Expanded Genetic Codes Create New Mutational Routes to Rifampicin Resistance in *Escherichia coli*

Michael J. Hammerling,*[,1] Jimmy Gollihar,[1] Catherine Mortensen,[1] Razan N. Alnahhas,[‡,1] Andrew D. Ellington,[1] and Jeffrey E. Barrick*[,1]

[1]Center for Systems and Synthetic Biology, Center for Computational Biology and Bioinformatics, Institute for Cellular and Molecular Biology, Department of Molecular BioSciences, The University of Texas at Austin

[‡]Present address: Department of BioSciences, Rice University, Houston, TX 77005

***Corresponding author:** E-mails: mhammerling@gmail.com; jbarrick@cm.utexas.edu.

Associate editor: Csaba Pal

Abstract

Until recently, evolutionary questions surrounding the nature of the genetic code have been mostly limited to the realm of conjecture, modeling, and simulation due to the difficulty of altering this fundamental property of living organisms. Concerted genome and protein engineering efforts now make it possible to experimentally study the impact of alternative genetic codes on the evolution of biological systems. We explored how *Escherichia coli* strains that incorporate a 21st nonstandard amino acid (nsAA) at the recoded amber (TAG) stop codon evolve resistance to the antibiotic rifampicin. Resistance to rifampicin arises from chromosomal mutations in the β subunit of RNA polymerase (RpoB). We found that a variety of mutations that lead to substitutions of nsAAs in the essential RpoB protein confer robust rifampicin resistance. We interpret these results in a framework in which an expanded code can increase evolvability in two distinct ways: by adding a new letter with unique chemical properties to the protein alphabet and by altering the mutational connectivity of amber-adjacent codons by converting a lethal nonsense mutation into a missense mutation. Finally, we consider the implications of these results for the evolution of alternative genetic codes. In our experiments, reliance on a mutation to a reassigned codon for a vital trait is not required for the long-term maintenance of an expanded genetic code and may even destabilize incorporation of an nsAA, a result that is consistent with the codon capture model of genetic code evolution.

Key words: **evolvability, codon capture, antibiotic resistance, sequence space.**

Article

Introduction

While a growing variety of nonstandard amino acids (nsAAs) can now be ribosomally incorporated into proteins in living cells using the tools of synthetic biology, the utility of expanded genetic codes for producing novel phenotypes and improving organismal fitness remains comparatively unexplored (Li and Liu 2014). Similarly, scenarios for how a newly expanded genetic code could become stabilized by progressive mutations in a genome over evolutionary time are only beginning to be experimentally investigated (Yu et al. 2014; Mandell et al. 2015; Tack et al. 2016). These gaps in knowledge have historically been due to considerable technical challenges in re-engineering a cell's translational machinery to code for a new 21st amino acid without negatively influencing its fitness (Lajoie et al. 2016). This barrier has now been largely surmounted in *Escherichia coli* with the creation of strains that incorporate nsAAs into proteins as efficiently and as accurately as natural amino acids (Mukai et al. 2010; Lajoie et al. 2013; Amiram et al. 2015). These advances have enabled the first studies of the impact of genetic code expansion on the evolutionary potential of model proteins and organisms. For example, mutations that lead to the incorporation of nsAAs have been found that increase the activity of the TEM-1 β-lactamase enzyme toward a novel substrate (Xiao et al. 2015) or that improve the competitive fitness of bacteriophage T7 (Hammerling et al. 2014), beyond what is possible by substituting canonical amino acids at the corresponding protein sites in both cases.

The genetic code defines the chemical building blocks and the network topology of sequence space available to evolve protein function in an organism (Freeland and Hurst 1998; Knight, Freeland, et al. 2001; Ilardo et al. 2015). Genetic code expansion, in which a redundant codon is recoded to a 21st nsAA, can theoretically make an organism more evolvable in two ways. First, the 21st amino acid adds a new dimension to sequence space by creating a new letter in the alphabet used to construct proteins. The unique geometry or chemical functionality of a nsAA side chain may enable fundamentally new properties to be encoded in proteins (Ambrogelly et al. 2007; Liu and Schultz 2010). Second, expanding—or even just reassigning—a genetic code alters which amino acid substitutions are accessible to evolution by single base changes in the DNA sequence encoding a protein. Such changes in connectivity may improve how robust protein sequences are to mutations or create new opportunities for adaptive mutations (Firnberg and Ostermeier 2013). This is particularly true

in the most common case in synthetic biology and nature wherein a stop codon, which would normally disrupt protein function, is reassigned to partially or fully code for a 21st amino acid (Ambrogelly et al. 2007).

Antibiotic resistance is an attractive phenotype to use to explore how alternative genetic codes affect evolvability because the molecular mechanisms of resistance are varied and well characterized. For example, rifampicin (Rif) is a broad-spectrum antibiotic that inhibits DNA-dependent RNA synthesis by binding to the bacterial RNA polymerase holoenzyme (RNAP) and preventing elongation of the nascent RNA strand (Campbell et al. 2001). Mutations at many locations within the *rpoB* gene, encoding the β-subunit of RNAP, can confer Rif resistance (Rif^R) by disrupting the hydrogen-bonding network that stabilizes Rif binding or by sterically occluding its binding site. Furthermore, the molecular effects of new Rif^R mutations can be readily interpreted in the context of a high-resolution crystal structure of *E. coli* RNAP bound to rifampicin (Molodtsov et al. 2013). Since Rif^R is conferred by single-base mutations that are always in the essential *rpoB* gene, we used it as a tractable model for studying the ability of recoded genetic codes containing nsAAs to confer a specific, beneficial phenotype. We further hypothesized that Rif^R might serve as a convenient "lever" to isolate *E. coli* variants that become permanently dependent on an expanded genetic code via a single-step genetic selection.

We examined spontaneous rifampicin resistance mutations in several *E. coli* strains engineered to recode the amber stop codon to different nsAAs. These mutants were screened to identify variants that required the nsAA for Rif^R due to amber codon substitutions in *rpoB*. In addition, we used recombineering of an amber codon sequence library into the *rpoB* gene to explore deeper into this region of sequence space. Several resistant strains that were single- or double-amber *rpoB* mutants were isolated and characterized. Structural modeling was used to predict the impact of these nsAA substitutions on the rifampicin-binding pocket of RpoB. Finally, an evolution experiment was performed to assess the stability of expanded genetic codes enforced via amber mutations in *rpoB*, with the unexpected result that some of these mutations actually destabilized the continued incorporation of the 21st amino acid. Our results have implications for understanding the stability and safety of engineered organisms with expanded genetic codes and the mechanisms by which new genetic codes may evolve in nature. In particular, they provide additional evidence that neutral processes leading to extinction of a codon and associated translational machinery from a genome followed by reassignment to a new amino acid may be sufficient for stably altering the meaning of a codon.

Results

Isolation and Characterization of a Rif^R *rpoB* Amber Mutant

In order to test whether an expanded genetic code can confer new mutational routes to rifampicin resistance, we transformed *E. coli* strain C321.ΔA (Lajoie et al. 2013) with a plasmid bearing the *Methanococcus jannaschii* tyrosyl tRNA/aminoacyl tRNA synthetase (aaRS) pair engineered to incorporate 3-iodotyrosine at the amber (TAG) codon (Sakamoto et al. 2009) (see Materials and Methods). We grew 960 cultures of this strain (MJH90) from small initial populations, such that any Rif^R mutants in different cultures were expected to result from independent mutations. Each culture was spotted onto LB-Rif plates with 3-iodotyrosine. This stringent, one-step selection typically produced 0–20 resistant colonies in each spot. Independent mutants were screened via replica plating to test their ability to grow on Rif plates in the absence of the nsAA. Mutants that were able to grow only if 3-iodotyrosine was present were candidates for having evolved Rif^R due to a mutation to a recoded amber codon in the *rpoB* gene.

Of the hundreds of colonies screened, three mutants were isolated which were unable to grow on Rif plates in the absence of 3-iodotyrosine. Sequencing the *rpoB* gene of these mutants revealed that two of three had mutations that converted the glutamine codon (CAG) for amino acid 513 to an amber codon (TAG), hereafter referred to as Q513X, indicating that it occurred via a C→T transition mutation of the first codon nucleotide. This glutamine is known to be important in the hydrogen-bonding network of rifampicin with the RpoB protein, and several mutations to canonical amino acids at this position are known to confer Rif^R (Campbell et al. 2001). The remaining mutant was found to have a D516G substitution in RpoB, a mutation that is known to confer robust rifampicin resistance. Despite containing no amber codon in the *rpoB* gene, this mutant was dependent on the presence of 3-iodotyrosine for growth, suggesting that it coincidentally contained a mutation to the amber codon in an unrelated but essential gene.

Normally, mutations to TAG in *rpoB* would be lethal to *E. coli* because they would create a stop codon in the middle of this essential gene. The Rif^R amber mutants demonstrate that the expanded genetic code increases evolvability in this system by providing another viable point mutation that confers Rif resistance (fig. 1*A* and *B*). Since substitutions of several canonical amino acids at position 513 of RpoB were already known to confer rifampicin resistance, we were interested in determining whether there were quantitative differences in the growth of strains incorporating different amino acids, including other nsAAs, at this position.

The 3-iodotyrosine aaRS synthetase is polyspecific: it is able to incorporate a variety of other 3-substituted tyrosine analogs (fig. 1*C*) with high efficiency and fidelity, while also excluding any of the canonical amino acids (Tack et al. 2016). We found that the *rpoB* Q513X mutant (strain MJH103) is viable in the presence of any one of 3-iodotyrosine, 3-chlorotyrosine, 3-bromotyrosine, or 3-nitrotyrosine; but not in media lacking one of these nsAAs. All of these nsAAs were also able to confer complete Rif^R. Moreover, the growth rate of the Q513X mutant in the presence of Rif was statistically indistinguishable for all four of the nsAAs tested (one-way ANOVA, $P = 0.458$) (fig. 1*D*).

To compare the growth rates of strains incorporating nsAAs at this position with strains incorporating similar

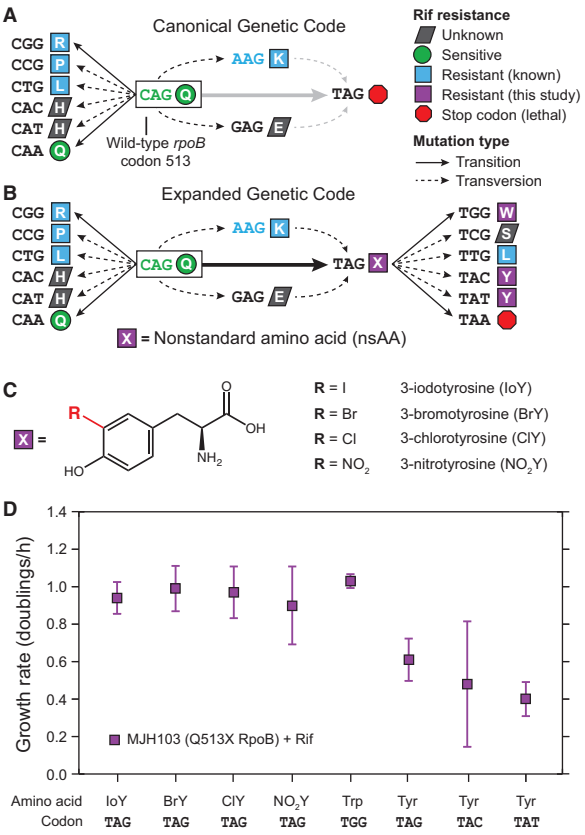

Fig. 1. Mutational network of rifampicin resistance at codon 513 of *rpoB*. (*A*) In the canonical genetic code, the rifampicin-sensitive wild-type glutamine codon CAG (boxed) has seven possible point-mutational routes to rifampicin resistance, with the CAG → TAG and the CAG → CAA mutations being inaccessible due to lethality and rifampicin sensitivity, respectively. Of the seven other possible single-base mutants, only four are reported to confer rifampicin resistance, and only one is a transition mutation. Since transitions occur at a higher rate than transversions, this single resistance mutation dominates the Rif^R fitness landscape at this amino acid position. (*B*) Expanding the genetic code to incorporate a 21st nsAA at the TAG codon alters the network topology surrounding the CAG codon in several ways. Most importantly, it provides an additional transition mutation, which confers resistance in the cases of the nsAAs studied here. In addition, it creates new mutational routes to the codons directly adjacent to the TAG codon, several of which have been shown to be resistant in this study. These alterations to network topology directly and indirectly increase evolvability. (*C*) We studied the rifampicin resistance phenotypes of cells incorporating 3-iodotyrosine and these other tyrosine derivatives incorporated in place of a canonical amino acid at residue 513 of RpoB due to a mutation to the amber codon at this position. (*D*) The growth rates of Rif^R mutants incorporating nsAAs are comparable to or better than that of mutants incorporating a standard amino acid (tyrosine) at this position. Values are the average of five biological replicates with error bars that are 95% confidence intervals. All nsAAs and tryptophan substituted at Q513 result in the same growth rate, while the incorporation of tyrosine results in a lower growth rate.

canonical amino acids, tryptophan (TGG codon) and tyrosine (TAC or TAT codon) mutants were isolated by plating Q513X mutants on media lacking any nsAA. In terms of mutational pathways, none of these codons are accessible by a single base change from the original glutamine codon (CAG) in *rpoB* (fig. 1*B*), so these amino acid substitutions would be exceptionally rare among spontaneous mutations conferring Rif^R in wild-type *E. coli* due to the organization of the genetic code. Finally, the wild-type *M. jannaschii* tyrosyl aaRS was tested within the Q513X mutant background to assess the growth of the strain when incorporating the standard amino acid tyrosine at the amber (TAG) codon.

Whether measured as a group or broken up into individual codons, the Rif^R tyrosine mutants were less fit in the presence of rifampicin (fig. 1*D*), with a significantly lower growth rate compared with growth of the strains incorporating 3-substituted nsAAs or tryptophan as a group (Welch's two-tailed *t*-test, $P < 0.004$). In contrast, the growth rate of the strain with the bulkier tryptophan incorporated at position 513 was indistinguishable from that of the amber mutants incorporating nsAAs (one-way ANOVA, $P = 0.229$). Thus, the 3-substituted nsAAs are not uniquely able to confer a better Rif^R cellular phenotype by this measure of fitness; their side chains do not provide a benefit that cannot also be achieved by the side chain of the standard amino acid tryptophan. However, as mutating to Trp requires two base substitutions compared with one for the nsAAs, the nsAA mutants still demonstrate how improving the connectivity of the genetic code increased the evolvability of the Rif^R phenotype (Firnberg and Ostermeier 2013).

Molecular Modeling of RNAP Mutants

To fully characterize the evolutionary possibilities at this codon within the context of an expanded genetic code, we next used molecular modeling to predict how substituting each canonical amino acid and the four tested nsAAs at position 513 of RpoB would change the binding energy between rifampicin and RNAP. For the substitution of 3-iodotyrosine for glutamine observed in the original Rif^R mutant, this modeling predicts that the iodine atom attached to the aromatic ring of the side chain at the *ortho* position will project directly into the rifampicin binding pocket and completely abolish binding (fig. 2*A*). Changes in the energy of the RNAP-Rif complex upon making each of these side chain substitutions ($\Delta\Delta G$) ranged from slightly stabilizing for some standard amino acids to an insurmountable steric occlusion of the rifampicin-binding pocket imposed by larger side chain substituents at position 513 (fig. 2*B*). The mutations known to confer rifampicin resistance display the most positive $\Delta\Delta G$ values, corresponding to large energetic barriers to rifampicin binding. Conversely, the susceptible wild-type amino acid and most substitutions that have not been observed as resistance mutations result in favorable, negative $\Delta\Delta G$ values for complex stability. These computational predictions are consistent with the superior growth rates in the presence of Rif of every *ortho*-substituted nsAA and tryptophan over tyrosine at this position, suggesting that the conformations identified in our structural modeling of the RNAP-rifampicin interaction capture relevant quantitative differences in resistance.

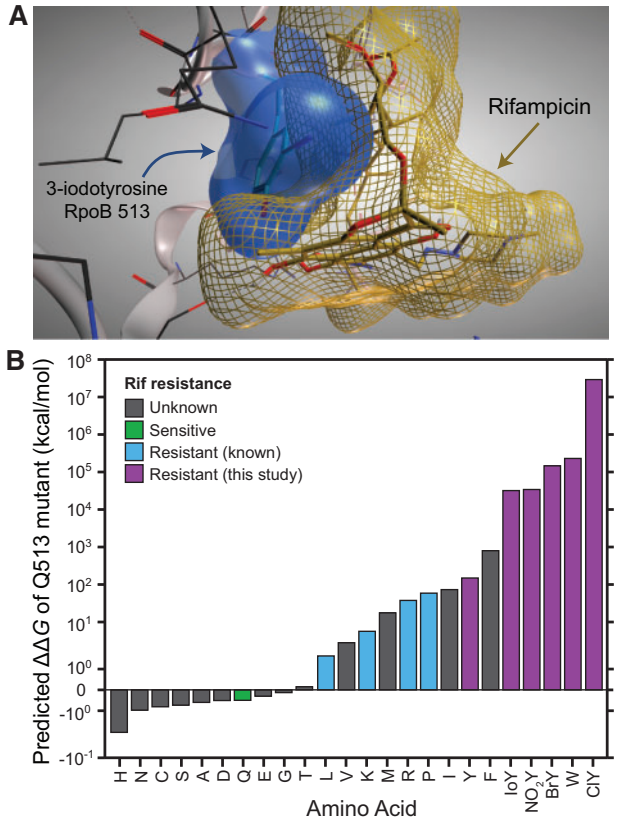

Fig. 2. Structural modeling predicts Rif^R mutations at codon 513 of *rpoB*. (*A*) Molecular modeling of the β-subunit of RNA polymerase (RpoB) with a mutation from glutamine to 3-iodotyrosine at position 513 (blue) in complex with rifampicin (gold) predicts that it blocks antibiotic binding. Space-filling models of the side chain and rifampicin are depicted. A steric clash is shown between the residue and antibiotic, presumably conferring resistance by occluding rifampicin from the complex. (*B*) Predicted $\Delta\Delta G$ values for changes in stability of the protein-rifampicin complex upon substituting different amino acids at position 513 of *rpoB*. The colors of each bar indicate the rifampicin-resistance phenotype of *E. coli* with that amino acid substitution. 3-iodotyrosine (IoY), 3-nitrotyrosine (NO_2Y), 3-bromotyrosine (BrY), and 3-chlorotyrosine (ClY) are modeled in addition to canonical amino acids. At the Q513 position, no amino acids with a predicted negative (stabilizing) or near-zero $\Delta\Delta G$ have been observed to confer rifampicin resistance, and are assumed to be rifampicin-sensitive. All reported resistant mutants have a positive value of $\Delta\Delta G$, with the nsAAs having the highest calculated destabilizations. These predicted $\Delta\Delta G$ values are consistent with the relative rankings of the experimentally measured growth rates of the tyrosine, tryptophan, and nsAA mutants.

Multiple-Amber Library Increases the Variety of Rif^R Mutants

Spontaneous mutations to nsAAs that confer rifampicin resistance are simple to obtain, but the scarcity of Rif^R *rpoB* amber mutants observed by this method (only two of hundreds of screened mutants) emphasizes the impact of the connectivity of the codon landscape on the likelihood that a gene will evolve to utilize an nsAA. If a given codon in a gene is not a single point mutation away from the codon used to incorporate the nsAA, spontaneous mutations to the nsAA will not be observed, even if they are beneficial. Compounding this limitation, biases in the mutational spectrum of an organism, like *E. coli* (Lee et al. 2012), can make even certain codon transitions that are possible via a single-base change so rare compared with others that they are unlikely to be observed in an experiment.

The majority of known Rif^R mutations fall within a 99-bp region of *rpoB* referred to as Region I (Campbell et al. 2001). Only four of the 33 codons of Region I are a single point mutation away from the amber codon, substantially limiting the impact of the expanded code on protein evolvability. In addition, one point mutation to an amber codon is unlikely to result in evolutionarily stable dependence on the nsAA, since it may simply revert by a single mutation to another resistant codon, as demonstrated by our isolation of rifampicin-resistant tyrosine and tryptophan mutants from Q513X mutants. Multiple amber codons in a single open reading frame would be expected to reduce the likelihood that an organism will be able to escape biocontainment, as a single point mutation within the ORF is no longer sufficient to revert the coding sequence to be translatable in the absence of the nsAA (Rovner et al. 2015).

To find rifampicin-resistant variants of *rpoB* containing multiple amber codons, a Region I library was generated using sets of overlapping oligonucleotides that included amber codons at a high frequency (see Materials and Methods) (fig. 3*A*). This library allows the protein to access the nsAA at the 29 codons in Region I that are not a single nucleotide change away from the amber codon, increasing the number of testable codons over 7-fold. We generated *rpoB* sequence variation in the genomes of *E. coli* C321.ΔA strains containing a plasmid with the machinery for incorporating either 3-iodotyrosine, 3-nitrotyrosine (using polyspecificity of the same 3-iodotyrosine aaRS/tRNA pair) (Tack et al. 2016), *p*-azidophenylalanine (Chin et al. 2002), or 5-hydroxytryptophan (Ellefson et al. 2014) (fig. 3*B*) by directly transforming the DNA library into this recombineering-proficient genetic background (Lajoie et al. 2013). After creating this more diverse pool of *rpoB* genes, we selected Rif^R mutants and screened 50 clones derived from each strain background to discover variants that were unable to grow in the absence of their cognate nsAA. From this screen, we identified five double-amber mutants (four unique genotypes) and one single-amber mutant (MJH226–MJH231) (fig. 3*C*). Four of the six new amber codons were more than one point mutation away from the original codons at these positions in *rpoB*.

This screen uncovered Rif^R amber mutations at several locations that are known to be able to mutate to canonical amino acids to confer resistance, including Q513, H526, and R529. Especially interesting from the perspective of evolvability, however, are positions F514, Q517, and T525, which to our knowledge have not been reported in the literature to be able to confer rifampicin resistance as single amino acid substitutions. The T525X amber mutation occurred alone in one clone selected on 5-hydroxytryptophan, so we can conclude that it is causative for rifampicin resistance. For the double mutants, it was unclear from these data alone whether one or both of the observed amber mutations contribute to resistance.

The growth rates of the previously described Rif^R single-amber 513X mutant (MJH103) and four unique double-amber mutants (MJH226–229) of *rpoB* were compared with *E. coli* with wild-type *rpoB* (MJH090), all bearing the polyspecific 3-iodotyrosine synthetase pair (fig. 4A). All of the mutants had a significantly slower growth rate compared with wild-type *rpoB* when rifampicin was not present (Welch's two-tailed *t*-test, $P < 0.05$). The decrease in growth rate of the amber-containing Rif^R mutants compared with wild-type in media lacking Rif is typical of other *E. coli* with Rif^R mutations involving substitutions of standard amino acids (Reynolds 2000; Barrick et al. 2010). This decrease in growth rate is typically associated with altered transcription termination and pausing, as well as with enzyme substrate recognition, whether or not Rif is present (Alifano et al. 2015). Accordingly, none of the Rif^R amber mutants displayed a significant difference in growth rate when cultured with rifampicin compared with when they were cultured without rifampicin (Welch's one-tailed *t*-test, $P > 0.05$).

To gain insight into the possible roles of nsAAs in double-amber mutants, we again modeled the molecular structures of each Rif^R variant and compared the predicted energies of their binding interactions with rifampicin. In each case, destabilization of the antibiotic-bound complex was observed, due to extensive steric clashes between Rif and the nsAA side chains (fig. 4B and C). We were unable to find rotamers of any of the nsAAs that would allow for proper ligand orientation and complex formation, and template-based docking of Rif into allowed conformations of each variant failed to produce productive binding complexes. In addition, each of the amino acid substitutions found in the amber double mutants were modeled separately to deconvolute their respective contributions to Rif^R. Interestingly, this modeling predicts that every one of these amber substitutions would result in Rif^R on its own ($\Delta\Delta G$ for the stability of the RNAP–Rif complex $>$ 1,000 kcal/mol in all cases).

Evolutionary Stability of Expanded Genetic Codes with Amber-Dependent Rif^R

Previous studies have found that *rpoB* mutants are typically able to restore their lost fitness after brief periods of evolution by acquiring compensatory mutations, rather than by reverting the resistance codon to restore Rif sensitivity, even when they are evolved in the absence of rifampicin (Reynolds 2000; Barrick et al. 2010). Consequently, we expected that culturing these mutants in the presence of rifampicin would prevent loss of the expanded genetic code, as the Rif^R phenotype that is dependent on incorporation of a 21st amino acid at the amber codon must be maintained. Thus, we wondered whether we would observe a difference in evolutionary stability of the expanded genetic code in strains with amber mutations conferring rifampicin resistance versus strains with a wild-type *rpoB* but the same expanded genetic code system.

We reasoned that strains with one or two amber codons in *rpoB* would be less likely to experience mutational failure of their expanded genetic code over evolutionary time. To test this hypothesis, we performed a serial transfer evolution

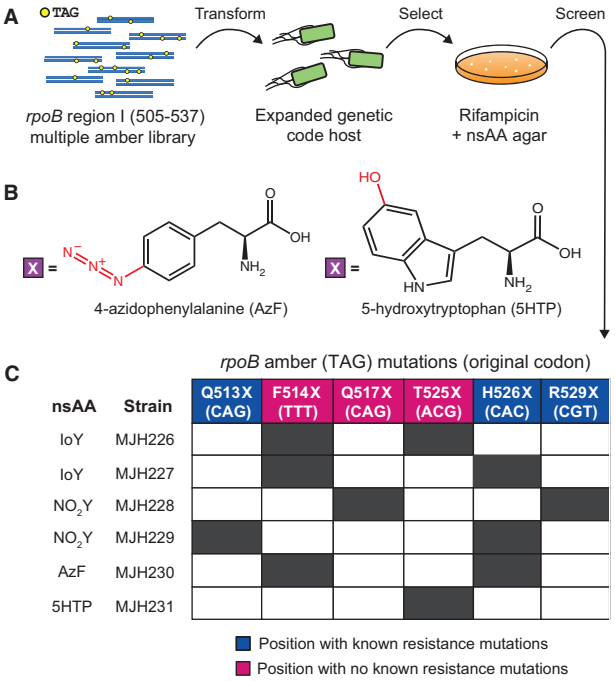

Fig. 3. One-step transformation and selection of rifampicin resistant *E. coli* with multiple amber codons in *rpoB*. (*A*) Strategy for selecting multiple-amber, rifampicin-resistant mutants of *rpoB*. Most rifampicin resistance mutations lie within a 33-codon segment of *rpoB* known as Region I. A Region I mutant library was transformed into C321.ΔA amberless *E. coli* and plated on LB-Rif plates containing the cognate nsAA. Fifty mutants from each transformation were then screened in LB with rifampicin and in the presence or absence of the cognate nsAA to screen for mutants that could not grow in the absence of the nsAA. (*B*) In addition to the 3-substituted tyrosine derivatives used in the first experiment, we used *E. coli* strains with genetic codes expanded with 4-azidophenylalanine and 5-hydroxytryptophan in this experiment. (*C*) Double- and one single-amber mutants obtained in this screen. nsAAs abbreviations are defined in figures 1C and 3B.

experiment in liquid medium on three replicates of each *rpoB* amber-mutant strain bearing the 3-iodotyrosine tRNA/synthetase pair. The *rpoB* genotypes included wild-type grown in the presence of 3-iodotyrosine or 3-nitrotyrosine, Q513X grown in 3-iodotyrosine or 3-nitrotyrosine, the double-mutants MJH226 and MJH227 grown in 3-iodotyrosine, and double-mutants MJH228 and MJH229 grown in 3-nitrotyrosine. Each of these strains was evolved for 100 generations by serial transfer in liquid medium containing the nsAA. Integrity of their expanded genetic codes was assessed every 20 generations by comparing viable cell counts in the presence of antibiotic, but without nsAA, to those when the nsAA was provided (as described in the Materials and Methods).

Growth of a cell on antibiotic without nsAA present indicates that mutations have converted all amber codons in the *rpoB* gene to other codons or that mutations have altered the translational machinery such that it incorporates one of the 20 standard amino acids at the amber codon. The latter scenario of expanded genetic

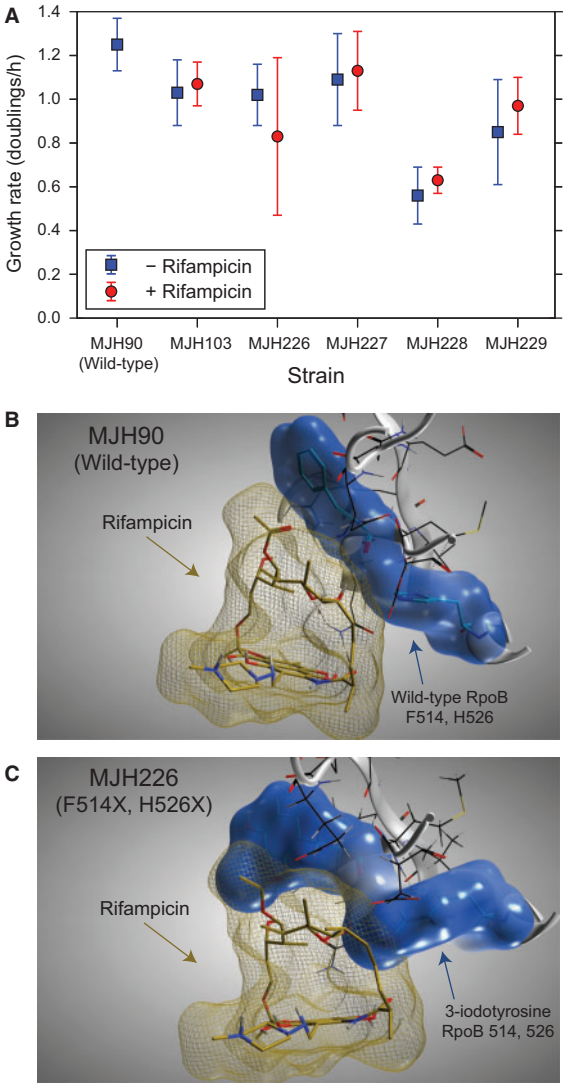

Fig. 4. Double-amber mutants confer robust rifampicin resistance. (*A*) Doubling times of Rif^R *rpoB* mutants incorporating nsAAs in the absence (blue) or presence (red) of Rif. Error bars represent 95% confidence intervals from five replicate measurements. The presence of Rif in the growth medium does not affect the growth rate of any Rif^R mutants. (*B*) Wild-type *rpoB* in complex with Rif (gold). Residues 514 and 526 are highlighted in blue. The orientation of the side chains allows for a productive binding complex and inhibition of the enzyme. (*C*) A double-amber mutant MJH227 (F514X, H526X) variant incorporating 3-iodotyrosine is shown with an overlay of Rif (gold) in the binding pocket. The occlusion of Rif from binding is achieved by steric occlusion involving the *ortho*-substituents of the nsAA. Rotamer exploration and docking failed to produce conformational variants that would allow for Rif–RpoB complex formation.

code failure may occur, for example, when mutations in the *E. coli* chromosome create amber suppressor tRNAs from native tRNAs. Without further genome engineering to remove such tRNAs, there will always be a rate at which these types of mutants arise in a population (Rovner et al. 2015). However, if this baseline does not increase during an experiment, it indicates that the expanded genetic code is evolutionarily stable on this timescale. That is, these types of mutants do not have a fitness advantage such that they are able to dominate and lead to a majority of the population no longer incorporating the nsAA.

Surprisingly, the expanded genetic code remains stable over evolutionary time in the absence of any *rpoB* amber mutations (fig. 5). That is, the frequency of mutant cells that were no longer dependent on the orthogonal translation system did not increase over time in populations of *E. coli* C321.ΔA with the polyspecific 3-iodotyrosine synthetase (MJH90), whether they were propagated with 3-iodotyrosine or 3-nitrotyrosine as the 21st amino acid. The Q513X RpoB single mutant (MJH103) was similarly stable for either of these expanded genetic codes, as were the double-mutants incorporating 3-iodotyrosine (MJH226 and MJH227). All of these Rif^R mutants with nsAA substitutions in RpoB were stable with or without maintaining selection for antibiotic-resistance by adding rifampicin during the entire 100 generations of this experiment.

The stability of the expanded genetic code in the absence of any amber codons in the host genome implies that though it is known that there are many mutations in both the genome and plasmid that result in the evolutionary failure of the system (Normanly and Abelson 1989; Rovner et al. 2015), there is not a sufficient selection pressure to bring these mutations to high frequency in the population, at least on this experimental timescale. Nor is nsAA dependence lost due to mutations that eliminate the amber codons present in *rpoB* in these Rif resistant mutants. In fact, the prevalence of nsAA independent mutants within each population day-to-day did follow the expected pattern, with a lower frequency of mutants as more amber codons were added to the *rpoB* gene (mutant frequency of MJH90 > MJH103 > MJH226/MJH227). As these mutants did not take over these populations in any case, it seems that the fitness burden of the altered genetic code in the context of this host is quite low, such that other beneficial mutations outcompete any genetic code "failure" mutations during the serial transfer experiment.

Interestingly, the 3-nitrotyrosine double-mutant genotypes reduced the stability of the expanded genetic code, resulting in significant loss of code fidelity at the amber codon over the course of the 100-generation experiment in at least one of the MJH228 lines, and in all of the MJH229 lines. We attribute this code instability to the lower fitness of these 3-nitrotyrosine double-mutants compared to the other Rif^R amber mutants (fig. 4A). This large initial fitness defect means that certain mutations that result in loss of the expanded genetic code—by evolving a native tRNA into an amber suppressor that decodes the TAG codon as a normal amino acid, for instance—may simultaneously alleviate the fitness burden associated with all amber mutations to *rpoB* by changing how this codon is read by the ribosome at all of these sites. Especially in the case of the 3-nitrotyrosine double mutants, it is reasonable to suspect that incorporation of this bulky amino acid with a charged ring decoration interferes with the normal function of the RNAP holoenzyme complex, resulting in globally altered transcription and a reduced cellular growth rate. Since this lower initial fitness results in more selection pressure to replace the nsAA with a different amino acid, mutations that "break" the genetic code would rise to high

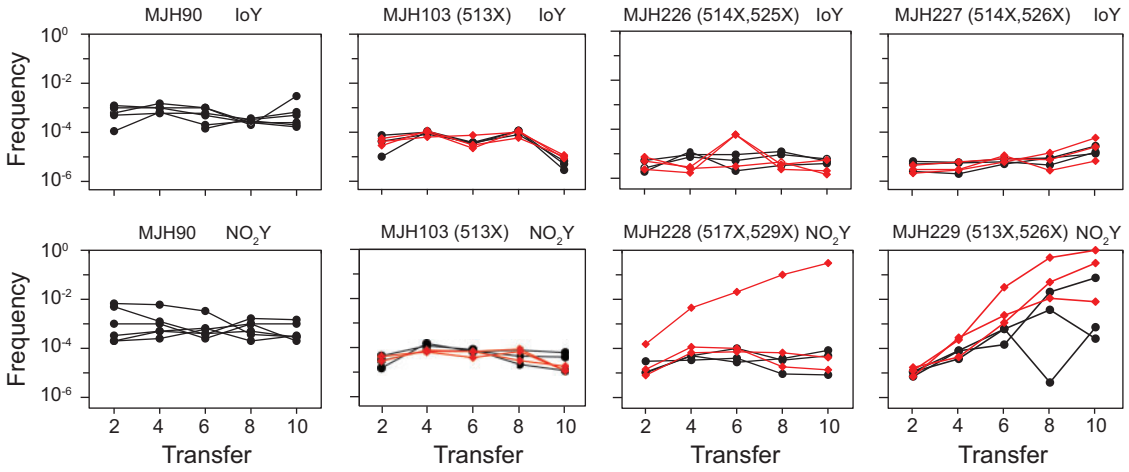

Fig. 5. Frequency of expanded genetic code loss mutants in evolved populations. Expanded genetic code strains with wild-type *rpoB* (MJH090), *rpoB* Q513X (MJH103), or *rpoB* double-amber mutants (MJH226–MJH229) were grown in the presence of 3-iodotyrosine (top row) or 3-nitrotyrosine (bottom row), and in the absence (black lines) or presence (red lines) of Rif. Most mutants maintained stable incorporation of the 21st amino acid over evolutionary time whether in the presence or absence of Rif, resulting in a low and approximately constant frequency of cells that exhibited nsAA-independent viability when grown in the absence of antibiotic across 10 transfers (as described in the Materials and Methods). Only 3-nitrotyrosine double mutants displayed an obvious decrease in expanded genetic code stability over time.

frequency much more rapidly in these population than the same mutations could in cells with wild-type *rpoB*. In contrast to the 3-nitrotyrosine double mutants, 3-iodotyrosine may be more readily accepted into protein structures, as has been demonstrated previously (Ohtake et al. 2015). Thus, surprisingly, including amber codons in essential genes to engineer dependence can actually decrease the stability of the new genetic code in certain cases.

Discussion

An expanded genetic code can increase the evolvability of proteins: both by adding new amino acid side chains to the protein alphabet and by increasing the connectivity of amino acid sequence space. However, the ability of an evolving system to make use of an expanded genetic code is also constrained by several factors. Most sense codons (53/61 = 87%) are not a single point mutation away from the amber codon used to expand the genetic code. As a result, the chemical properties of the amino acids encoded by these few adjacent codons are important in determining the likelihood that a mutation to the nsAA will be accessible to evolution (Firnberg and Ostermeier 2013). Introducing a 21st amino acid with atoms and functional groups that are radically different from the chemistry coded by adjacent codons will make it unlikely that point mutations to this codon will be tolerated in native protein structures, while introducing a new side chain with properties that are similar to those of the 20 standard amino acids may add mostly neutral connections in sequence space. A compromise between these two extremes may be optimal for an organism in the long run, as evidenced by the balance of mutational robustness and chemical diversity in the canonical genetic code (Freeland and Hurst 1998; Freeland et al. 2002). Alternatively, since adaptive evolution (particularly in the short-term) acts on the most beneficial extreme tail of phenotypic variance, it may be better to introduce the rare possibility of novel, highly beneficial mutations afforded by adding a new chemical group into a genetic code at the expense of somewhat reduced robustness against deleterious mutations. This latter line of reasoning is especially relevant when recoding a stop codon, as most mutations to the stop codon will already completely disrupt protein function.

The only known natural genetic code expansions, with the 21st amino acids selenocysteine and pyrrolysine, introduce novel side chains that are used to create unique phenotypes in proteins that cannot be achieved by any amino acid in the canonical code (Ambrogelly et al. 2007; Rother and Krzycki 2010). The rifampicin resistance mutations documented here do not fall into this category, as other amino acids are able to confer equivalent antibiotic resistance phenotypes. Presumably, if these strains were evolved over longer periods of time with an intact expanded genetic code, additional nsAA substitutions would accumulate in many more essential genes. Eventually, failure mutations that resulted in the loss of the expanded code would no longer be beneficial or even neutral to organismal fitness, but this process may require more than single nsAA substitutions in one protein's structure. For example, strong addiction to a new code can be enforced by introducing a nsAA residue and then evolving compensatory changes in residues that interact with the novel side chain (Tack et al. 2016).

The observation that an expanded genetic code is stable in the highly engineered *E. coli* strain C321.ΔA lends experimental support to the plausibility of one hypothesis for the natural evolution of altered genetic codes. The "codon capture" theory posits that natural alterations to the genetic code occur in a two-step process (Osawa and Jukes 1989; Santos et al. 2004) . First, a codon becomes extinct from a genome. Some combination of gene loss, which reduces the number of every codon in a genome, and a mutational bias, which favors

extreme AT or GC base content, leads to the disappearance of all examples of a particular codon via genetic drift. Second, the meaning of the now extinct codon is reassigned by mutations to the anticodon loop of existing tRNAs or horizontal gene transfer of decoding machinery, which allows the organism or organelle to recognize this codon as a different amino acid.

The *E. coli* genome was not greatly reduced in the process of generating of C321.ΔA (Lajoie et al. 2013). Yet, the removal of all amber codons from the genome and deletion of protein release factor RF-1 enables codon meaning to be altered without competition with the native function of the amber codon as a translation terminator. In this way, the engineering procedure was similar to the mechanism put forth in the codon capture theory, in that the extinction of the codon from the genome occurred via a process that was approximately neutral with respect to organism fitness. This is in contrast to the "ambiguous intermediate" theory, an alternative hypothesis which proposes that genetic code expansion often proceeds through an evolutionary intermediate in which the recoded codon has multiple meanings (Schultz and Yarus 1996). That is, the codon leads to the incorporation of either the old or the new amino acid with some probability each time that it is read by a ribosome. One challenge to this hypothesis is the contention that this ambiguous, intermediate code would likely negatively impact the fitness of an organism because the resultant "statistical" proteome would contain aberrant sequences (Bacher et al. 2004; Santos et al. 2004; Yu et al. 2014). Since codon capture, in contrast, occurs by a process that is neutral and stepwise, it does not suffer from this drawback. Indeed, a comparative analysis of genetic code reassignment in mitochondria suggests that disappearance of a codon from this small genome is often an important component of codon reassignment, but that it may not strictly be required or necessarily predict the likelihood that a particular codon will be successfully reassigned (Knight, Landweber, et al. 2001). This study provides experimental support for the notion that the removal of all codons and associated decoding machinery from a genome can be sufficient for the evolutionary stability of a nascent codon reassignment. This occurs even in the absence of substantial genome reduction or a phenotype that requires selection on a unique chemical property of a 21st amino acid.

In summary, this is a case study of how a panel of expanded genetic codes can confer a well-understood phenotype in a model organism, and the relationship between that phenotype and the stability of the expanded genetic code itself. It offers several lessons for understanding the principles governing the evolvability of the genetic code. First, while an expanded genetic code can demonstrably improve evolvability of a selectable phenotype, for a mutation to a repurposed codon to stabilize the expanded code against mutational failure it must provide a fitness benefit that is strictly dependent on the unique chemistry or geometry of the new amino acid. Even the presence of multiple mutations to the altered codon may not be enough to reinforce the new code if standard amino acids may be substituted to achieve a similar phenotype. Second, removal of a codon and its associated decoding machinery from a genome appears to be a sufficient condition for the evolutionary stability of an alteration to that codon's meaning. This experimental result supports the plausibility of the codon capture theory. Finally, while central housekeeping proteins (such as the component of RNA polymerase studied here) present attractive targets for substituting nsAAs to achieve dependence on expanded genetic codes—due to their irreplaceable functions in the cell—the fitness costs associated with compromising their functions in any way may result in destabilizing rather than reinforcing a new genetic code.

Materials and Methods

Media and Strains

All bacterial growth was conducted at 30°C in LB (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl). Unless otherwise stated, media supplements were included at the following concentrations: 100 μg/mL spectinomycin, 100 μg/mL carbenicillin, 100 μg/mL rifampicin, 20 μg/mL gentamicin, and 500 μM nsAA. The tRNA/synthetase expression plasmids were generated using Gibson isothermal assembly (Gibson 2011) to combine *M. jannaschii* tRNA-synthetase pairs, with a spectinomycin-resistant plasmid backbone and the p15A origin of replication from pACYC184. In this family of plasmids, the aaRS and the Nap3 tRNA (Guo et al. 2009) are expressed from copies of the native *E. coli* TyrRS and *lpp* promoters, respectively, and these co-linear transcripts are followed by one copy of the *rrnB* terminator. These plasmids were transformed into *E. coli* strain, C321.ΔA, for all experiments in this study (Addgene: #48998) (Lajoie et al. 2013).

Selection for Rifampicin-Resistant Mutants

Frozen stocks of C321.ΔA bearing the plasmids with each tRNA/synthetase pair were revived in LB with spectinomycin (to select for the tRNA/aaRS plasmid), carbenicillin (TEM-1 is present in the C321.ΔA genome), and the appropriate nsAA and grown overnight. The next day, each culture was diluted into 200 mL of the same medium to create a solution with ~100 cells/mL. Then, 400 μL of this mixture was aliquoted into each well of ten 96-well microplates and incubated overnight. The next day, 2.5 μL of culture from each of the 960 total wells was spotted onto LB agar plates containing all previous supplements as well as rifampicin. These plates were then grown 48 hours, and Rif^R colonies from each spot were patched onto an LB agar plate with all previous antibiotics, and with or without the relevant nsAA to observe independent, Rif^R colonies that were only viable in the presence of the appropriate nsAA. These patch plates were assessed after 36 hours of growth. Patches that grew only in the presence of the nsAA were inoculated in liquid media with the appropriate antibiotics and with or without nsAA. Region I of the *rpoB* gene of mutants that grew only when supplemented with the nsAA was PCR amplified and Sanger sequenced to identify resistance mutations.

Obtaining *rpoB* Codon 513 Rescue Mutants

The original *rpoB* Q513X mutant was revived overnight in 5mL LB broth with appropriate antibiotics and 3-iodotyrosine. This culture was then diluted 1:10,000 into sterile saline and again 1:1,000 into 50 mL of the same medium. Then, 400 μL of this mixture was aliquoted into each well of a 96-well culture plate and incubated overnight to obtain independent rescue mutants. The next day, 2.5 μL of culture from each well was spot-plated onto LB-Rif plates with no nsAA. These plates were then grown for 48 h and 93 independent resistant colonies were picked and grown in 5 mL LB–Rif cultures. Of these cultures, 47 were screened for rescue mutations in the *rpoB* gene by PCR amplification and Sanger sequencing.

Growth Rate Assessment

The log-phase growth rate of mutants was assessed in a TECAN M200 plate reader at 30°C with periodic shaking and measurements every 5 min. Growth rate was assessed for each sample within the OD600 window between 0.02 and 0.04 by computing an exponential doubling time using a freely available web tool (http://www.doubling-time.com).

Molecular Modeling

The wild-type structure of *E. coli* RNA polymerase (PDB: 4KMU chains C and D) (Molodtsov et al. 2013) was prepared for mutational analyses using the Molecular Operating Environment (MOE.09.2014) software package from Chemical Computing Group. The structure was inspected for anomalies and protonated/charged with the Protonate3D subroutine (310K, pH 7.4, 0.1 M salt) (Labute 2009). The protonated structure was then lightly tethered to reduce significant deviation from the empirically determined coordinates and minimized using the Amber10:EHT force field with R-field treatment of electrostatics to an RMS gradient of 0.1 kcal/mol/Å. Next, we created rotamer libraries for each noncanonical amino acid using the low-mode molecular dynamics method (LowModeMD) (Labute 2010). Library positions of representative variants from the selection were mutated and repacked within the local environment. The final model for each variant was further refined using explicit solvent and minimizing the solvent enclosed structure to an RMS gradient of 0.001 kcal/mol/Å. Models were evaluated by calculating phi–psi angles and superimposed against the reference structure to ensure that the native fold was maintained (Clark and Labute 2007). Measurements of stability, potential energy, and contact interactions on a fixed-ligand structure for each variant were scored within MOE using an apo structure with each substitution and no Rif present as a reference.

Generation of Region 1 Mutant Library

Using four overlapping segments consisting of 37 oligonucleotides (33 amber mutant and four wild-type oligos), a 99-bp library of dsDNA consisting of a mix of single, double, triple, and quadruple amber-mutants was built. Unique oligonucleotides for each of the 33 amber codon substitutions were combined at a total frequency of 60% in a primer extension reaction with Phusion DNA polymerase. The remaining 40% consisted wild-type oligonucleotides to generate a distribution of higher-order mutant combinations. The sequences of all 37 oligonucleotides are provided in supplementary table S1, Supplementary Material online. Assuming nonbiased annealing and extension, this reaction would produce a variety of sequences, with single, double, triple, and quadruple amber mutants of *rpoB*, constituting 12.96%, 15.36%, 34.56%, and 34.56% of the total library, respectively, and wild-type fragments making up the remaining 2.56%. This library was purified and transformed directly into recombineering-proficient C321.ΔA electrocompetent cells as previously described (Lajoie et al. 2013). Rif^R mutants were selected by plating dilutions of the resulting transformation on LB–Rif plates containing the nsAA. These mutants were screened for growth in liquid LB with the cognate nsAA and no growth without the nsAA to verify that they contained amber substitutions before sequencing Region I of *rpoB* and further characterization.

Evolutionary Stability Experiment

For each independently evolved line, a clone of the parent strain was isolated on LB agar medium to serve as the ancestor, and inoculated into 5 mL of LB medium with plasmid antibiotics, nsAA, and rifampicin when appropriate. A total of 10 serial transfers were performed over 10 days with a 1:100 dilution of saturated culture into fresh media, and frozen stocks were saved every two transfers. The frequency of failure mutants in a population (cells that have lost the expanded genetic code or mutated TAG codons in *rpoB*) was determined by plating a dilution of the mixed population on LB–Rif agar lacking nsAA, and comparing the number of cells observed to the total viable cell count determined by plating a dilution on LB with the nsAA. The failure rate of expanded genetic code hosts with wild-type *rpoB* was determined by transforming with a dual-selectable kanamycin/chloramphenicol plasmid containing an amber codon at a neutral position (Y8X mutation) in the chloramphenicol acetyltransferase gene. In this case, failure of the expanded genetic code host was monitored as the frequency of observed transformants growing on LB agar with kanamycin and chloramphenicol but lacking nsAA compared with the total observed transformants on kanamycin alone. When we observed zero failure mutants at the counted dilution, the 95% confidence bound on the maximum frequency of mutants in the population was calculated from the Poisson distribution and used in place of the missing data.

Supplementary Material

Supplementary table S1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This study was supported by the National Science Foundation BEACON Center for the Study of Evolution in Action (Grant number DBI-0939454); the Defense Advanced Research Projects Agency (Grant number HR0011-15-C0095); and the Welch Foundation (Grant number F-1780). We thank Jared Ellefson for the 5-hydroxytryptophan incorporation

plasmid and members of the Barrick and Ellington labs for helpful suggestions and discussion.

References

Alifano P, Palumbo C, Pasanisi D, Talà A. 2015. Rifampicin-resistance, *rpoB* polymorphism and RNA polymerase genetic engineering. *J Biotechnol*. 202:60–77.

Ambrogelly A, Palioura S, Söll D. 2007. Natural expansion of the genetic code. *Nat Chem Biol*. 3:29–35.

Amiram M, Haimovich AD, Fan C, Wang Y-S, Ntai I, Moonan DW, Ma NJ, Rovner AJ, Hoon S, Kelleher NL, et al. 2015. Evolution of translation components in recoded organisms enables multi-site non-standard amino acid incorporation in proteins at high yield and purity. *Nat Biotechnol*. 33:1272–1278.

Bacher JM, Hughes RA, Tze-Fei Wong J, Ellington AD. 2004. Evolving new genetic codes. *Trends Ecol Evol*. 19:69–75.

Barrick JE, Kauth MR, Strelioff CC, Lenski RE. 2010. *Escherichia coli rpoB* mutants have increased evolvability in proportion to their fitness defects. *Mol Biol Evol*. 27:1338–1347.

Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, Darst SA. 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 104:901–912.

Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. 2002. Addition of *p*-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J Am Chem Soc*. 124:9026–9027.

Clark AM, Labute P. 2007. 2D depiction of protein–ligand complexes. *J Chem Inf Model* 47:1933–1944.

Ellefson JW, Meyer AJ, Hughes RA, Cannon JR, Brodbelt JS, Ellington AD. 2014. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat Biotechnol*. 32:97–101.

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

Freeland SJ, Hurst LD. 1998. The genetic code is one in a million. *J Mol Evol*. 47:238–248.

Freeland SJ, Prog G, Machines E. 2002. The Darwinian genetic code: an adaptation for adapting? Genet. *Program Evol Mach*. 3:113–127.

Gibson DG. 2011. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol*. 498:349–361.

Guo J, Melançon CE, Lee HS, Groff D, Schultz PG. 2009. Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew Chemie - Int Ed*. 48:9148–9151.

Hammerling MJ, Ellefson JW, Boutz DR, Marcotte EM, Ellington AD, Barrick JE. 2014. Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. *Nat Chem Biol*. 10:178–180.

Ilardo M, Meringer M, Freeland S, Rasulev B, Cleaves HJ. 2015. Extraordinarily adaptive properties of the genetically encoded amino acids. *Sci Rep*. 5:9414.

Knight RD, Freeland SJ, Landweber LF. 2001. Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet*. 2:49–58.

Knight RD, Landweber LF, Yarus M. 2001. How mitochondria redefine the code. *J Mol Evol*. 53:299–313.

Labute P. 2009. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* 75:187–205.

Labute P. 2010. LowModeMD—implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. *J Chem Inf Model* 50:792–800.

Lajoie MJ, Söll D, Church GM. 2016. Overcoming challenges in engineering the genetic code. *J Mol Biol*. 428:1004–1021.

Lajoie MJ, Rovner AJ, Goodman DB, Aerni H-R, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, et al. 2013. Genomically recoded organisms expand biological functions. *Science* 342:357–360.

Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A*. 109:E2774–E2783.

Li X, Liu CC. 2014. Biological applications of expanded genetic codes. *Chembiochem* 15:2335–2341.

Liu CC, Schultz PG. 2010. Adding new chemistries to the genetic code. *Annu Rev Biochem*. 79:413–444.

Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, Norville JE, Gregg CJ, Stoddard BL, Church GM. 2015. Biocontainment of genetically modified organisms by synthetic protein design. *Nature* 518:55–60.

Molodtsov V, Nawarathne IN, Scharf NT, Kirchhoff PD, Showalter HDH, Garcia GA, Murakami KS. 2013. X-ray crystal structures of the *Escherichia coli* RNA polymerase in complex with benzoxazinorifamycins. *J Med Chem*. 56:4758–4763.

Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S, Sakamoto K. 2010. Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res*. 38:8188–8195.

Normanly J, Abelson J. 1989. tRNA identity. *Annu Rev Biochem*. 58:1029–1049.

Ohtake K, Yamaguchi A, Mukai T, Kashimura H, Hirano N, Haruki M, Kohashi S, Yamagishi K, Murayama K, Tomabechi Y, et al. 2015. Protein stabilization utilizing a redefined codon. *Sci Rep*. 5:9762.

Osawa S, Jukes TH. 1989. Codon reassignment (codon capture) in evolution. *J Mol Evol*. 28:271–278.

Reynolds MG. 2000. Compensatory evolution in rifampin-resistant *Escherichia coli*. *Genetics* 156:1471–1481.

Rother M, Krzycki JA. 2010. Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea* 2010:297–308.

Rovner AJ, Haimovich AD, Katz SR, Li Z, Grome MW, Gassaway BM, Amiram M, Patel JR, Gallagher RR, Rinehart J, et al. 2015. Recoded organisms engineered to depend on synthetic amino acids. *Nature* 518:89–93.

Sakamoto K, Murayama K, Oki K, Iraha F, Kato-Murayama M, Takahashi M, Ohtake K, Kobayashi T, Kuramitsu S, Shirouzu M, et al. 2009. Genetic encoding of 3-iodo-L-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography. *Structure* 17:335–344.

Santos MAS, Moura G, Massey SE, Tuite MF. 2004. Driving change: the evolution of alternative genetic codes. *Trends Genet*. 20:95–102.

Schultz DW, Yarus M. 1996. On malleability in the genetic code. *J Mol Evol*. 42:597–601.

Tack DS, Ellefson JW, Thyer R, Wang B, Gollihar J, Forster MT, Ellington AD. 2016. Addicting diverse bacteria to a noncanonical amino acid. *Nat Chem Biol*. 12:138–140.

Xiao H, Nasertorabi F, Choi S, Han GW, Reed SA, Stevens RC, Schultz PG. 2015. Exploring the potential impact of an expanded genetic code on protein function. *Proc Natl Acad Sci U S A*. 112:6961–6966.

Yu ACS, Yim AKY, Mat WK, Tong AHY, Lok S, Xue H, Tsui SKW, Wong JTF, Chan TF. 2014. Mutations enabling displacement of tryptophan by 4-fluorotryptophan as a canonical amino acid of the genetic code. *Genome Biol. Evol*. 6:629–641.