

Release Factor One Is Nonessential in *Escherichia coli*

David B. F. Johnson,^{†,⊥} Chong Wang,^{†,⊥} Jianfeng Xu,[†] Matthew D. Schultz,^{‡,§} Robert J. Schmitz,[‡] Joseph R. Ecker,^{‡,||} and Lei Wang^{*,†}

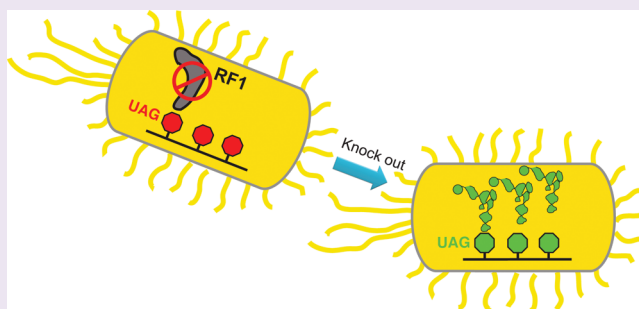
[†]The Jack H. Skirball Center for Chemical Biology and Proteomics and [‡]Genomic Analysis Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, California 92037, United States

[§]Bioinformatics Program, University of California at San Diego, La Jolla, California 92093, United States

^{||}Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, Maryland 20815, United States

Supporting Information

ABSTRACT: Recoding a stop codon to an amino acid may afford orthogonal genetic systems for biosynthesizing new protein and organism properties. Although reassignment of stop codons has been found in extant organisms, a model organism is lacking to investigate the reassignment process and to direct code evolution. Complete reassignment of a stop codon is precluded by release factors (RFs), which recognize stop codons to terminate translation. Here we discovered that RF1 could be unconditionally knocked out from various *Escherichia coli* strains, demonstrating that the reportedly essential RF1 is generally dispensable for the *E. coli* species. The apparent essentiality of RF1 was found to be caused by the inefficiency of a mutant RF2 in terminating all UAA stop codons; a wild type RF2 was sufficient for RF1 knockout. The RF1-knockout strains were autonomous and unambiguously reassigned UAG to encode natural or unnatural amino acids (Uaas) at multiple sites, affording a previously unavailable model for studying code evolution and a unique host for exploiting Uaas to evolve new biological functions.



Although the canonical genetic code is preserved in almost all organisms, small deviations have been discovered, including the reassignment of sense codons from one amino acid to another and the reassignment between stop and sense codons.^{1,2} Stop codons are decoded by class I release factors (RFs).³ Whereas eukaryotes and archaea use a single RF to recognize all three stop codons,^{4,5} bacteria use two: RF1 is specific for UAA/UAG, and RF2 is specific for UAA/UGA.⁶ It is unknown why there are two class I RFs in bacteria while a single class I RF is sufficient for organisms from the other two domains. The process for stop codon reassignment and its potential association with RF evolution are also unclear. Natural code evolution occurs over millions of years. Extant organisms harboring altered genetic codes are at the end-point of the code evolution. There are no records of the initial causes of or responses to such altered genetic codes; further adaptations to and details of eventual fixation are completely unknown. To enable in-depth investigation of code change and any concurrent cellular adaptations in real time, it is necessary to generate a model organism that is able to undergo such evolutionary processes in the laboratory.

Synthetically recoding a genome may afford new properties to the organism through encoding unnatural amino acids (Uaas) and preventing cross-contamination with wild type life forms.⁷ For successful genome recoding, the target codon must be reassigned to the new meaning in high efficiency and

without ambiguity. An attractive route is to reassign the UAG stop codon to a Uaa in bacteria. Orthogonal tRNA/synthetase pairs have been engineered to incorporate Uaas into proteins in response to UAG,^{8–12} yet the presence of RF1 makes the meaning of UAG ambiguous, being a stop signal and a Uaa simultaneously. RF1 competition limits the incorporation of Uaas at a single UAG site with low efficiency; the addition of even a second UAG codon decreases protein yields precipitously.¹³ Although Uaas can be incorporated at more than one site into a protein,^{14,15} such low UAG-encoding efficiency prevents effective use of Uaas at multiple sites to explore novel protein and organismal properties through directed evolution. In addition, the ambiguity of the UAG codon may hinder the eventual fixation of an altered genetic code to an organism, because protein products truncated at the UAG sites can interfere with normal protein functions and have detrimental effects on the host cell, thus preventing advantageous coding from being inherited and selected in directed evolution. To exclude Uaa incorporation at legitimate termination sites specified by UAG, endogenous UAG codons in the *E. coli* genome can be replaced with a synonymous UAA stop codon through genome engineering.¹⁶ However, for

Received: May 9, 2012

Accepted: June 4, 2012

Published: June 5, 2012

complete reassignment of UAG to a sense codon, a necessary and critical step is to knock out RF1 from the *E. coli* genome.

In some eukaryotes such as ciliates and green algae, the reassignment of a stop codon to a sense codon is accompanied by convergent changes in eRF1.² For instance, the eRF1 of *Tetrahymena* restricts its recognition to UGA, and UAA/UAG are reassigned to Gln; the eRF1 of *Euplotes* recognizes only UAA/UAG as stop codons, and UGA is used to encode Cys.^{17,18} In bacteria, *Mycoplasma* species have lost the RF2 gene, and the UGA codon encodes Trp instead.¹⁹ However, *Mycoplasma* are obligatory pathogens with highly reduced genomes. To date, no free-living bacterium has been found lacking either RF1 or RF2.¹⁹ For *E. coli*, RF1 has been reported to be essential,²⁰ and only conditionally lethal knockouts have been described.^{21,22} Recently we managed to knock out RF1 from a special *E. coli* strain that has a reduced genome and a mutated RF2 gene.¹³ Herein we discovered that the dispensability of RF1 is a general property of wild type *E. coli*, arguing against the paradigm that RF1 is essential. We revealed the underlying mechanism for preventing RF1 knockout, and generated autonomous RF1 deletion strains valuable for studying code evolution and for genome recoding.

We attempted to replace the RF1-encoding gene, *prfA*, with the chloramphenicol acetyltransferase gene in a variety of *E. coli* strains (Figure 1) using the established λ red recombinase-based homologous recombination method.²³ *E. coli* K-12 and B are the two progenitors from which most *E. coli* strains are derived.²⁴ Reported attempts to knockout RF1 have used only *E. coli* K-12 strains and were unsuccessful.^{20,25} We initially attempted to delete RF1 in the three common K-12 strains MG1655, DH10 β , and HT115. Knockout of RF1 was assayed by genomic PCR amplifying the *prfA* locus (Figure 1A). Consistent with previous reports, we could not knock out RF1 in any of these K-12 strains (Figure 1B).

We next attempted the RF1 knockout with two other K-12 strains that contain alterations related to translational termination. The BP5 α strain harbors a glutamyl amber suppressor tRNA, making the UAG coding ambiguous for either a stop signal or Gln. This strain would test if stop codon ambiguity is a factor to prime RF1 removal. The MDS42 strain has nearly 700 nonessential genes deleted²⁶ and was used to assess if a reduced genome size and termination load could assist with RF1 removal. However, no RF1 knockouts were generated from these two strains either, suggesting that amber suppression and a minimal genome are insufficient for RF1 removal in K-12 derivatives.

All K-12 strains contain a peculiar A246T mutation in the RF2-encoding gene *prfB*, lowering the release activity of RF2 for UAA 5-fold.^{27,28} The UAA codon accounts for the termination of ~64% of *E. coli* genes.²⁹ We reasoned that the A246T mutation might severely impair the ability of RF2 to recognize all UAA stop codons upon removal of RF1 and thus prevent the RF1 knockout. We therefore tested three common *E. coli* B strains because B strains encode wild type RF2 that contains Ala246.²⁸ Indeed, RF1 knockout was successful with all three B strains tested; REL606, BL21, and BL21(DE3) were used to generate the knockout strains CW1.0, CW2.0, and JX1.0, respectively (Figure 1A). In addition, to determine if the A246T mutation in RF2 prevents RF1 removal in K-12 strains, we reverted the A246T mutation to Ala in the K-12 strain DH10 β to generate the DH10 β f strain. This strain also permitted the direct knockout of RF1 (Figure 1B). These

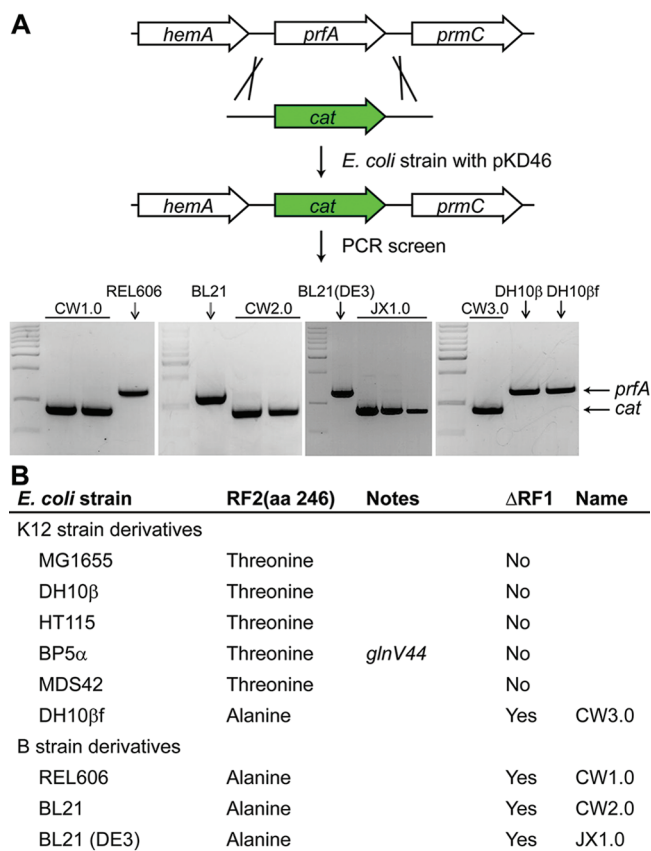


Figure 1. RF1 can be knocked out from *E. coli* strains containing wild type RF2. (A) The RF1-encoding gene *prfA* was replaced with the chloramphenicol acetyltransferase (*cat*) gene through homologous recombination using the phage λ red recombinase expressed with plasmid pKD46.²³ Chloramphenicol-resistant colonies were screened using PCR for *cat* replacement and confirmed with sequencing. The agarose gels show the PCR results for parental and knockout strains as labeled. (B) A list of *E. coli* strains tested for RF1 knockout. K-12 and B strains are the most widely used *E. coli* and are the subjects of classical experiments. Among other genomic differences between K-12 and B strains,²⁴ A246T in RF2 is a peculiar mutation in K-12 strains that reduces the release activity of RF2 for the stop codon UAA. This mutation was reverted to wild type Ala in DH10 β f. The *glnV44* gene encodes a glutamyl amber suppressor tRNA. RF1 was successfully knocked out in three B strains and the DH10 β f strain.

results indicate that RF1 is not essential in *E. coli*, in contrast to the previous conclusions.^{20,25}

Full genomic sequencing was performed on RF1 knockout strains JX1.0, CW1.0, and CW2.0 and compared to their respective parental strains.^{30,31} The RF1 deletion was verified in all cases. For CW1.0 and CW2.0, no other mutations were found throughout the genome. For JX1.0, only seven additional single nucleotide polymorphisms (SNPs) were found (Supporting Table S1). Six of these SNPs are silent mutations in two genes of phage origin. None of the SNPs correspond to known mutations that complement an RF1 deficiency.^{32–35} These results indicate that RF1 was knocked out from the parental *E. coli* strains without incurring compensatory mutations.

What would happen to the UAG site during protein translation in the absence of RF1? We mutated one and three tyrosine codons to TAG in the enhanced green fluorescent protein (EGFP) gene³⁶ to generate the 1-TAG and 3-TAG EGFP reporters, respectively, and tested their

expression in BL21(DE3) and the RF1 knockout strain JX1.0 (Figure 2A). As expected, BL21(DE3) cells expressing either

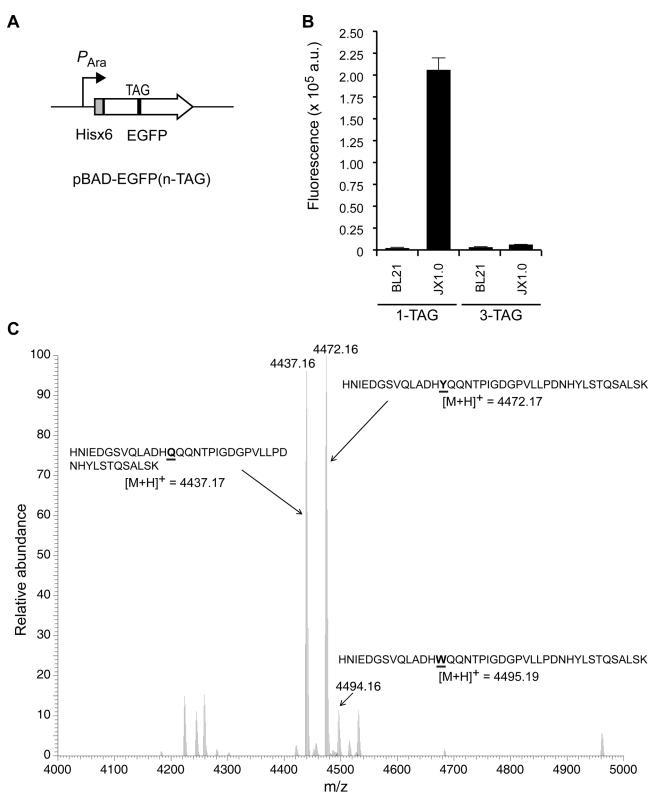


Figure 2. UAG codon in the RF1 knockout strain JX1.0 is suppressed by endogenous tRNAs with mispairing anticodons. (A) An EGFP gene with the TAG codon at permissive sites was driven by the arabinose promoter in the plasmid pBAD-EGFP(n-TAG) to test EGFP expression. (B) In-cell fluorescence intensity measured with fluorometry for BL21(DE3) and JX1.0 cells transformed with pBAD-EGFP(1-TAG) or pBAD-EGFP(3-TAG). Samples were normalized for cell numbers. (C) Fourier transform mass spectrometric analysis of EGFP protein purified from JX1.0 expressing pBAD-EGFP(1-TAG). The spectrum shows the intact precursor ions (charge deconvoluted) for the tryptic fragment of EGFP (HNIEDGVSQQLADHXQNTPIGDGPVLLPDNHYLSTQSALS, X representing the UAG encoded site). Monoisotopic masses measured with high accuracy indicate that Gln, Tyr, and Trp (expected $[M + H]^+ = 4437.17, 4472.17, \text{ and } 4495.19$ Da; measured 4437.16, 4472.16, and 4494.16 Da, respectively) were incorporated at the UAG site. No other amino acids were detected at the UAG site.

the 1- or 3-TAG reporter showed no EGFP fluorescence, indicating that RF1 terminates EGFP translation at UAG normally. Surprisingly, JX1.0 showed a high level of fluorescence from the 1-TAG EGFP mutant and a small but reproducible level of fluorescence from the 3-TAG EGFP mutant (Figure 2B). These data suggest that some endogenous tRNAs can suppress the UAG codon in JX1.0.

To reveal the identity of these tRNAs, we purified EGFP protein expressed with the 1-TAG reporter in JX1.0 and identified the amino acid incorporated at the UAG site using mass spectrometry (Figure 2C). Tyr, Gln, and Trp were found at the UAG site through Fourier transform mass spectrometric analysis of the tryptic fragments of EGFP. To ensure that this is not from a suppressor mutation in the endogenous tRNAs, chromosomal tRNA^{Tyr}, tRNA^{Gln}, and tRNA^{Trp} were resequenced, and all were confirmed to be wild type in JX1.0. The

anticodons of these three tRNAs have only a single base mispairing with UAG, so these tRNAs can weakly misread the UAG codon.³⁷ Misreading of UAG by these tRNAs was not obvious in the presence of RF1 in BL21(DE3) but became marked in JX1.0 after RF1 removal.

The unconditional knockout of RF1 opens up the possibility of reassigning the meaning of the UAG codon completely, so that the UAG codon does not ambiguously encode a stop signal and an amino acid simultaneously. We attempted to recruit UAG for an amino acid in JX1.0, so as to mimic a possible code evolution pathway implicated by RF constriction in stop codon recognition in *Tetrahymena*, *Euplotes*, and other eukaryotes.^{17,18} An orthogonal tRNA/synthetase pair, the tRNA^{Tyr}_{CUA}/LWIRS pair, was introduced into JX1.0. This pair does not crosstalk with endogenous *E. coli* tRNA/synthetase pairs and functionally couples with *E. coli*'s protein translational machinery.^{8,38} The tRNA^{Tyr}_{CUA} decodes the UAG codon specifically through its anticodon CUA, and LWIRS is engineered to charge the tRNA^{Tyr}_{CUA} with the Uaa *p*-acetyl-L-phenylalanine (pActF).³⁹ An EGFP gene containing 1-, 2-, 3-, or 10-TAG codons was co-expressed with tRNA^{Tyr}_{CUA}/LWIRS in a single plasmid pAIO-EGFP(n-TAG) (Figure 3A).

EGFP expression was assayed by Western blotting (Figure 3B) and in-cell fluorescence (Figure 3C). When pActF was not added to the growth media, BL21(DE3) cells expressed a small amount of full-length EGFP with the 1-TAG reporter only, suggesting that the tRNA^{Tyr}_{CUA}/LWIRS pair incorporates a natural amino acid in very low efficiency in the absence of the cognate pActF. No full-length EGFP was detected for the 2-, 3-, or 10-TAG reporters. In contrast, JX1.0 expressed full-length EGFP for 1-, 2-, 3-, and 10-TAG mutants, although the efficiency decreased with the number of UAG codons. We then purified the EGFP protein from JX1.0 expressing pAIO-EGFP(1-TAG) in the absence of pActF (Figure 3D) and analyzed it with mass spectrometry. Consistently, Tyr, Gln, and Trp were found at the UAG site as observed in Figure 2C, confirming misreading by endogenous tRNAs in JX1.0.

When pActF was supplied in the growth medium, BL21(DE3) cells showed efficient expression of the 1-TAG EGFP mutant, but EGFP expression decreased precipitously with the addition of each UAG codon due to the competition from RF1 termination. The use of 3 UAG codons in EGFP virtually abolished protein expression, and no protein could be detected at all in the 10-TAG mutant. In stark contrast, the RF1-deletion strain JX1.0 showed high expression of EGFP in all mutants in the presence of pActF, as indicated by Western (Figure 3B) and in-cell fluorescence (Figure 3C). EGFP proteins were purified with yields of 8.5 (± 0.4), 7.1 (± 0.4), 9.7 (± 0.3), and 1.2 (± 0.1) mg/L for the 1-, 2-, 3-, and 10-TAG EGFP samples, respectively (Figure 3D). There was no decrease in incorporation efficiency when the UAG codon was increased from 1 to 2 or 3, indicating that UAG is changed to a sense codon in JX1.0. The drop off in yield in the 10-TAG EGFP sample is likely because 10 pActF interferes with EGFP folding and/or stability, which is corroborated with the lack of fluorescence from 10-pActF EGFP (Figure 3C).

Although JX1.0 had no RF1 to terminate protein translation at the UAG codon, a very small amount of protein was detected on the Western blot but could not be detected with mass spectrometry that seemed to be truncated at the UAG site (Figure 3B). We think that the decoding of the UAG codon as the Uaa pActF may not be as efficient as the decoding of canonical sense codons as natural amino acids. First, the pActF-

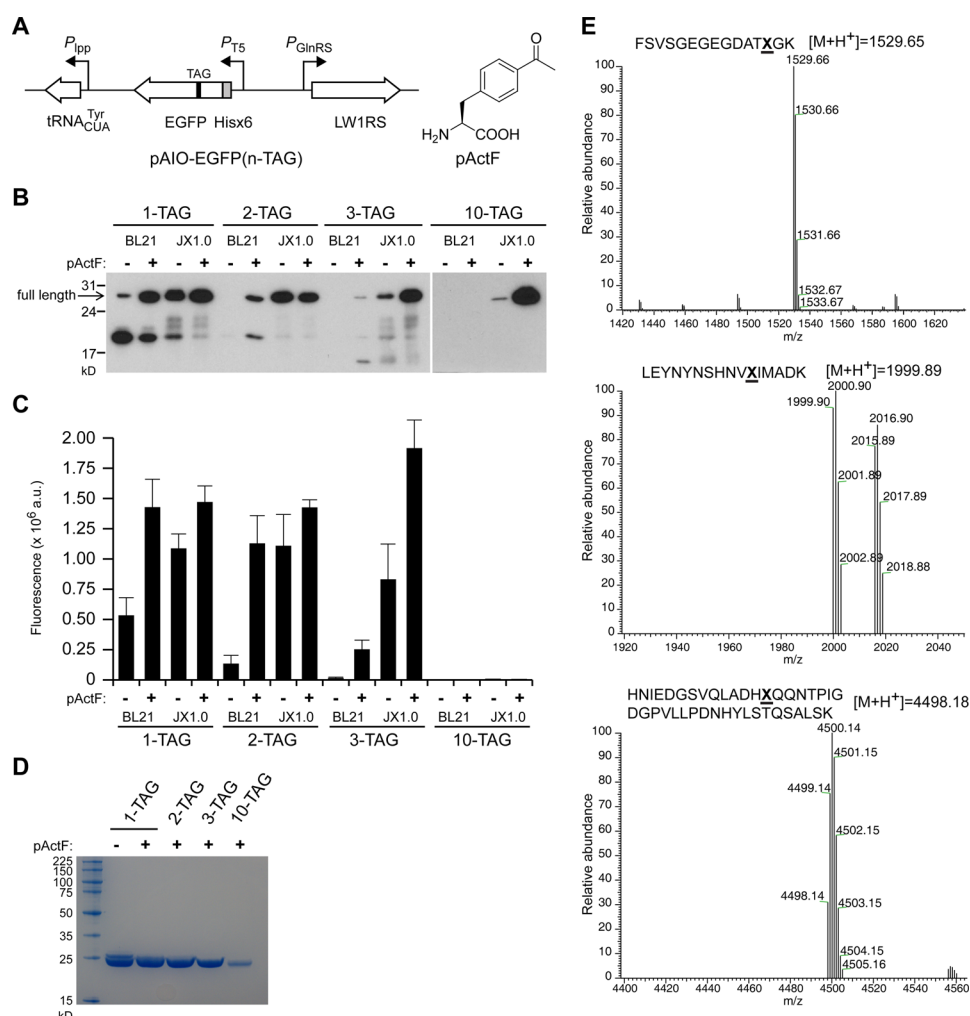


Figure 3. UAG-decoding orthogonal tRNA/synthetase pair reassigns the UAG codon to an amino acid in JX1.0. (A) The plasmid pAIO-EGFP(n-TAG) encodes an orthogonal tRNA^{Tyr}_{CUA}/LWIRS pair and the EGFP reporter with various numbers of TAG codons. Structure of pActF is shown to the right. (B) Western blot analysis of EGFP expression by pAIO-EGFP(n-TAG) in BL21(DE3) and JX1.0 cells in the presence or absence of pActF. A penta-His antibody was used to detect the Hisx6 tag fused at the N-terminus of EGFP. (C) In-cell fluorescence intensity of EGFP expressed by pAIO-EGFP(n-TAG) in BL21(DE3) and JX1.0 cells in the presence or absence of pActF. Samples were normalized for cell numbers in each lane. Measurement was performed on three independent batches of cells, and error bars represent the SEM. (D) SDS-PAGE analysis of EGFP proteins purified from JX1.0 cells expressing pAIO-EGFP(n-TAG) in the absence or presence of pActF. (E) Fourier transform mass spectrometric analysis of EGFP expressed by pAIO-EGFP(3-TAG) in JX1.0 with pActF in the growth medium. The spectra show the intact precursor ions (charge deconvoluted) for the tryptic fragments of EGFP. Monoisotopic masses measured with high accuracy indicate that pActF was incorporated at all three UAG sites. Peptide sequences are shown, and X represents the UAG encoded site. Top: site 39, expected [M + H]⁺ = 1529.65 Da, measured 1529.66 Da. Middle: site 151, expected [M + H]⁺ = 1999.89 Da, measured 1999.90 Da; the 2015.89 peak corresponds to Met oxidation of this pActF-containing peptide. Bottom: site 182, expected [M + H]⁺ = 4498.18 Da, measured 4498.14 Da. No peaks corresponding to any other amino acids at these UAG positions were detected. The signal-to-noise ratio is >1000, translating to a fidelity for pActF incorporation of >99.9%. Tandem MS spectra of these peptides are shown in Supporting Figure S1, which unambiguously confirm the peptide sequences and pActF at position X.

specific LWIRS is less active than the wild type synthetases^{39,40} and thus may generate fewer aminoacylated orthogonal tRNAs. Second, the orthogonal tRNA^{Tyr}_{CUA} has not been evolutionarily optimized for UAG decoding, whereas many natural tRNAs are posttranscriptionally modified through evolution for efficient codon recognition.⁴¹ Third, natural aminoacyl-tRNAs have also been fine-tuned for binding to the elongation factor Tu and the ribosome to achieve efficient translation,^{42,43} whereas the pActF-charged orthogonal tRNA has not been optimized for binding to either. All of these factors can lead to less efficient decoding of UAG as pActF and ribosome drop-off during translation, resulting in truncated protein products. The less efficient decoding of UAG to pActF can also have accumulative effect to account for the decrease of protein yield for the 10-

TAG mutant. Nonetheless, the underlying mechanism for generating these truncated products in the absence of RF1 is intriguing and warrants further studies.

To identify the amino acid incorporated in response to UAG in JX1.0, we purified EGFP expressed in JX1.0 using pAIO-EGFP(3-TAG) in the presence of pActF and analyzed it with mass spectrometry (Figure 3E). The monoisotopic masses of the tryptic peptides clearly showed that pActF was incorporated at all three UAG sites. No peaks corresponding to the incorporation of other amino acids at the UAG sites were detected. The precursor ions of the peptides containing the UAG sites were individually fragmented with an ion trap mass spectrometer. The fragment ion masses were unambiguously assigned, confirming that pActF was incorporated at the UAG

sites (Supporting Figure S1). These results indicate that misreading of UAG in JX1.0 by endogenous tRNAs was outcompeted by the tRNA^{Tyr}_{CUA}/LW1RS pair, which specifically decodes UAG as pActF.

To evaluate the initial response of *E. coli* to RF1 deletion and subsequent UAG reassignment, we assessed the health of JX1.0 using a growth assay. JX1.0 was healthy, cloneable, and stable in culture; no changes in phenotype or genotype were observed after growing over 200 generations. Compared to parental BL21(DE3), JX1.0 showed a slower doubling rate (Figure 4).

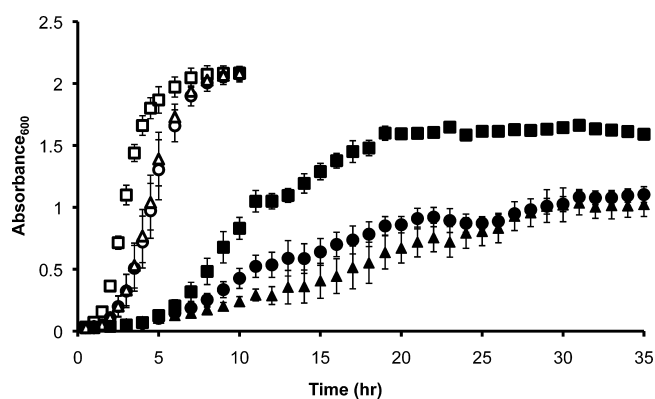


Figure 4. RF1 knockout and UAG reassignment slow down the doubling of JX1.0. The growth curves for the RF1 knockout strain JX1.0 are represented with closed symbols and for the parental BL21(DE3) strain with open symbols. Squares represent cells only; circles represent cells transformed with the pAIO plasmid, which expresses the orthogonal tRNA^{Tyr}_{CUA}/LW1RS pair; and triangles represent cells transformed with the pAIO plasmid and grown in the presence of pActF. Error bars represent the s.e.m., $n = 3$.

The doubling time for JX1.0 was 91 min compared to 26 min for BL21(DE3). This difference suggests that RF1 knockout creates burdens for cells due to the lack of proper termination and relatively weak missuppression of UAGs by near-cognate tRNAs, yet the effect is not lethal as perceived before. Introduction of the pAIO plasmid expressing the orthogonal tRNA^{Tyr}_{CUA}/LW1RS pair in the absence of pActF increased the doubling of JX1.0 and BL21(DE3) to 135 and 41 min, respectively, possibly due to a general toxicity of expressing unacylated tRNAs. The addition of pActF to the growth medium together with the expression of tRNA^{Tyr}_{CUA}/LW1RS further increased the doubling time to 253 min for JX1.0, whereas BL21(DE3) was not affected. This dramatic doubling reduction of JX1.0 is most likely from the efficient incorporation of pActF at UAG positions throughout the proteome, reflecting the pressure generated by UAG reassignment to a sense codon. In contrast, BL21(DE3) expresses RF1, which competes at UAG positions for termination, thereby mitigating this pressure.

We show here that RF1 is nonessential for the *E. coli* species, in contrast to the previous conclusion that RF1 is essential. By reverting Thr246 to Ala and removing the autoregulation of RF2 expression, we recently knocked out RF1 from an engineered *E. coli* strain MDS42 that has a reduced genome.¹³ Because ~700 nonessential genes are deleted in MDS42 and multiple mutations are introduced into its RF2 gene, it is difficult to conclude what factors specifically contribute to RF1 knockout and whether RF1 dispensability is a general property of *E. coli*. Moreover, MDS42 is not a wild type *E. coli* and not suitable for the investigation of code evolution and adaptation

since a variety of genes have been removed. By studying various nonengineered *E. coli* strains, we discovered here that a wild type RF2 is sufficient and necessary for RF1 removal without incurring compensatory mutations. As no other mutations are artificially introduced into these strains, our results clearly demonstrate that RF1 is nonessential for wild type *E. coli*. The successful RF1 knockout in multiple strains indicates that RF1 dispensability is a general property of *E. coli* without peculiar requirements.

Unconditional RF1 removal can be explained by the usage of UAG in *E. coli* genome. UAG is the least-used stop codon in *E. coli*, terminating only ~7% of the total genes.²⁹ Among 302 essential genes in *E. coli*,⁴⁴ only 7 genes are ended with the UAG codon, and these 7 genes all have either a UGA or UAA stop codon downstream in short distances (Supporting Figure S2). If the UAG is readthrough, this second different stop codon can ensure the termination of these proteins with only a small number of amino acids appended. The additional amino acids may not completely disable the function of these proteins, allowing *E. coli* to survive as observed in this study. The few UAG-ending essential genes and the presence of double stop codons can account for the nonessentiality of RF1. Consistently, we show here that a wild type RF2 but not the weakened A246T mutant permitted RF1 knockout, suggesting that efficient termination of the dominant UAA and UGA stop codons is required for *E. coli* survival.

Knockout of RF1 provides novel insights on the evolution of RFs and the genetic code. A major distinction of bacteria from eukaryotes and archaea in protein translation is that bacteria use two different RFs to recognize the stop codons, whereas a single RF is sufficient for organisms in other domains. This difference has served as evidence that the evolution of RF and translation termination between bacteria and eukaryotes/archaea is nonconserved.^{45,46} Our results show that an autonomous *E. coli* strain with a single class I RF can be generated. The viability of such a bacterium blurs the apparent two-*vs*-one RF distinction between prokaryotes and eukaryotes/archaea, suggesting that RF evolution in the three domains of life might be more similar than current differences suggest. In addition, certain eukaryotes have been found to reassign a stop codon to a sense codon and restrict the recognition of eRF1 to the rest of the stop codons.^{2,17,18} It is unclear whether stop codon reassignment caused eRF1 restriction or *vice versa* during the code evolution. We show here that bacteria, in addition to eukaryotes found in nature, can also undergo such stop codon reassignment accompanied by RF changes. Moreover, our results demonstrate that RF changes can synthetically drive the reassignment of a stop codon to an amino acid, providing experimental evidence for this hypothetical evolution pathway.

An autonomous RF1 knockout bacterium will enable new research for understanding code evolution. A major challenge in studying the evolution of the genetic code is that many questions are out of reach of direct experimentation.⁴⁷ No organisms exist containing a primitive or intermediate genetic code for comparison, and natural code evolution takes millions of years. The RF1 knockout strain reported here now afford a previously unavailable model organism to study otherwise intractable questions on the codon reassignment process in real time. We show here that RF1 knockout and the introduction of an orthogonal tRNA/synthetase pair completely reassigned UAG from a stop signal to an amino acid, setting an initial stage for a bacterium to adopt this altered genetic code. Such a stage

had been buried in the recess of evolution and experimentally inaccessible. Using the RF1 knockout strain we may now be able to address questions such as whether the altered code can eventually be fixed, how long this process would take, and what physiological changes are necessary for such adaptations. Answers to these questions would not only help us understand the fundamentals of code evolution but also provide rare empirical data to guide code optimization for specific synthetic purposes. In addition, it is hypothesized that the genetic code started with a set of primitive amino acids and that others were added until the total 20 was reached.^{48,49} It remains mysterious whether the addition of new amino acids to the repertoire affords evolutionary advantage to drive code expansion. This challenging fundamental question can now be investigated experimentally with strains reported here.

Our finding will guide genome recoding to reassign stop codons successfully. It is promising to generate an *E. coli* through genome engineering that has endogenous UAG codons replaced with the synonymous UAA stop codon.¹⁶ However, to reach the final goal of setting aside the UAG codon to encode a Uaa unambiguously, RF1 must be removed from the host strain. MG1655, a K-12 derivative strain, has been used in the current genome engineering effort,¹⁶ which will not permit RF1 knockout on the basis of our findings here. Our results provide a solution for this impasse: reversion of the Thr246 to Ala in the RF2 gene should enable RF1 knockout and clean recoding, which can be conveniently applied at any genome engineering stage before the final RF1 gene removal.

An autonomous RF1 knockout bacterium will afford a unique host for synthesizing and evolving new protein functions and biological properties through Uaa exploitation. To date, directed laboratory evolution of new biosynthesis ability using genetically encoded Uaas has not been feasible due to two demanding requirements: simultaneous Uaa incorporation at multiple sites and an autonomous host. Multisite Uaa incorporation enables synergy and maximizes the exploration of protein sequence space, and a self-sustaining host with such ability is necessary for experimental evolution. The RF1 deletion strains we generated here fulfill both requirements and thus should open the field for Uaa-based directed evolution. Since a variety of new functionalities can be introduced into proteins through Uaas, organisms containing an expanded genetic repertoire have greater potential for novel biosynthesis *via* directed evolution, such as biorenewable production of chemicals and fuels.

METHODS

Strain Construction. Knockout of the *prfA* gene was attempted using a chloramphenicol acetyltransferase (*cat*) cassette *via* established procedures.²³ Briefly, 51 nucleotide overhangs homologous to the regions immediately 5' and 3' of *prfA* were appended to the *cat* gene. One microgram of this cassette was electroporated into various strains harboring the pKD46 plasmid, which expresses the phage λ red recombinase. Chloramphenicol-resistant clones were screened for knockout by genomic PCR using primers 5'-GGA TAA CGA ACG CCT GAA TA-3' and 5'-TCC AGC AGG ATT TCA GCA TC-3'. Positive clones were verified by DNA sequencing and genomic sequencing.

DH10 β was constructed from DH10 β as follows to revert the Thr246 in *prfB* to Ala. A knockin cassette was first generated containing the *prfB* gene from BL21(DE3) transcriptionally coupled to a kanamycin resistant (Kan^R) cassette. The Kan^R cassette was flanked on the 3' end by a 51-nucleotide region homologous to the 3' end of the endogenous *prfB* gene. One microgram of this cassette was

electroporated into DH10 β harboring the pKD46 plasmid. Kan^R clones were screened using PCR and sequence verified for mutation of position 246 to Ala.

Plasmid Construction. All plasmids were assembled by standard cloning methods and confirmed by DNA sequencing. pAIO plasmids containing an EGFP gene with different TAG codons were synthesized as follows: EGFP cassettes with an N-terminal Hisx6 tag and TAG codons at various positions were created using overlapping PCRs. The following sites were used: Y182 for 1-TAG; Y39 and Y182 for 2-TAG; Y39, Y182, and Y151 for 3-TAG; Y39, K101, D102, E132, D133, K140, E172, D173, D190, and V193 for 10-TAG. These cassettes were first cloned into pBP-Blunt (Biopioneer, San Diego, CA) and then digested and ligated into pBK-AIO vectors containing the orthogonal tRNA^{Tyr}_{CUA} and LW1RS³⁹ using *Spe* I and *Bgl*III restriction sites. pBAD vectors containing the 1- or 3-TAG EGFP genes were constructed by inserting the EGFP cassettes into the pBAD/His vector (Invitrogen life technologies) using the *Nco* I and *Hind* III restriction sites.

In-Cell Fluorescence. In-cell fluorescence intensity was determined using a FluoroLog-3 (Horiba Jobin Yvon). *E. coli* colonies were picked and grown in 2xYT medium for 16 h with or without Uaas. Cells were washed twice with PBS buffer and diluted in PBS to an OD₆₀₀ of 0.1. The emission spectrum of EGFP was recorded from 503 to 560 nm using an excitation wavelength of 488 nm. The fluorescence intensity of each sample was compared using the intensity at the 511 nm emission peak.

Growth Assay. A colony was picked for each *E. coli* strain and grown overnight in 2xYT medium with the appropriate antibiotics. Cells were normalized to an OD₆₀₀ of 1 and diluted 1:50 in fresh 2xYT medium with antibiotics and 1 mM Uaa (when applicable). For BL21(DE3) strains, OD₆₀₀ was then measured every 30 min for 10 h. For JX1.0 strains, OD₆₀₀ was measured every 60 min for 48 h. Doubling times were calculated from the exponential growth phase in each strain.

ASSOCIATED CONTENT

Supporting Information

Supporting methods for genome sequencing, Western analyses, protein purification, mass spectrometry, and supporting figures and table. This material is available free of charge *via* the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: lwang@salk.edu.

Author Contributions

[†]These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank W. Fischer for mass spectrometric analyses. This work was supported, in part, by the Howard Hughes Medical Institute and a grant from the Gordon and Betty Moore foundation to J.R.E. L.W. acknowledges support from the Searle Scholar Program (06-L-119), Beckman Young Investigator Program, March of Dimes Foundation (#5-FY08-110), California Institute for Regenerative Medicine (RN1-00577-1) and National Institutes of Health (1DP2OD004744-01).

REFERENCES

- (1) Osawa, S., Jukes, T., Watanabe, K., and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiol. Rev.* 56, 229–264.
- (2) Knight, R. D., Freeland, S. J., and Landweber, L. F. (2001) Rewiring the keyboard: evolvability of the genetic code. *Nat. Rev. Genet.* 2, 49–58.

- (3) Youngman, E. M., McDonald, M. E., and Green, R. (2008) Peptide release on the ribosome: mechanism and implications for translational control. *Annu. Rev. Microbiol.* 62, 353–373.
- (4) Frolova, L., Le Goff, X., Rasmussen, H. H., Cheperegin, S., Drugeon, G., Kress, M., Arman, I., Haenni, A. L., Celis, J. E., Philippe, M., Justesen, J., and Kisselev, L. (1994) A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature* 372, 701–703.
- (5) Dontsova, M., Frolova, L., Vassilieva, J., Piendl, W., Kisselev, L., and Garber, M. (2000) Translation termination factor aRF1 from the archaeon *Methanococcus jannaschii* is active with eukaryotic ribosomes. *FEBS Lett.* 472, 213–216.
- (6) Scolnick, E., Tompkins, R., Caskey, T., and Nirenberg, M. (1968) Release factors differing in specificity for terminator codons. *Proc. Natl. Acad. Sci. U.S.A.* 61, 768–774.
- (7) Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z. Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., 3rd, Smith, H. O., and Venter, J. C. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56.
- (8) Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of *Escherichia coli*. *Science* 292, 498–500.
- (9) Wang, L., and Schultz, P. G. (2005) Expanding the genetic code. *Angew. Chem., Int. Ed.* 44, 34–66.
- (10) Wang, Q., Parrish, A. R., and Wang, L. (2009) Expanding the genetic code for biological studies. *Chem. Biol.* 16, 323–336.
- (11) Wu, X., and Schultz, P. G. (2009) Synthesis at the interface of chemistry and biology. *J. Am. Chem. Soc.* 131, 12497–12515.
- (12) Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* 79, 413–444.
- (13) Johnson, D. B., Xu, J., Shen, Z., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., Xiang, Z., Ecker, J. R., Briggs, S. P., and Wang, L. (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* 7, 779–786.
- (14) Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S., and Schultz, P. G. (2004) An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7566–7571.
- (15) Huang, Y., Russell, W. K., Wan, W., Pai, P. J., Russell, D. H., and Liu, W. (2010) A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in *Escherichia coli*. *Mol. Biosyst.* 6, 683–686.
- (16) Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., Tolonen, A. C., Gianoulis, T. A., Goodman, D. B., Reppas, N. B., Emig, C. J., Bang, D., Hwang, S. J., Jewett, M. C., Jacobson, J. M., and Church, G. M. (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* 333, 348–353.
- (17) Lozupone, C., Knight, R., and Landweber, L. (2001) The molecular basis of nuclear genetic code change in ciliates. *Curr. Biol.* 11, 65–74.
- (18) Inagaki, Y., and Doolittle, W. F. (2001) Class I release factors in ciliates with variant genetic codes. *Nucleic Acids Res.* 29, 921–927.
- (19) Inagaki, Y., Bessho, Y., and Osawa, S. (1993) Lack of peptide-release activity responding to codon UGA in *Mycoplasma capricolum*. *Nucleic Acids Res.* 21, 1335–1338.
- (20) Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M. V., Grechkin, Y., Mseeh, F., Fonstein, M. Y., Overbeek, R., Barabasi, A. L., Oltvai, Z. N., and Osterman, A. L. (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* 185, 5673–5684.
- (21) Rydén, S., and Isaksson, L. (1984) A temperature-sensitive mutant of *Escherichia coli* that shows enhanced misreading of UAG/A and increased efficiency for some tRNA nonsense suppressors. *Mol. Gen. Genet.* 193, 38–45.
- (22) Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S., and Sakamoto, K. (2010) Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res.* 38, 8188–8195.
- (23) Datsenko, K., and Wanner, B. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6640–6645.
- (24) Studier, F. W., Daegelen, P., Lenski, R. E., Maslov, S., and Kim, J. F. (2009) Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. *J. Mol. Biol.* 394, 653–680.
- (25) Kang, Y., Durfee, T., Glasner, J. D., Qiu, Y., Frisch, D., Winterberg, K. M., and Blattner, F. R. (2004) Systematic mutagenesis of the *Escherichia coli* genome. *J. Bacteriol.* 186, 4921–4930.
- (26) Posfai, G., Plunkett, G., 3rd, Feher, T., Frisch, D., Keil, G. M., Umehoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S., de Arruda, M., Burland, V., Harcum, S. W., and Blattner, F. R. (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science* 312, 1044–1046.
- (27) Uno, M., Ito, K., and Nakamura, Y. (1996) Functional specificity of amino acid at position 246 in the tRNA mimicry domain of bacterial release factor 2. *Biochimie* 78, 935–943.
- (28) Dincbas-Renqvist, V., Engstrom, A., Mora, L., Heurgue-Hamard, V., Buckingham, R., and Ehrenberg, M. (2000) A post-translational modification in the GGQ motif of RF2 from *Escherichia coli* stimulates termination of translation. *EMBO J.* 19, 6900–6907.
- (29) Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 28, 292.
- (30) Jeong, H., Barbe, V., Lee, C. H., Vallenet, D., Yu, D. S., Choi, S. H., Couloux, A., Lee, S. W., Yoon, S. H., Cattolico, L., Hur, C. G., Park, H. S., Segurens, B., Kim, S. C., Oh, T. K., Lenski, R. E., Studier, F. W., Daegelen, P., and Kim, J. F. (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol.* 394, 644–652.
- (31) Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., 3rd, Burland, V., Mau, B., Petrosino, J. F., Qin, X., Muzny, D. M., Ayele, M., Gibbs, R. A., Csorgo, B., Posfai, G., Weinstock, G. M., and Blattner, F. R. (2008) The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J. Bacteriol.* 190, 2597–2606.
- (32) Zhang, S., Ryden-Aulin, M., Kirsebom, L. A., and Isaksson, L. A. (1994) Genetic implication for an interaction between release factor one and ribosomal protein L7/L12 in vivo. *J. Mol. Biol.* 242, 614–618.
- (33) Ito, K., Uno, M., and Nakamura, Y. (1998) Single amino acid substitution in prokaryote polypeptide release factor 2 permits it to terminate translation at all three stop codons. *Proc. Natl. Acad. Sci. U.S.A.* 95, 8165–8169.
- (34) Dahlgren, A., and Ryden-Aulin, M. (2000) A novel mutation in ribosomal protein S4 that affects the function of a mutated RF1. *Biochimie* 82, 683–691.
- (35) Kaczanowska, M., and Ryden-Aulin, M. (2004) Temperature sensitivity caused by mutant release factor 1 is suppressed by mutations that affect 16S rRNA maturation. *J. Bacteriol.* 186, 3046–3055.
- (36) Yang, T. T., Cheng, L., and Kain, S. R. (1996) Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* 24, 4592–4593.
- (37) Murgola, E. J. (2005) Translational suppression: when two wrongs do make a right, in *tRNA: Structure, Biosynthesis, and Function* (Soll, D., and RajBhandary, U. L., Eds.), pp 491–509, ASM Press, Washington, DC.
- (38) Wang, L., and Schultz, P. G. (2001) A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* 8, 883–890.
- (39) Wang, L., Zhang, Z., Brock, A., and Schultz, P. G. (2003) Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 56–61.
- (40) Takimoto, J. K., Adams, K. L., Xiang, Z., and Wang, L. (2009) Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Mol. Biosyst.* 5, 931–934.

(41) Yokoyama, S., Nishimura, S. (1995) Modified nucleosides and codon recognition, in *tRNA: Structure, Biosynthesis, and Function* (Soll, D., and RajBhandary, U. L., Eds.), pp 207–223, ASM Press, Washington, DC.

(42) Schrader, J. M., Chapman, S. J., and Uhlenbeck, O. C. (2011) Tuning the affinity of aminoacyl-tRNA to elongation factor Tu for optimal decoding. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5215–5220.

(43) Ledoux, S., and Uhlenbeck, O. C. (2008) Different aa-tRNAs are selected uniformly on the ribosome. *Mol. Cell* 31, 114–123.

(44) Kato, J., and Hashimoto, M. (2007) Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* 3, 132.

(45) Inagaki, Y., and Ford Doolittle, W. (2000) Evolution of the eukaryotic translation termination system: origins of release factors. *Mol. Biol. Evol.* 17, 882–889.

(46) Moreira, D., Kervestin, S., Jean-Jean, O., and Philippe, H. (2002) Evolution of eukaryotic translation elongation and termination factors: variations of evolutionary rate and genetic code deviations. *Mol. Biol. Evol.* 19, 189–200.

(47) Johnson, D. B., and Wang, L. (2010) Imprints of the genetic code in the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8298–8303.

(48) Crick, F. H. (1968) The origin of the genetic code. *J. Mol. Biol.* 38, 367–379.

(49) Trifonov, E. N. (2004) The triplet code from first principles. *J. Biomol. Struct. Dyn.* 22, 1–11.