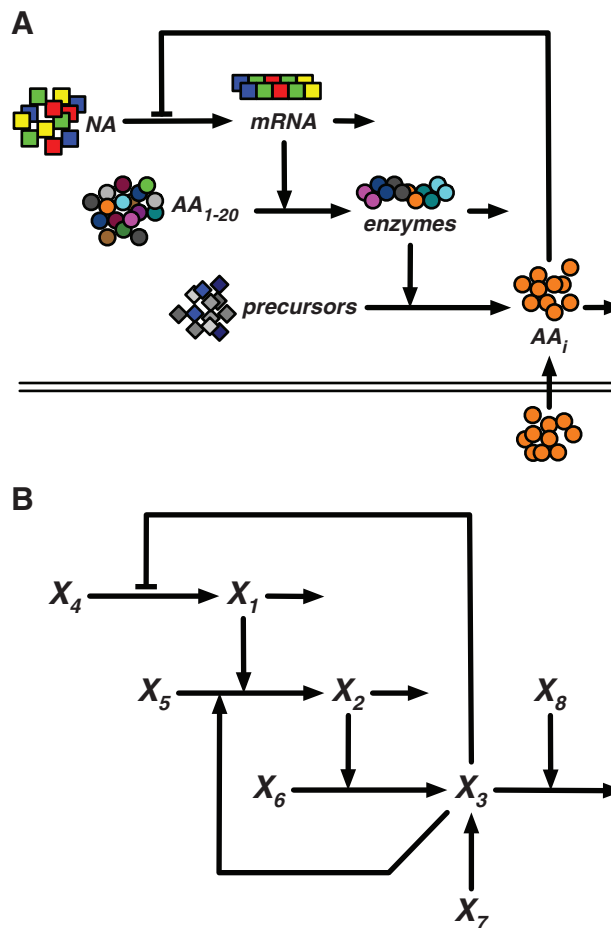

Fig. 3. Model of amino acid biosynthesis and regulation during starvation. (*A*) Model illustration of amino acid biosynthesis and regulation with an external supply. NA, nucleic acid precursors; mRNA, messenger RNA for the enzymes of the amino acid biosynthetic pathway; AA$_{1-20}$, free amino acid; AA$_i$, free cognate amino acid. (*B*) Abstract model of species concentrations and interactions used here. X_1, mRNA; X_2, critical enzyme of the amino acid biosynthetic pathway; X_3, free cognate amino acid; X_4, nucleic acid precursors; X_5, protein precursors; X_6, cognate amino acid precursors; X_7, external cognate amino acid; X_8, total cellular mRNA.

is described by a rational function with a Hill number of g_{13}, and a ratio between minimum and maximum rates $\sigma = V_{1L}/V_{1H}$. The loss of X_1 is dominated by first-order degradation of mRNA. X_2 represents the concentration of the critical enzyme, which is assumed to be stable. As was shown in the previous section, the rate of translation, or protein production, is drastically affected by the limiting amino acid concentration, X_3, and the exponent n is the number of cognate amino acids in the critical enzyme. The loss of X_2 is dominated by first-order dilution in an exponentially growing cell, and therefore β_2 is equal to the growth rate constant, which is in turn affected by the availability of free amino acid, or $\beta_2 = \mu = \mu_M [X_3/(X_3 + K_{23})]^m$, where μ_M is the maximum growth rate constant when X_3 is in excess. The free cognate amino acid concentration, X_3, can be increased by biosynthesis or import from an external supply, each represented by a positive term. The free cognate amino acid pool is

depleted by the cellular demand for amino acid, which we assume is dominated by protein synthesis. If the amino acids are quickly recycled from the aborted ribosome–peptide complex, then intuitively the rate of amino acid utilization is the rate of successful protein production, and the final term of equation (6) is therefore similar to the first term of equation (5). We assume that the exponent m is the average number of cognate amino acids in each protein of the expressed proteome, rather than n, the average number in the critical enzyme of the corresponding biosynthetic pathway. Furthermore, if the cell produces P proteins, each with an average number of cognate amino acids m, at a rate equal to that of the critical enzyme in the biosynthetic pathway, then $\beta_3 = Pm\alpha_2 X_5$.

To simplify the analysis, and without loss of generality, the variables and parameters are normalized with respect to initial concentrations and a chosen time constant: $x_i = X_i/X_{i0}$, $k_i = K_i/X_{30}$, and $\tau = t\mu_M$. The normalized system is described by equations (7–9):

$$\frac{dx_1}{d\tau} = A\left[\left(\frac{1+k_{13}}{\sigma+k_{13}}\right)\left(\frac{\sigma x_3^{g_{13}} + k_{13}}{x_3^{g_{13}} + k_{13}}\right) - x_1\right], \quad (7)$$

$$\frac{dx_2}{d\tau} = \left[x_1(1+k_{23})^n\left(\frac{x_3}{x_3+k_{23}}\right)^n - x_2(1+k_{23})^m\left(\frac{x_3}{x_3+k_{23}}\right)^m\right], \quad (8)$$

$$\frac{dx_3}{d\tau} = B\left[Cx_2 + (1-C)x_7 - (1+k_{23})^m\left(\frac{x_3}{x_3+k_{23}}\right)^m\right], \quad (9)$$

where $A = \frac{\beta_1}{\mu_M}$, $B = \frac{\beta_3 X_{80}}{\mu_M X_{30}}$, and $C = \frac{\alpha_{31}X_{60}X_{20}}{\alpha_{31}X_{60}X_{20} + \alpha_{32}X_{70}}$.

It should be noted that care is taken when normalizing the system. To create a well-controlled comparison between systems with different parameters, the terms were chosen to ensure that the corresponding reaction rates in each system were equal at the initial conditions. Simple inspection confirms that the gain and loss of each species at the initial conditions is unity, no matter what the parameter values.

Parameter Estimation

The mathematical model described by equations (7–9) contains nine parameters: A, B, C, g_{13}, σ, k_{13}, k_{23}, m, and n. Of the nine, five are aggregates of other parameters. Table 1 lists the estimated parameter values. Where possible, the parameters are estimated based on published data for *E. coli*. In the remaining cases, reasonable estimates are made to reflect expected operating conditions. The aggregate parameter A is calculated based on published values. Similarly, B is based on published values and a reasonable estimate for P. C is chosen to reflect a heavily repressed biosynthetic pathway during growth in an initial state of amino acid abundance. k_{13} and k_{23} represent the critical thresholds of transcriptional and translational regulation, respectively, and are normalized relative to the initial cognate amino acid concentration, X_{30}.

Their values are chosen so that the initial rate of transcription, based on X_{30}, is near minimum but still within the regulatory regime, whereas the initial rate of global translation, also based on X_{30}, is near maximum but still within the regulatory regime.

Dynamic Response to Starvation

We define the cognate bias as $n - m$, or the difference between the number of cognate amino acids in a critical enzyme of the corresponding biosynthetic pathway, n, and the average number in the proteome, m. If the critical enzyme of the pathway is compositionally similar to the rest of the proteome, then there is no bias, or $n - m = 0$. Low bias represents the case where there are relatively fewer cognate amino acids in the biosynthetic pathway, or $n - m < 0$, whereas high bias represents the case where there are more, or $n - m > 0$. To examine the effect of cognate bias, various systems with different biases were computationally simulated in response to rapid and complete starvation: The normalized external supply of amino acid, x_7, was decreased from 1 to 0 at $\tau = 0$. The results are shown in figure 4. All of the systems immediately experience a rapid drop in free cognate amino acid and a commensurate rise in the mRNA that encodes the biosynthetic pathway. The response is expected, as the large cellular demand quickly depletes the free amino acid reserves. Systems with low (blue) or no (green) cognate bias compensate by derepression and stabilize at a new steady-state amino acid concentration. Systems with a slight high bias (red and light blue) recover more slowly, but eventually stabilize as well (see the final values in fig. 4C). Systems with slightly higher bias (pink and yellow) do not recover, but do stabilize at lower amino acid concentrations. Figure 4C and D clearly shows how the final steady-state concentration and response times vary with bias: The higher the bias, the slower the response and the lower the final steady-state concentration of free amino acid. Similar behavior is observed when the parameter values are varied—A between 5 and 100, B between 10^4 and 10^6, and m between 4 and 16—indicating that the result is insensitive to the values of the parameters. Furthermore, the behavior changes significantly between $n = 32$ (yellow) and 36 (black), at which point the amino acid concentration does not appear to stabilize at all, but rather continues on a trajectory toward zero, suggesting the full failure of the system. However, the response time is extremely slow, and the large exponents magnify rounding errors, which may introduce uncertainty in the final steady-state values. Nevertheless, the concentrations do eventually reach zero, which is confirmed by the following analysis of the steady states.

Steady-State Concentrations after Starvation

At steady state, the derivatives that represent the changing concentrations vanish in equations (7–9). Also, complete starvation sends the external amino acid supply x_7 to zero. Simple inspection of the equations when $dx_i/d\tau$ and x_7 are zero reveals that $x_3 = 0$ is always a solution, meaning that a steady-state amino acid concentration of zero is a possibility. Subsequent manipulation of equations (7–9)

Table 1. Parameters Estimates for Amino Acid Biosynthesis.

Parameter	Description	Value Used	Sources
β_1	Rate constant of mRNA degradation	$1\,\text{min}^{-1}$	Leive and Kollin (1967); Blundell and Kennell (1974)
μ_M	Rate constant of maximum growth and enzyme dilution	$0.01\,\text{min}^{-1a}$	Bremer and Dennis (1996)
$\alpha_2 X_5$	Rate of initiation of translation	$20\,\text{min}^{-1b}$	Draper (1996)
X_{30}	Initial concentration of free cognate amino acid	$75\,\mu\text{M}$	Sundararaj et al. (2004)
X_{80}	Initial concentration of total mRNA	$1\,\mu\text{M}$	Sundararaj et al. (2004)
A	Normalized rate constant of mRNA degradation	100	
B	Normalized initial maximum supply of cognate amino acid	10^6	
C	Initial fraction of cognate amino acid autosynthesized	0.01	
g_{13}	Strength of repression	2	
σ	Potential capacity (inverse)	10^{-4c}	
k_{13}	Concentration for half-maximal velocity raised to the power g_{13} (K_{13} relative to $X_{30}^{g_{13}}$)	0.001	
k_{23}	Concentration for half-maximal velocity through step i (K_{23} relative to X_{30})	0.1	
m	Average number of cognate amino acids in a protein	16^d	Tatusova et al. (1999); Rudd (2000); Pruitt et al. (2005)

NOTE.—The parameter values are estimated based on published data for *Escherichia coli*, or reasonable estimates are made to reflect expected operating conditions.
[a]Rate of dilution dominates rate of degradation; slowest growth rate as observed on minimal media.
[b]Rate at which efficient mRNAs initiate translation of trp.
[c]Potential capacity may not be observed capacity.
[d]Based on average protein length and an amino acid composition of 0.05.

yields the following equation, in terms of the free amino acid concentration x_3 and the estimated parameters of the system:

$$C\left(\frac{1+k_{13}}{\sigma+k_{13}}\right)\left(\frac{\sigma x_3^{g_{13}}+k_{13}}{x_3^{g_{13}}+k_{13}}\right) = (1+k_{23})^{2m-n}\left(\frac{x_3}{x_3+k_{23}}\right)^{2m-n}. \tag{10}$$

There is no closed-form solution for x_3 given the general form of equation (10). However, when both sides of the equation, f_{left} and f_{right}, are plotted as shown in figure 5A, the intersections identify the values of x_3, or the normalized steady-state concentrations, that satisfy equation (10). In figure 5A, f_{left} is drawn for $g_{13} = 2$, whereas f_{right} is drawn for several different values of cognate bias, or $n - m$. Overall, there are six labeled intersections in figure 5A, and a numerical solver was used to verify and refine each of them, producing six potential steady-state concentrations for x_3: 0.013, 0.047, 0.090, 0.12, 0.17, and 0.28. However, the number of intersections is not equal to the number of curves—for some values of bias, f_{left} and f_{right} clearly intersect at a single point; for other values, the curves intersect at two points or not at all.

Figure 5A reveals three distinct cases of interest. When $n \leq 2m$, the system has only one intersection, and therefore one steady state, in addition to the zero steady state that was previously identified. The eigenvalues of the system linearized near the steady states confirm, as the simulations indicated, that the positive steady state is stable whereas the steady state at zero is unstable. This implies only one possible outcome for $n \leq 2m$: A low, albeit nonzero, stable steady-state amino acid concentration, which is considered safe. When n is slightly greater than $2m$, but still below some threshold m_{cr}, the system has two intersections in figure 5A, or two nonzero steady states, in addition to the ever-present steady state at

zero. The eigenvalues of the system linearized near these steady states confirm, as the simulations indicated, that the high steady state is stable, the intermediate steady state is unstable, and the zero steady state is stable. Thus, there are two possible outcomes: Recovery or full system failure. When n is much greater than $2m$, or greater than the threshold m_{cr}, there are no intersections in figure 5A, but the steady state at zero is still stable, implying that the only possible outcome is full failure. In this final case, the high cognate bias is fatal. Indeed, the steady-state analysis confirms the results indicated by the dynamic simulations. The key determinant of survival is the cognate bias, or the relative values of n and m. In particular, the key measure is $n - 2m$, or the "critical bias." Values of $n - 2m \leq 0$ are safe, whereas values of $n - 2m > 0$ are potentially fatal. Furthermore, figure 5B shows that even when $n < m_{cr}$ and the system therefore stabilizes at a nonzero steady state amino acid concentration, the normalized growth rate $\mu/\mu_M = [x_3/(x_3 + k_{23})]^m$ varies with n, and the lower the bias, the higher the growth rate.

Statistical Evidence of Bias across Genomes

A biological design that would protect against full failure during starvation would be a low cognate bias for the critical enzymes of the amino acid biosynthetic pathways. Past work suggests that there is statistical evidence of low cognate bias in the pathways of some organisms, including *B. subtilis*, *E. coli*, and *S. enterica* (serovar Typhimurium) (Alves and Savageau 2005). However, that analysis used a different model and a different measure of bias based on relative percentage of amino acid composition. Our model presented here indicates that the key measure of bias should instead be based on relative number of cognate amino acids. Furthermore, there is now a substantially larger body of genomic data to mine.
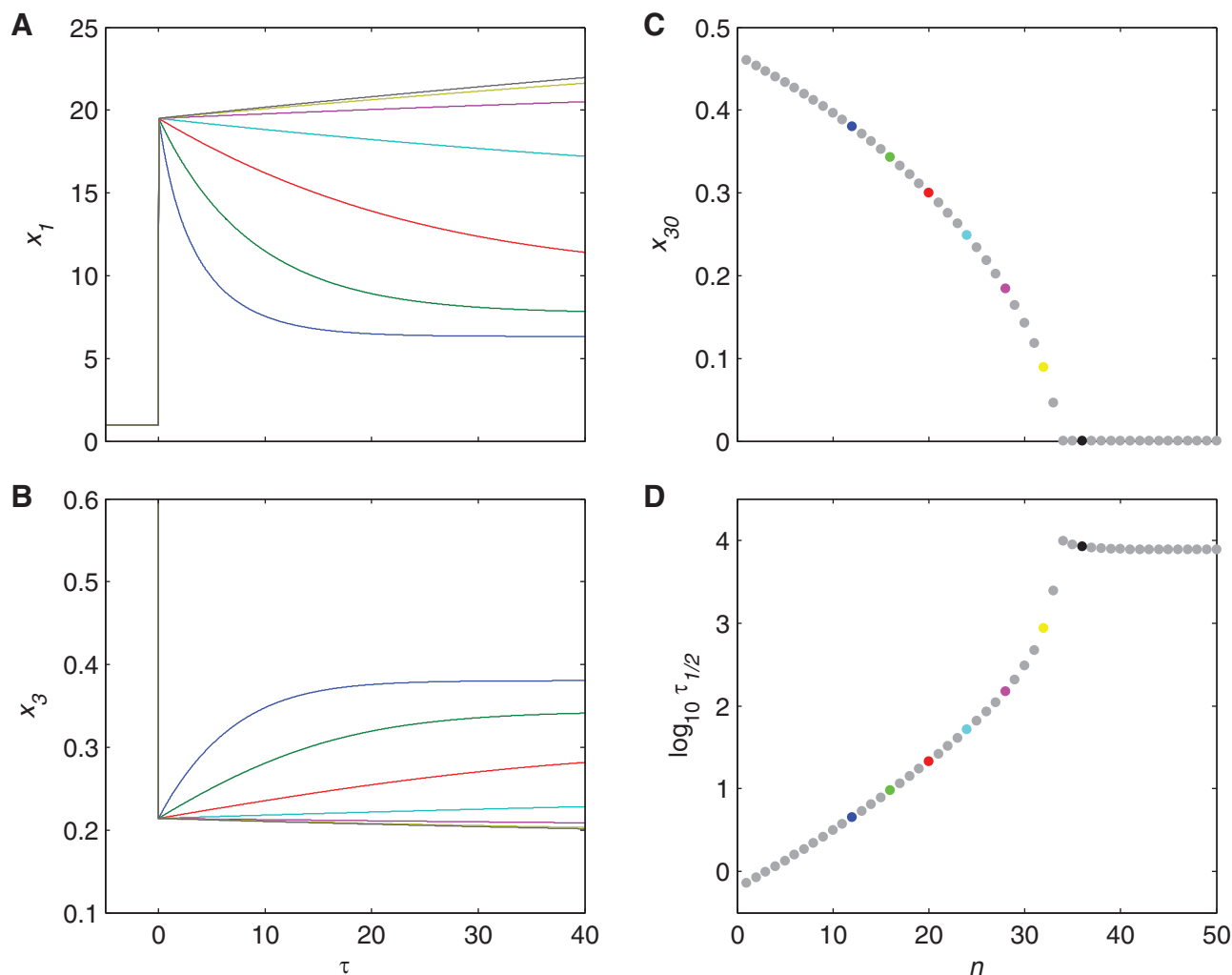
FIG. 4. Dynamics of cognate amino acid concentration during starvation. (A) Normalized concentration of mRNA for the critical enzyme, x_1, and (B) normalized concentration of free cognate amino acid, x_3, during complete and instantaneous starvation. The time scale is normalized to the rate of dilution in the growing cell, or $\tau = t\mu_M$. (C) Final concentration of free cognate amino acid, x_{30}, when the simulation continues to steady state, and (D) the speed of the response after the initial rapid drop, or the time $\tau_{1/2}$ required to cover half the distance to the final steady-state concentration. The number of cognate amino acids in the critical enzyme of the biosynthetic pathway varies from 1 to 50; specific values are $n = 12$ (blue), 16 (green), 20 (red), 24 (light blue), 28 (pink), 32 (yellow), and 36 (black). The average number of cognate amino acids in the proteins of the proteome, m, is 16. Values of $n < m$ represent a low cognate bias in the biosynthetic pathway, whereas values of $n > m$ represent a high cognate bias.

To search for evidence of cognate bias in the genomes of sequenced prokaryotes, we utilized the MetaCyc pathway database (Caspi et al. 2011) and the UniProt protein database (UniProt Consortium 2014). In our model, n represents the number of cognate amino acids in some critical enzyme of the amino acid biosynthetic pathway. However, the identification of a critical enzyme is problematic. It could be a rate-limiting enzyme—and yet different enzymes may be rate limiting under different conditions. The selection could be based on other factors, including activity, half-life, normal concentration, or regulative capacity, each of which is relatively uncharacterized when compared with the wealth of sequence data. Furthermore, amino acid biosynthetic pathways are intertwined, which complicates the identification of a critical enzyme for a specific biosynthetic pathway. Nevertheless, it can be argued that the selective pressure for cognate amino acid bias applies to all of the regulated enzymes of the pathway—after all, the entire pathway must be upregulated in response to starvation—and the last enzyme in each pathway can be uniquely and easily identified. Using MetaCyc, we compiled a list of the last enzymes in all of the known pathways leading to one of the 20 fundamental amino acids, shown in supplementary table S1, Supplementary Material online. From UniProt, we downloaded the complete proteomes of 1,816 completely sequenced prokaryotes. Within each proteome, we searched for the biosynthetic enzymes identified in supplementary table S1, Supplementary Material online. If an enzyme was found, we used the number of cognate amino acids in the enzyme to represent n in our model. If more than one enzyme was found for the production of a particular amino acid, we assumed that the organism has multiple alternative pathways and the pathway with the lowest
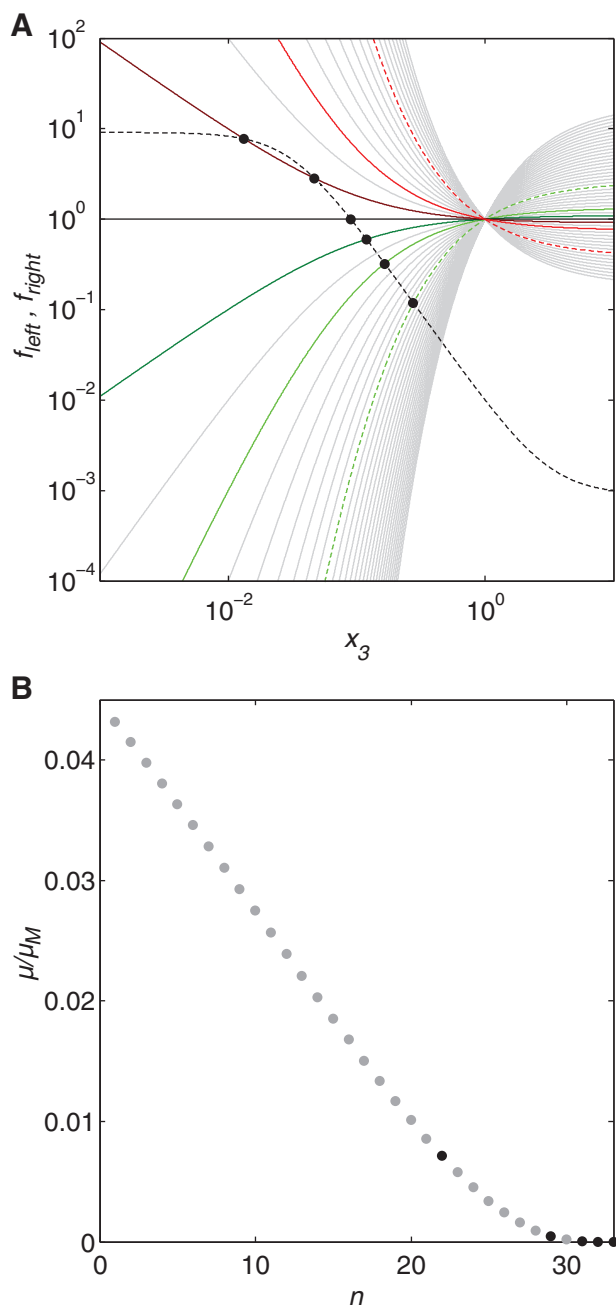
number of cognate amino acids is the most resilient; therefore the lowest of the numbers was used to represent n. Finally, we counted the number of cognate amino acids in each protein of the proteome and used the average number to represent m. The cognate bias, as we have shown, is $n - m$, and the critical bias is $n - 2m$. The resulting data, used in subsequent analyses, are included in supplementary table S2, Supplementary Material online.

Figure 6 displays histograms of the cognate biases measured for each of the 20 amino acids over all 1,816 proteomes. As expected, histograms of the critical biases are similar in shape and shown in supplementary figure S2, Supplementary Material online. There are obvious cases of extreme cognate bias in figure 6: Tryptophan, in almost every case, has a low bias, whereas Arginine, in almost every case, has a high bias. To statistically analyze the significance of the biases, we performed a sign test (see Materials and Methods), a nonparametric test that does not assume a particular population distribution and measures the probability that the values are drawn from a population with a median value of zero, or no bias. The results of the test are listed in table 2. Low P values indicate that the population is biased, and the sample median indicates whether it is biased high or low. The results indicate that there is significant statistical evidence of low cognate bias ($P < 0.001$) in the biosynthetic pathways of six amino acids: Asparagine, Tryptophan, Proline, Leucine, Serine, and Cysteine. All but one of the pathways—Glutamate—show statistically significant evidence of low, or safe, critical bias. Organisms that have a close phylogenetic relationship might be expected to have similar biases and might not be considered independent samples. However, we obtained essentially the same results for the UniProt reference proteomes, which have been selected specifically to provide a wide phylogenetic distribution (UniProt Consortium 2014). Moreover, when clustering on the basis of cognate bias, even organisms that have a close phylogenetic relationship split into different clusters, as shown in the following section.

Clustering of Cognate Bias Compared with Taxonomy

Without any prior knowledge, each protein of the proteome would naively be expected to contain equal amounts, or 5%, of each amino acid. However, it is widely know that this is not the case, and figure 7A depicts our calculation of the amino acid composition biases for each of the completely sequenced prokaryotes found in the UniProt database. The composition bias is measured as the average difference between the number of an amino acid in each protein of the proteome and the expected number, or 5% of the length of the protein. A positive number, or high composition bias, indicates a higher than expected number of a given amino acid in the proteins of the proteome; a low bias indicates a lower than expected number. Figure 7A shows that Cysteine, Tryptophan, Histidine, and Methionine are consistently underrepresented in the proteomes, whereas Leucine is consistently overrepresented. The remaining amino acids vary between a high and low composition bias, depending on the organism. The vertical ordering is sorted by the similarity of the organisms

FIG. 5. Steady-state cognate amino acid concentration and growth rate after starvation. (A) After starvation ($x_7 = 0$), the normalized, steady-state value of free amino acid concentration x_3 is found at the intersection of f_{left}, the value of the left side of equation (10) (dotted black) and f_{right}, the value of the right side of equation (10) for $n = 1$–50 (gray), specifically $n = 22$ (dashed green), 29 (green), 31 (dark green), 32 (black), 33 (dark red), 35 (red), and 42 (dashed red). The average number of cognate amino acids in the proteins of the proteome, m, is 16. Cases where $n < 2m$ (dashed green, green, and dark green) have a single stable, positive steady state, and are considered safe. Cases where $2m < n < m_c$ (dark red) have multiple potential outcomes. Cases where $n > m_c$ (red and dashed red) have only one steady state at zero (not evident on the logarithmic axis), and are therefore fatal. (B) Normalized growth rate $\mu/\mu_M = [x_3/(x_3 + k_{23})]^m$ given the steady-state amino acid concentrations found for $n = 1$–33 (gray), specifically for $n = 22$, 29, 31, 32, and the coincident doublet at 33 (black).
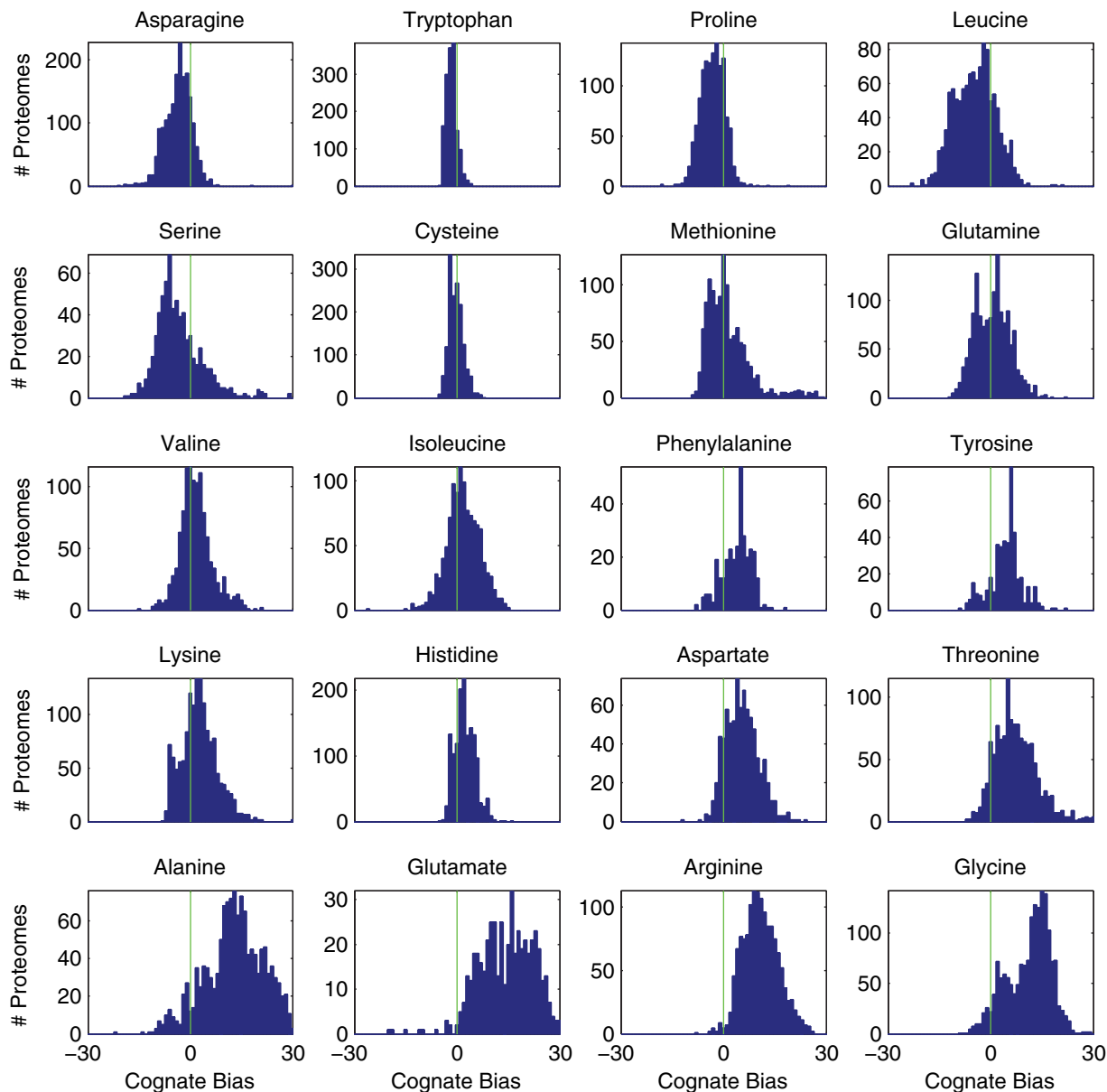
Fig. 6. Histograms of cognate bias measured in complete proteomes. Each histogram describes the distribution of cognate biases across the proteomes of fully sequenced prokaryotes in the UniProt database. Here, the cognate bias is the difference between the number of cognate amino acids in the final enzyme of the biosynthetic pathway n (or the minimum n if there are multiple pathways) and the average number of cognate amino acids in each protein of the putative proteome m. A low bias is less than zero ($n - m < 0$), whereas a high bias is greater than zero ($n - m > 0$). With the exception of Alanine and Glutamate, only 19 points fall outside the depicted range; the full ranges for Alanine and Glutamate are shown in supplementary figure S1, Supplementary Material online.

across all 20 amino acids, and reveals large and small clusters with similar bias profiles. The adjacent bar indicates the taxonomic phylum of each organism, and in several cases the bias clusters roughly correspond to phyla, especially in the highly represented cases of Proteobacteria, Fimicutes, and Actinobacteria.

On the other hand, figure 7B depicts the measured cognate biases for each of the complete proteomes. A positive number, or high cognate bias, indicates a higher than expected number of a given amino acid in the final enzyme of the biosynthetic pathway; a low bias indicates a lower than expected number. Missing values indicate that none of the

known enzymes, and presumably none of the known pathways, was found in the proteome. Note that the cognate bias, by definition, is measured with respect to the composition bias, and so it is remarkable, for example, that Tryptophan, which is already underrepresented in the proteome, tends to have an even lower number in the biosynthetic pathway. The vertical ordering of the cognate bias profiles, based on similarity, reveals fewer obvious clusters than the ordering based on composition bias, although a small cluster in the center includes E. coli and other organisms that can synthesize all 20 amino acids, and a larger cluster at the bottom includes proteomes with a very high Glutamate cognate bias. Likewise,

Table 2. Statistical Measures of Bias.

Amino Acid	Number of Organisms	Median Cognate Bias	P Value	Median Critical Bias	P Value
Asparagine	1,790	−3.1	1e-166	−15.7	0
Tryptophan	1,526	−1.7	6e-161	−5.4	0
Proline	1,288	−3.0	3e-112	−16.6	5e-281
Leucine	1,142	−4.4	2e-80	−36.3	2e-249
Serine	694	−4.5	6e-40	−23.0	1e-142
Cysteine	1,509	−0.5	8e-12	−3.3	8e-246
Methionine	1,229	0.1	0.2	−7.9	3e-141
Glutamine	1,443	0.7	0.0005	−10.3	1e-273
Valine	1,142	1.3	1e-17	−20.7	3e-250
Isoleucine	1,142	1.6	2e-23	−18.0	7e-246
Phenylalanine	315	4.6	3e-29	−7.4	9e-68
Tyrosine	433	5.1	3e-46	−3.7	1e-55
Lysine	1,470	2.3	8e-52	−13.4	1e-307
Histidine	1,409	2.1	6e-91	−4.2	5e-209
Aspartate	799	5.2	5e-99	−11.7	2e-165
Threonine	1,253	6.6	9e-181	−9.7	7e-188
Alanine	1,427	13.8	1e-216	−12.7	7e-204
Glutamate	1,046	45.0	7e-220	25.4	3e-23
Arginine	1,438	10.5	1e-295	−7.2	3e-166
Glycine	1,727	12.6	3e-303	−10.1	4e-264

NOTE.—For each amino acid, the table lists the number of organisms in which a biosynthetic pathway enzyme was found, the median cognate bias and critical bias of the population, and their P values based on a sign test (see Materials and Methods). Low P values ($P < 0.001$) indicate a statistically significant high or low (positive or negative) bias.

clustering based on cognate bias does not appear to correlate as strongly with the taxonomic phyla as does the clustering based on composition bias, especially in the highly represented cases of Proteobacteria, Fimicutes, and Actinobacteria. A similar result is apparent when the proteomes and pathways are clustered by critical bias, as shown in supplementary figure S3, Supplementary Material online. (Raw data for fig. 7A and B, and supplementary fig. S3, can be found in supplementary tables S3–S5, respectively.) Indeed, the larger clusters apparent in figure 7A are broken up and scattered in figure 7B and supplementary figure S3, suggesting that any similarity in cognate bias may depend less on taxonomy and more on the environment.

Discussion

Part of the cell's stringent response to starvation is the upregulation of amino acid biosynthetic pathways, and yet the stringent response might not be enough to protect the cell from the shock of rapid starvation. Cognate bias—relatively fewer cognate amino acids in the corresponding pathway synthesizing the amino acid—could avoid a potential catch-22, in which the emergency response requires the very amino acid that has disappeared from the environment. Past work (Alves and Savageau 2005) suggested that there is a low cognate bias in the pathways of some amino acids, that the bias tends to be lower for the key enzymes in the pathway, and that the profile of bias differs between *E. coli*, *S. enterica* (serovar Typhimurium), and *B. subtillis*—organisms from

different environmental niches. However, the analysis used a different model and a different measure of bias based on relative percentage of amino acid composition, whereas our more detailed model indicates that the key measure of bias should be based on the relative number of cognate amino acids. To illustrate the difference, consider three proteins of 100, 200, and 400 amino acids, each containing two cognate amino acids. The first consists of 2% cognate amino acid, the second consists of 1%, and the third 0.5%. Based on relative percentages, the first protein has the highest bias, whereas the third has the lowest bias; but based on relative numbers of cognate amino acids, their bias is the same. The impact of the different definitions can be significant.

Dynamic simulations and a steady-state analysis of our model show that important aspects of the system depend on the cognate bias, the number of cognate amino acids in the corresponding biosynthetic pathway (n) relative to the number in the expressed proteome (m), or n-m. The lower the cognate bias, the faster the system responds to starvation and the higher the recovered concentration of the free amino acid pool will be, potentially creating a selective evolutionary pressure to lower the cognate bias. Furthermore, our results confirm that a key determinant of the cell fate is the cognate bias. The crucial measure, based on cognate bias, is the critical bias, or $n - 2m$. A low critical bias is always safe, whereas a very high critical bias is fatal. An ambiguous critical bias—slightly high, but below some threshold—can lead to either recovery or failure, depending on the initial conditions and the dynamics of the system. The results suggest immediate predictions that can be experimentally tested. Although single amino acid starvation experiments have been performed in bacteria over the years (Venetianer 1969; Stephens et al. 1975; Shand et al. 1989; Parish 2003; Ohashi et al. 2008), they have not comprehensively tested all of the amino acids. We propose that samples taken from an unstressed culture, grown on rich medium, in steady-state exponential growth, could be used to inoculate a series of cultures grown on 20 different chemically defined media, where each medium has an excess of all amino acids except one. We predict new cultures that upregulate pathways with low critical bias will recover, whereas those that upregulate pathways with high critical bias may fail, or at least experience a long lag in recovery. Furthermore, the model is general enough to make predictions for a variety of organisms. Measures of critical bias should predict relative growth rates and differential recovery times between two organisms starving for the same amino acid or the same organism starving for different amino acids.

A key aspect of our analysis is the changing environment. In a broader context, three classes of amino acids likely correspond to different environmental effects on the proteome: For amino acids that are never required by the organism, the biosynthetic pathway can be dispensed with entirely. For amino acids that are always required by the organism, regulation can be dispensed with, and the amino acids synthesized constitutively without regard for cognate bias. For all other environments, regulation would be advantageous, and a compensating cognate bias would likely exist for amino acids that experience the most frequent and extreme fluctuations in the
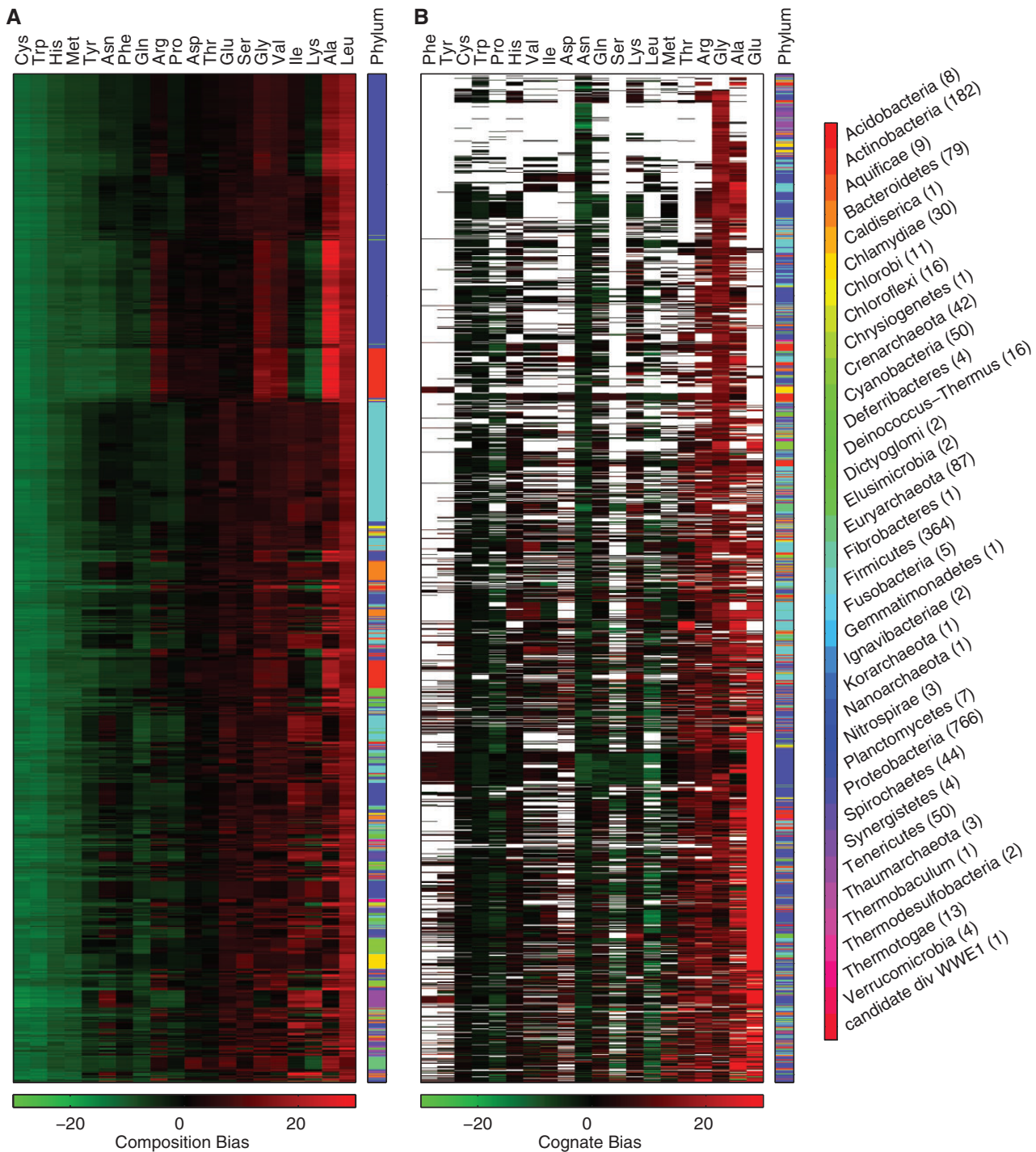
FIG. 7. Hierarchical clustering of composition bias and cognate bias profiles, based on bias similarity in the complete proteomes of the UniProt database. (A) The clustering is based on the composition bias: The average difference between the number of an amino acid in each protein and the expected number, or 5% of the protein length. High composition bias (red) indicates that the amino acid is relatively overrepresented in the proteins of the proteome; low bias (green) indicates that the amino acid is relatively underrepresented. Each row describes the biases measured for a sequenced organism, and the vertical ordering is based on the similarity of the bias profile across all 20 amino acids, with rows closer to the top being more similar. Each column represents one of the 20 amino acids, and the horizontal ordering is based on the similarity of the bias values for that given amino acid over all of the organisms, with columns to the left being more similar. The phylum of each row, or organism, is plotted in a distinct color to the right. The number of organisms in each phylum is shown in parenthesis in the legend. (B) The clustering is based on the cognate bias: The difference between the average number of a given amino acid in each protein of the proteome and the number in the final enzyme of the amino acid biosynthetic pathway. A high cognate bias (red) represents a higher number in the pathway; a low bias (green) represents a lower number in the pathway. Missing values (white) indicate that the biosynthetic pathway was not found in the proteome. The vertical and horizontal ordering are based on similarity, as described in (A). The phylum of each row, or organism, is plotted to the right as in (A).

organism's natural environment. In figure 7B, the missing data points may indicate biosynthetic pathways the organisms have eliminated. High bias may represent the second case: Amino acids that are constitutively produced without regard to bias because they are absent in the environment. In this case, the evidence might suggest that other mechanisms lead to the selection for high cognate bias. Our analysis and results shed light on the last case, a dynamic environment where the amino acid supply is in flux.

An analysis of the proteomes for over 1,800 fully sequenced organisms indicates a statistically low cognate bias in six amino acid biosynthetic pathways, and a low critical bias in all but one. The results open up the possibility that an organism's genome may reflect the nutritional stress that the organism naturally encounters. It should be noted that large protein data sets such as UniProt are built in part by homology with known proteins, which could skew the results toward a few well-studied organisms such as *E. coli*. On the other hand, we obtained similar results when analyzing only the reference proteomes in UniProt, which represent a small, diverse set of well-studied model organisms. Furthermore, we have only considered the case of immediate and complete starvation, whereas organisms naturally experience a variety of transitions from abundance to scarcity and vice versa, sometimes in patterns. For example, *E. coli* encounters periods of starvation and abundance while passing through the guts of successive hosts (Savageau 1983). Future work could explore a full range of environmental conditions and patterns of stress. Ideally, transcriptomic, proteomic, and metabolomic studies would be performed over a full range of concentrations for a single amino acid. In fact, genome-wide studies continue to advance our understanding of the stringent response in *E. coli* (DeRisi et al. 1997; Chang et al. 2002; Durfee et al. 2008; Traxler et al. 2008; Jozefczuk et al. 2010) as well as other organisms (Brockmann-Gretza and Kalinowski 2006; Vercruysse et al. 2011; Gaca et al. 2012), but the studies have not focused on individual amino acid starvation.

The stringent response is a drastic and complex measure taken to cope with a changing environment. Likewise determining the natural environment of most organisms is a challenging task. Our model of starvation and the stringent response establishes a link between the genome and the environment. The results suggest at least one design principle—a low cognate bias—that cells may use to cope with nutritional stress. Armed with that knowledge, we may be able to mine our growing genomic storehouse for clues to an organism's environmental niche.

Materials and Methods

All analyses were performed with custom scripts and the built-in functions of MATLAB R2009a (7.8). The dynamic simulations utilized the built-in MATLAB solver for stiff ODEs: ode15s. The steady-state solutions were found by sampling equation (10), and then verified and polished using the built-in MATLAB function fzero. The linearization of the system and the evaluation of the local stability of the steady states were accomplished with the aid of the MATLAB Symbolic Math Toolbox.

Information on amino acid biosynthetic pathways and their final enzymes was curated manually from the MetaCyc pathway database (Caspi et al. 2011). Protein data were retrieved, parsed, and analyzed using custom MATLAB scripts from the UniProt protein database (UniProt Consortium 2014). The resulting data, used in the subsequent analyses, can be found in supplementary tables S1–S5, Supplementary Material online. The statistical tests for significant cognate bias and critical bias utilized the sign test, a nonparametric test that does not make assumptions about the population distribution, and measures the probability that values were drawn from a population with a median value of zero, or in this case, no bias. The sign test used here was one-sided, in order to measure whether the population deviated in a particular direction, either high or low. The sign tests were performed through the built-in MATLAB function signtest. The clustering analysis was based on nearest Euclidean distance and performed through the built-in MATLAB function linkage.

Supplementary Material

Supplementary figures S1–S3 and tables S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

References

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2002. Molecular biology of the cell. New York: Garland Science. p. 335–365.

Alves R, Savageau MA. 2005. Evidence of selection for low cognate amino acid bias in amino acid biosynthetic enzymes. *Mol Microbiol.* 56:1017–1034.

Blundell M, Kennell D. 1974. Evidence for endonucleolytic attack in decay of *lac* messenger RNA in *Escherichia coli*. J Mol Biol. 83: 143–161.

Brackley CA, Broomhead DS, Romano MC, Thiel M. 2012. A max-plus model of ribosome dynamics during mRNA translation. *J Theor Biol.* 303:128–140.

Brackley CA, Romano MC, Thiel M. 2011. The dynamics of supply and demand in mRNA translation. *PLoS Comput Biol.* 7:e1002203.

Bremer H, Dennis PP. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt FC, Curtiss R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE, editors. *Escherichia coli* and *Salmonella*: cellular and molecular biology. Washington (DC): ASM Press. p. 1553–1569.

Brockmann-Gretza O, Kalinowski J. 2006. Global gene expression during stringent response in *Corynebacterium glutamicum* in presence and absence of the rel gene encoding (p)ppGpp synthase. *BMC Genomics* 7:1–15.

Cashel M. 1969. The control of ribonucleic acid synthesis in *Escherichia coli*: IV. Relevance of unusual phosphorylated compounds from amino acid-starved stringent strains. *J Biol Chem.* 244:3133–3141.

Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, et al. 2011. The MetaCyc database of metabolic pathways and enzymes

and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 40:D742–D753.

Chang D-E, Smalley DJ, Conway T. 2002. Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Mol Microbiol.* 45:289–306.

Chatterji D, Ojha AK. 2001. Revisiting the stringent response, ppGpp and starvation signaling. *Curr Opin Microbiol.* 4:160–165.

Chen W, Shao Y, Chen F. 2013. Evolution of complete proteomes: guanine-cytosine pressure, phylogeny and environmental influences blend the proteomic architecture. *BMC Evol Biol.* 13:1–15.

Curtis TP, Sloan WT. 2005. Exploring microbial diversity—a vast below. *Science* 309:1331–1333.

Dalebroux ZD, Swanson MS. 2012. ppGpp: magic beyond RNA polymerase. *Nat Rev Microbiol.* 10:203–212.

DeRisi JL, Iyer VR, Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680–686.

Draper DE. 1996. Translational initiation. In: Neidhardt FC, Curtiss R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE, editors. *Escherichia coli* and *Salmonella*: cellular and molecular biology. Washington (DC): ASM Press. p. 902–908.

Durfee T, Hansen A-M, Zhi H, Blattner FR, Jin DJ. 2008. Transcription profiling of the stringent response in *Escherichia coli*. *J Bacteriol.* 190: 1084–1096.

Elf J, Ehrenberg M. 2005. Near-critical behavior of aminoacyl-tRNA pools in *E. coli* at rate-limiting supply of amino acids. *Biophys J.* 88:132–146.

Fluitt A, Pienaar E, Viljoen H. 2007. Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput Biol Chem.* 31:335–346.

Friesen JD, Fiil NP, Parker JM, Haseltine WA. 1974. A new relaxed mutant of *Escherichia coli* with an altered 50S ribosomal subunit. *Proc Natl Acad Sci U S A.* 71:3465–3469.

Gaca AO, Abranches J, Kajfasz JK, Lemos JA. 2012. Global transcriptional analysis of the stringent response in *Enterococcus faecalis*. *Microbiology* 158:1994–2004.

Gilchrist MA, Wagner A. 2006. A model of protein translation including codon bias, nonsense errors, and ribosome recycling. *J Theor Biol.* 239:417–434.

Gromadski KB, Rodnina MV. 2004. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol Cell.* 13:191–200.

Guet CC, Bruneaux L, Min TL, Siegal-Gaskins D, Figueroa I, Emonet T, Cluzel P. 2008. Minimally invasive determination of mRNA concentration in single living bacteria. *Nucleic Acids Res.* 36:e73.

Harris BZ, Kaiser D, Singer M. 1998. The guanosine nucleotide (p)ppGpp initiates development and A-factor production in *Myxococcus xanthus*. *Genes Dev.* 12:1022–1035.

Hlavacek WS, Savageau MA. 1996. Rules for coupled expression of regulator and effector genes in inducible circuits. *J Mol Biol.* 255: 121–139.

Hlavacek WS, Savageau MA. 1997. Completely uncoupled and perfectly coupled gene expression in repressible systems. *J Mol Biol.* 266: 538–558.

Jørgensen F, Kurland CG. 1990. Processivity errors of gene expression in *Escherichia coli*. *J Mol Biol.* 215:511–521.

Jozefczuk S, Klie S, Catchpole G, Szymanski J, Cuadros-Inostroza A, Steinhauser D, Selbig J, Willmitzer L. 2010. Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol Syst Biol.* 6:364.

Keiler KC. 2008. Biology of trans-translation. *Annu Rev Microbiol.* 62: 133–151.

Leive L, Kollin V. 1967. Synthesis, utilization and degradation of lactose operon mRNA in *Escherichia coli*. *J Mol Biol.* 24:247–259.

Magnusson LU, Farewell A, Nyström T. 2005. ppGpp: a global regulator in *Escherichia coli*. *Trends Microbiol.* 13:236–242.

Meiler A, Klinger C, Kaufmann M. 2012. ANCAC: amino acid, nucleotide, and codon analysis of COGs—a tool for sequence bias analysis in microbial orthologs. *BMC Bioinformatics* 13:1–8.

Moore SD, Sauer RT. 2007. The tmRNA system for translational surveillance and ribosome rescue. *Annu Rev Biochem.* 76:101–124.

Moura A, Savageau MA, Alves R. 2013. Relative amino acid composition signatures of organisms and environments. *PLoS One* 8:e77319.

Neidhardt FC, Curtiss Iii R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE, editors. 1996. *Escherichia coli* and *Salmonella*: cellular and molecular biology. Washington (DC): ASM Press. p. 391–611.

Ohashi Y, Hirayama A, Ishikawa T, Nakamura S, Shimizu K, Ueno Y, Tomita M, Soga T. 2008. Depiction of metabolome changes in histidine-starved *Escherichia coli* by CE-TOFMS. *Mol BioSyst.* 4:135–147.

Parish T. 2003. Starvation survival response of *Mycobacterium tuberculosis*. *J Bacteriol.* 185:6702–6706.

Pedersen S. 1984. *Escherichia coli* ribosomes translate in vivo with variable rate. *EMBO J.* 3:2895–2898.

Perlstein EO, de Bivort BL, Kunes S, Schreiber SL. 2007. Evolutionarily conserved optimization of amino acid biosynthesis. *J Mol Evol.* 65: 186–196.

Potrykus K, Cashel M. 2008. (p)ppGpp: still magical? *Annu Rev Microbiol.* 62:35–51.

Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 33:D501–D504.

Roche ED, Sauer RT. 1999. SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. *EMBO J.* 18:4579–4589.

Rudd KE. 2000. EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res.* 28:60–64.

Savageau MA. 1983. *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. *Am Nat.* 122:732–744.

Shah P, Gilchrist MA. 2010. Effect of correlated tRNA abundances on translation errors and evolution of codon usage bias. *PLoS Genet.* 6: e1001128.

Shand RF, Blum PH, Mueller RD, Riggs DL, Artz SW. 1989. Correlation between histidine operon expression and guanosine 5′-diphosphate-3′-diphosphate levels during amino acid downshift in stringent and relaxed strains of *Salmonella typhimurium*. *J Bacteriol.* 171: 737–743.

Sørensen MA, Kurland CG, Pedersen S. 1989. Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol.* 207:365–377.

Stent GS, Brenner S. 1961. A genetic locus for the regulation of ribonucleic acid synthesis. *Proc Natl Acad Sci U S A.* 47:2005–2014.

Stephens JC, Artz SW, Ames BN. 1975. Guanosine 5′-diphosphate 3′-diphosphate (ppGpp): positive effector for histidine operon transcription and general signal for amino-acid deficiency. *Proc Natl Acad Sci U S A.* 72:4389–4393.

Sundararaj S, Guo A, Habibi-Nazhad B, Rouani M, Stothard P, Ellison M, Wishart DS. 2004. The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate *in silico* modeling of *Escherichia coli*. *Nucleic Acids Res.* 32: D293–D295.

Tatusova TA, Karsch-Mizrachi I, Ostell JA. 1999. Complete genomes in WWW Entrez: data representation and analysis. *Bioinformatics* 15: 536–543.

Traxler MF, Summers SM, Nguyen H-T, Zacharia VM, Hightower GA, Smith JT, Conway T. 2008. The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol Microbiol.* 68:1128–1148.

UniProt Consortium. 2014. Activities at the universal protein resource (UniProt). *Nucleic Acids Res.* 42:D191–D198.

van der Biezen EA, Sun J, Coleman MJ, Bibb MJ, Jones JDG. 2000. *Arabidopsis* RelA/SpoT homologs implicate (p)ppGpp in plant signaling. *Proc Natl Acad Sci U S A.* 97:3747–3752.

Varenne S, Buc J, Lloubes R, Lazdunski C. 1984. Translation is a non-uniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol.* 180:549–576.

Venetianer P. 1969. Level of messenger RNA transcribed from the histidine operon in repressed, derepressed and histidine-starved *Salmonella typhimurium*. *J Mol Biol.* 45:375–384.

Vercruysse M, Fauvart M, Jans A, Beullens S, Braeken K, Cloots L, Engelen K, Marchal K, Michiels J. 2011. Stress response regulators identified through genome-wide transcriptome analysis of the

(p)ppGpp-dependent response in *Rhizobium etli*. *Genome Biol.* 12: 1–19.

Wall ME, Hlavacek WS, Savageau MA. 2003. Design principles for regulator gene expression in a repressible gene circuit. *J Mol Biol.* 332: 861–876.

Wall ME, Hlavacek WS, Savageau MA. 2004. Design of gene circuits: lessons from bacteria. *Nat Rev Genet.* 5:34–42.

Ward DM, Ferris MJ, Nold SC, Bateson MM. 1998. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol Mol Biol Rev.* 62:1353–1370.

Wendrich TM, Blaha G, Wilson DN, Marahiel MA, Nierhaus KH. 2002. Dissection of the mechanism for the stringent factor RelA. *Mol Cell.* 10:779–788.

Xiu Z-L, Chang Z-Y, Zeng A-P. 2002. Nonlinear dynamics of regulation of bacterial *trp* operon: model analysis of integrated effects of repression, feedback inhibition, and attenuation. *Biotechnol Prog.* 18: 686–693.

Xu J. 2006. Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol.* 15:1713–1731.

Yanofsky C. 1988. Transcription attenuation. *J Biol Chem.* 263:609–612.