# Codon usage of highly expressed genes affects proteome-wide translation efficiency

Idan Frumkin[a], Marc J. Lajoie[b], Christopher J. Gregg[b], Gil Hornung[c], George M. Church[b,1], and Yitzhak Pilpel[a,1]

[a]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel; [b]Department of Genetics, Harvard Medical School, Boston, MA 02115; and [c]The Nancy and Stephen Grand Israel National Center for Personalized Medicine, Rehovot 76100, Israel

Contributed by George M. Church, March 24, 2018 (sent for review November 20, 2017; reviewed by Grzegorz Kudla and Dmitri A. Petrov)

**Although the genetic code is redundant, synonymous codons for the same amino acid are not used with equal frequencies in genomes, a phenomenon termed "codon usage bias." Previous studies have demonstrated that synonymous changes in a coding sequence can exert significant *cis* effects on the gene's expression level. However, whether the codon composition of a gene can also affect the translation efficiency of other genes has not been thoroughly explored. To study how codon usage bias influences the cellular economy of translation, we massively converted abundant codons to their rare synonymous counterpart in several highly expressed genes in *Escherichia coli*. This perturbation reduces both the cellular fitness and the translation efficiency of genes that have high initiation rates and are naturally enriched with the manipulated codon, in agreement with theoretical predictions. Interestingly, we could alleviate the observed phenotypes by increasing the supply of the tRNA for the highly demanded codon, thus demonstrating that the codon usage of highly expressed genes was selected in evolution to maintain the efficiency of global protein translation.**

codon usage evolution | tRNA | codon-to-tRNA balance | translation efficiency | genome engineering

Since there are 61 sense codons but only 20 amino acids, most amino acids are encoded by more than a single codon. However, synonymous codons for the same amino acid are not utilized to the same extent across different genes or genomes. This phenomenon, termed "codon usage bias," has been the subject of intense research and was shown to affect gene expression and cellular function through varied processes in bacteria, yeast, and mammals (1–4).

Although differential codon usage can result from neutral processes of mutational biases and drift (5–7), certain codon choices could be specifically favored as they increase the efficiency (8–12) or accuracy (13–17) of protein synthesis. These forces would typically lead to codon biases in a gene because they locally exert their effect on the gene in which the codons reside. Indeed, there is a positive correlation between a gene's expression level and the degree of its codon bias (1). Various systems have demonstrated how altering the codon usage synonymously can alter the expression levels of the manipulated genes (18–21), an effect that could reach more than 1,000-fold (22).

In addition to such *cis* effects, it is possible that codon usage also acts in *trans*, namely, that the codon choice of some genes would affect the translation of others due to a "shared economy" of the entire translation apparatus (23–25). Previous theoretical works have suggested that an increase in the elongation rate may reduce the number of ribosomes on mRNAs and therefore may indirectly increase the rate of initiation of other transcripts due to an increase in the pool of free ribosomes (6, 26). In addition, a recent computational study in yeast has also examined the indirect effects of synonymous codon changes on the translation of the entire transcriptome (27). However, experimental evidence of such changes is absent. Here we ask how manipulating the frequency of a single codon on a small subset of genes influences the synthesis of other proteins.

To tackle this question, we replaced common codons with a synonymous, rare counterpart in several highly expressed genes. We then asked how this massive change in the codon representation in the transcriptome would affect the manipulated genes, other genes, and the physiology and well-being of the cell (Fig. 1). Interestingly, our genetic manipulation did not consistently affect the translation efficiency of the mutated genes, but it did show a profound proteome-wide effect on the translation process. Importantly, the translation efficiency of genes changed in a way that was dependent on the extent to which they contained the affected codons. These observations demonstrate that *trans* effects of codon usage could have strong implications in the cell. We could alleviate these physiological and molecular defects by increasing the tRNA supply for the manipulated codon in a manner that restored codon-to-tRNA balance. Our work demonstrates that codon choice not only tunes the expression level of individual genes but also maintains the efficiency of global protein translation in the cell.

## Results

**Codon Usage Manipulation Leads to Proteome-Wide Changes in Translation Efficiencies in a Codon-Dependent Manner.** We asked how the codon usage of a small subset of genes affects the translation of other genes. To this end, we manipulated the

### Significance

Highly expressed genes are encoded by codons that correspond to abundant tRNAs, a phenomenon thought to ensure high expression levels. An alternative interpretation is that highly expressed genes are codon-biased to support efficient translation of the rest of the proteome. Until recently, it was impossible to examine these alternatives, since statistical analyses provided correlations but not causal mechanistic explanations. Massive genome engineering now allows recoding genes and examining effects on cellular physiology and protein translation. We engineered the *Escherichia coli* genome by changing the codon bias of highly expressed genes. The perturbation affected the translation of other genes, depending on their codon demand, suggesting that codon bias of highly expressed genes ensures translation integrity of the rest of the proteome.

Author contributions: I.F., M.J.L., C.J.G., G.M.C., and Y.P. designed research; I.F., M.J.L., and C.J.G. performed research; M.J.L., C.J.G., G.H., and G.M.C. contributed new reagents/analytic tools; I.F., M.J.L., C.J.G., and Y.P. analyzed data; and I.F. and Y.P. wrote the paper.

Reviewers: G.K., University of Edinburgh; and D.A.P., Stanford University.

Conflict of interest statement: G.M.C. is a co-founder of EnEvolv.



Data deposition: All raw files have been uploaded to the National Center for Biotechnology Information Sequence Read Archive (accession no. SRP142627).

[1]To whom correspondence may be addressed. Email: gchurch@genetics.med.harvard.edu or pilpel@weizmann.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1719375115/-/DCSupplemental.

Published online May 7, 2018.

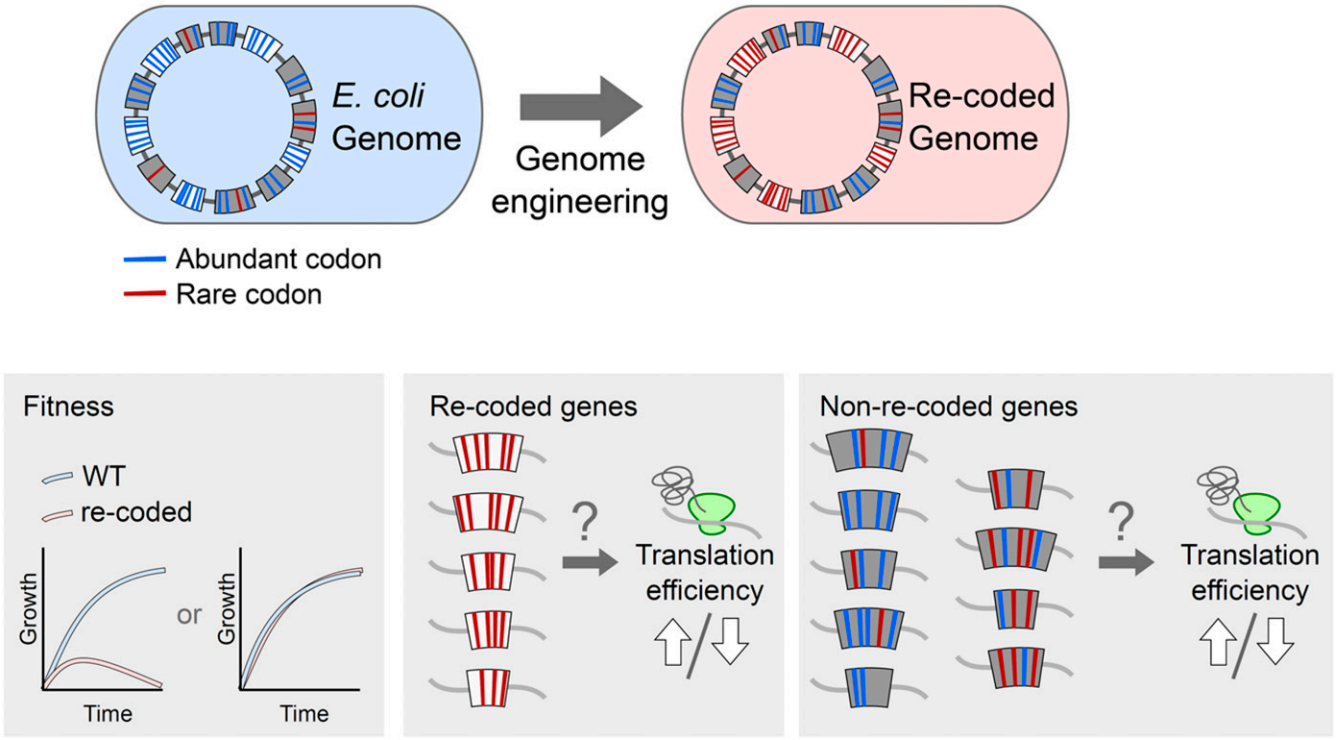


**Fig. 1.** Does the codon usage of a subset of genes affect the translation efficiency of other genes? (*Upper*) Hypothetical genomes of the wild-type and recoded strains are shown. Using genome engineering, we replaced abundant codons (origin codon, blue lines) with rare codons (destination codon, red lines) in highly expressed genes (white background). (*Lower Left*) Two potential effects of recoding on fitness: Recoding could either reduce or not affect the fitness. (*Lower Center*) The translation efficiency of recoded genes could be increased, decreased, or not changed. (*Lower Right*) The translation efficiency of non-recoded genes that have the origin (blue) or destination (red) codon could be increased, decreased, or not changed.

frequency of the arginine codon CGG, since it is the only codon in *Escherichia coli* that is translated by a single-copy tRNA gene and whose tRNA does not translate other codons (see Fig. 2 for codon–anticodon interactions for CGN codons in *E. coli*) (28). Using genome editing, we were able to introduce 60 synonymous mutations into a single genome of an *E. coli* strain that converted CGU and CGC (the "origin codons") to CGG (the "destination codon"). To maximize the effects of our manipulations on the proteome and on the cell, we recoded genes that show high mRNA levels and are highly occupied by ribosomes. Notably, we avoided any recoding of essential, ribosomal, or global regulatory genes, as manipulating these genes might influence the cell directly, hence masking potential effects due to changes in codon usage. We introduced synonymous mutations in the eight genes with the highest ribosome-profiling occupancy score that are not essential and that do not relate directly to the above functions (Table 1). Following our manipulation, the translation demand for the ACG anticodons is reduced by ∼5%, the demand for the CCG anticodon is elevated by ∼3.5-fold, and our recoded genes constitute ∼70% of the new total demand for this codon in the cell (Table 1). See *Materials and Methods* for a full description of the recoded process.

We then asked how our manipulation on the CGG representation in the transcriptome influences translation efficiency in the cell. To this end, we analyzed the transcriptome by RNA-sequencing (RNA-seq) and the proteome by mass spectrometry of the original wild type and the recoded strains (each strain was analyzed with three independent repetitions for both the transcriptome and proteome; see *Materials and Methods*). Then we calculated the translation efficiency of each gene by normalizing the protein level to its corresponding mRNA level based on the three independent repetitions.

Notably, only one of the eight recoded genes showed reduced translation efficiency (Fig. 3*A*), suggesting that the effects of our codon-usage manipulation on the genes that harbor the manipulation are weak. A possible reason for this weak effect is that in the current experiment only a single codon type was manipulated in each recoded gene, in contrast to prior studies in which entire ORFs have been manipulated (18, 21). It is also possible that our manipulations did affect translation efficiency in *cis*, although some compensatory effect, e.g., acting on the initiation level, may have acted to counteract the reduction in elongation. Ultimately, this observation reassures us that our codon manipulations successfully increased translation demand for the CGG codon and provide a unique opportunity to elucidate any *trans* effects of codon usage in highly expressed genes.

We postulated that the increased usage of CGG at the expense of the CGU and CGC codons might reduce the translation efficiency of other genes in the genome that were not mutated, in particular genes that naturally have high usage of CGG. Indeed, we observed 455 genes with increased translation efficiency and 566 genes with decreased translation efficiency at a fold change above or below 1.5 in the recoded strain compared with the wild type (Fig. 3*A*). Strikingly, genes with high (more than five) occurrences of the CGG codon that were not engineered by us demonstrated lower translation efficiencies in the recoded strain compared with the wild-type strain than genes that do not use this codon (Fig. 3*A*, *Inset*). This observation suggests that our CGG codon manipulation affected in *trans* the translation of other, non-recoded genes in the recoded strain. In support of this result, the hundreds of genes that showed reduced translation efficiency demonstrated higher occurrences of the CGG codon compared with the genes with increased translation efficiency (Fig. 3*B*). On the other hand, we observed that genes with increased translation efficiency were enriched with the CGU,

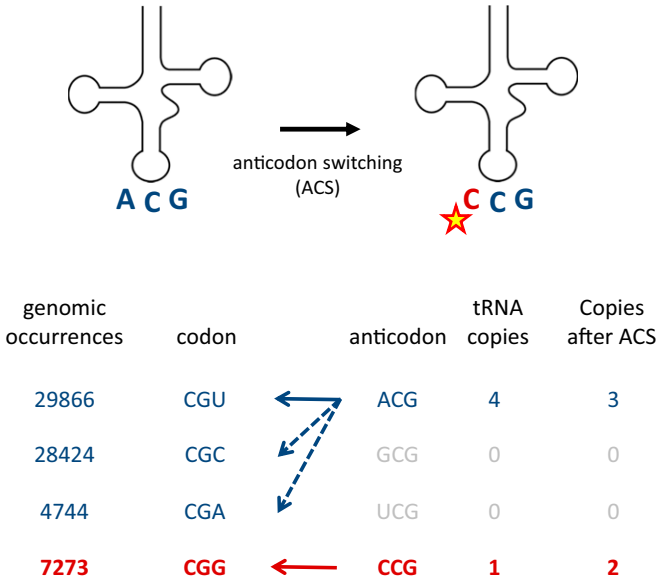


**Fig. 2.** The Arginine CGN box. We recoded CGU and CGC (origin codons) to CGG (destination codon). In *E. coli*, both origin codons are translated by $tRNA^{ACG}$ with the anticodon ACG due to an A-to-I modification that is mediated by the enzyme tRNA-specific adenosine deaminase (tadA). The destination codon is translated solely by $tRNA^{CCG}$, which translates no other codons. $tRNA^{ACG}$ and $tRNA^{CCG}$ appear in the genome with four copies and one copy, respectively. A solid arrow symbolizes fully matched interactions between the codon and anticodon; dashed arrows represent wobble interactions, which are enabled by modifying the ACG anticodon to ICG.

CGC, and CGA codons (Fig. 3*C*). We thus conclude that the increased demand on the CGG codon due to our recoding reduced the translation efficiency of genes that were enriched with this codon, while the relief of demand from the CGU, CGC, and CGA codons increased the translation efficiency of genes that utilize these codons. While most studies measure the resulting change in expression level of a gene whose different codons were synonymously manipulated (18, 21), our results demonstrate how the manipulation of a codon's frequency can affect global translation patterns by changing the translation efficiency of other genes according to their codon usage.

Theory predicts that changes in the elongation rate should have the largest expression effects on genes with high rates of translation initiation because these genes are more likely to suffer from traffic jams and ribosomal collisions (10, 27). Thus, we hypothesized that genes with reduced translation efficiency in the recoded strain should have higher translation initiation rates than genes whose translation efficiency did not decrease. Indeed, genes with reduced translation efficiency demonstrate higher initiation rates, as calculated with the Ribosome Binding Site Calculator (29), than unaffected genes or genes with increased translation efficiency (Fig. 3*G*). The observations that genes with reduced translation efficiency are more enriched with the CGG codon, on one hand, and have higher initiation rates, on the other, strengthens our conclusion that the recoded strain suffers from ribosomal elongation changes compared with wild-type cells. In line with theoretical predictions (10, 27), increasing the dwell time of the ribosome during elongation reduces translation efficiency, provided the initiation rate is sufficiently high.

**Proteome-Wide Changes in Translation Efficiencies Are Alleviated by Increased tRNA Supply.** To confirm our hypothesis that the changes in translation efficiencies resulted from the increased cellular demand for $tRNA^{CCG}$, the tRNA which translates CGG, we decided to elevate the availability of this tRNA and examine the effect on the translation phenotype. We, and others, have recently shown that a mechanism to increase tRNA availability is a mutation in the anticodon that changes the codon specificity of the tRNA (30, 31). We have shown that such anticodon-switching mutations can maintain the functionality of tRNA genes and are utilized by many species as an adaptive mechanism of the cellular tRNA pool.

Thus, we mutated the anticodon of one of the four copies of the $tRNA^{ACG}$ gene from ACG to CCG on the background of the recoded strain (Fig. 2). We then analyzed the transcriptome and proteome of this anticodon-switched strain (based on three independent repetitions) and compared it with both the recoded and wild-type strains. Strikingly, although the genome of the anticodon-switched strain is more similar to the recoded strain, its global translation efficiency pattern clustered together with the wild-type strain and away from the recoded strain (Fig. 3*H*). This observation suggests that manipulating the tRNA pool of the recoded strain restored the translation efficiency of genes back to their normal state.

Indeed, only 124 genes with increased translation efficiency and 408 genes with decreased translation efficiency were identified between the wild-type and anticodon-switched strains (Fig. 3*D*), further demonstrating that the translation efficiency defect in the recoded strain was alleviated upon anticodon switching. Strikingly, while CGG-enriched genes particularly tended to have reduced translation efficiencies in the recoded strain, they demonstrated efficiencies similar to the wild-type strain in the anticodon-switched strain, and the difference in translation efficiency ratios between these genes and CGG-depleted genes was not observed (Fig. 3*D*, *Inset*). Consistently, the genes with increased or decreased translation efficiency between the wild-type and the anticodon-switched strain demonstrated the same distribution of codon occurrences for CGG or CGU + CGC + CGA (Fig. 3 *E* and *F*). These observations suggest that the additional supply of $tRNA^{CCG}$ at the expense of $tRNA^{ACG}$ in the anticodon-switched strain resulted in a more efficient translation of CGG-enriched genes.

We next wanted to examine whether particular codons, especially those involved in the recoding process (CGN codons), were enriched in or depleted from the proteome of the recoded strain. We defined "proteomic codon usage" as the codon occurrences in each gene multiplied by the measured expression level of its protein product. We then calculated this index for each codon and calculated its recoded/wild type ratio (Fig. 4*A*). Remarkably, the CGG codon has the lowest recoded/wild type ratio of all 61 codons, further showing how the introduction of this codon on highly expressed genes resulted in a global proteomic effect. Notably, the two origin codons, namely CGU and CGC, behaved similarly to all other sense codons in this measurement. When the same comparison was performed between the recoded and the anticodon-switched strains, the observed ratio for CGG was significantly increased, while the ratio for the CGU and

**Table 1. Occurrences of origin codons on recoded genes and their contribution to CGG translation demand following codon replacement**

| Gene | No. CGU codons | No. CGC codons | % total CGG translation demand after recoding |
|---|---|---|---|
| *ompA* | 3 | 10 | 25.8 |
| *ompC* | 1 | 12 | 18.5 |
| *ompF* | 2 | 10 | 9.1 |
| *ompX* | 2 | 3 | 5.7 |
| *pal* | 0 | 8 | 4.5 |
| *ahpC* | 1 | 5 | 3.3 |
| *atpE* | 0 | 2 | 2.7 |
| *cspE* | 1 | 0 | 1.8 |
| Total | 10 | 50 | 71.3 |

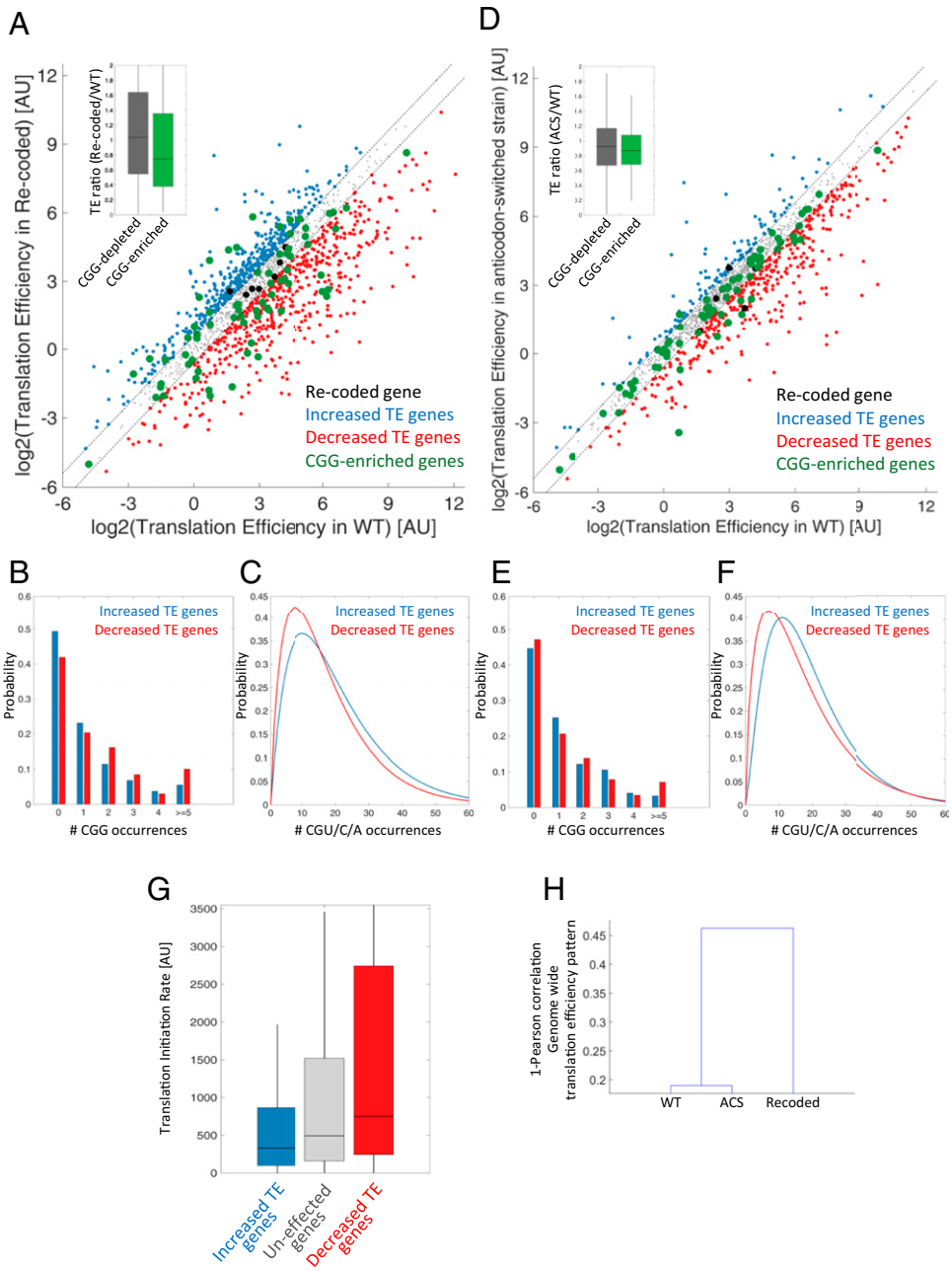


**Fig. 3.** Manipulating the codon frequency of CGG results in global translation efficiency changes. (*A*) We carried out RNA-seq analysis of the transcriptome and mass spectrometry analysis of the proteome for both the wild-type and recoded strains. This allowed us to calculate the translation efficiency (protein/mRNA) for each gene and to classify two gene groups of increased or decreased translation efficiency with a fold-change threshold of 1.5. The eight recoded genes are colored black, the increased translation efficiency group is colored blue, the decreased translation efficiency group is colored red, and CGG-enriched genes are colored green. (*Inset*) Ratios of translation efficiency between recoded and wild-type cells for CGG-enriched genes (more than five occurrences of CGG) and CGG-depleted genes (no occurrences of CGG). CGG-enriched genes show lower translation efficiency ratios ($P$ value = 0.01). (*B*) Distribution of CGG occurrences, translated by $tRNA^{CCG}$, for genes with increased (blue) or decreased (red) translation efficiency (TE) in the recoded strain compared with the wild-type strain. The group of genes with decreased translation efficiency demonstrates higher CGG occurrences ($P$ value = 0.0018). (*C*) Distribution of CGU + CGC + CGA occurrences, all translated by $tRNA^{ACG}$, in genes with increased (blue) or decreased (red) translation efficiency in the recoded strain compared with the wild-type strain. The group of genes with increased translation efficiency demonstrates more codon CGU + CGC + CGA occurrences ($P$ value = $6.79 \times 10^{-5}$). (*D*) To increase the $tRNA^{CCG}$ supply, we mutated the anticodon of $tRNA^{ACG}$ from ACG to CCG on the background of the recoded strain and termed this strain the "anticodon-switched strain." We then analyzed its transcriptome and proteome. Note that many fewer genes, and particularly the CGG-enriched genes in green, now deviate from the diagonal, suggesting that the anticodon-switching mutation alleviated the translational difficulty of the recoded strain. The color code is as in *A*. (*Inset*) CGG-enriched genes now show translation efficiency ratios similar to those of CGG-depleted genes ($P$ value > 0.05). (*E*) As in *B*, but for the genes with increased and decreased translation efficiency in the anticodon-switched strain compared with the wild-type strain. In contrast to the previous comparison in *B*, these two groups utilize the CGG codon to the same extent ($P$ value > 0.05). (*F*) As in *C*, but for the genes with increased and decreased translation efficiency between the wild-type and anticodon-switched strain. In contrast to the previous comparison in *C*, these two groups utilize the CGU + CGC + CGA codon to the same extent ($P$ value > 0.05). (*G*) Translation initiation rates for genes with increased, decreased, or unaffected translation efficiency in the recoded and wild-type strains, as defined in *A*. Note that genes with decreased translation efficiency, which are also enriched with CGG, also show higher initiation rates ($P$ value = 0.01), in agreement with theoretical predictions. (*H*) The translation efficacy pattern of the anticodon-switched (ACS) strain clustered closer to the wild-type strain and away from the recoded strain.

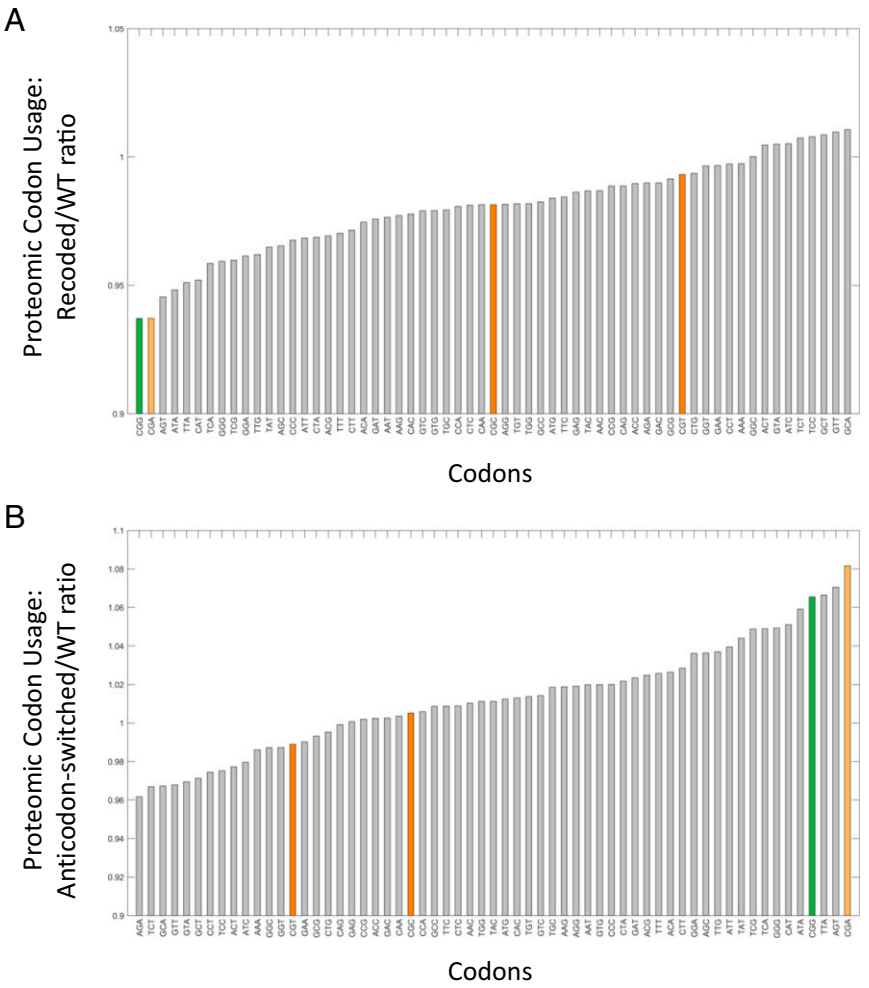


**Fig. 4.** Codon manipulation affects proteomic codon usage. (*A*) We defined the codon proteomic usage as the number of codon occurrences in each gene multiplied by the measured expression level of its protein product. We calculated the recoded/wild type ratio of this index for each of the 61 sense codons and observed that the CGG codon has the lowest value. (*B*) As in *A*, but comparing the anticodon-switched strain with the wild-type strain. Due to the additional supply of $tRNA^{CCG}$, at the expense of $tRNA^{ACG}$, the CGG codon is among the highest codons, an indication that the additional tRNA supply improves the translation efficiency of genes containing CGG. Consistently, CGU and CGC show lower values than in *A* (see text for discussion on CGA codon).

CGC codons was reduced (Fig. 4*B*). These observations are consistent with the additional supply of $tRNA^{CCG}$ at the expense of $tRNA^{ACG}$ in the anticodon-switched strain. Interestingly, the codon CGA, which is translated by $tRNA^{ACG}$, demonstrated behavior similar to CGG and not to CGU or CGC, probably because CGA is the rarest CGN codon and usually co-occurs with CGG on the same genes.

**The Recoded Strain Suffers from Reduced Ability to Translate Transcripts with the CGG Codon.** To directly demonstrate that the codon manipulation in the recoded strain indeed hampered the translation of other genes in a codon-dependent manner, we sought a reporter that would read out the effects of recoding on the translation of the CGG codon. To this end, we used two previously published versions of a YFP reporter with six occurrences of arginine (32), each version having an alternative codon choice, either CGU or CGG (Fig. 5*A*). Since both YFP variants are not under any specific regulation by the cell and were shown to have similar mRNA levels (32), they can serve as a direct proxy for protein synthesis in each strain.

Low-copy plasmids carrying either YFP-CGU or YFP-CGG were transformed to the wild-type and recoded strains. Then YFP production was measured (*Materials and Methods*), and a YFP-CGG/YFP-CGU ratio was calculated for each strain (Fig. 5*A*). The wild-type strain demonstrated maximum YFP production values of 1,224 arbitrary units (AU) and 1,405 AU for YFP-CGU and YFP-CGG, respectively, leading to a YFP-CGG/YFP-CGU ratio of 1.15 (Fig. 5*B*), in agreement with a previous measurement of codon translation speeds in *E. coli* (33). In comparison, the recoded strain showed maximum YFP production values that were consistent with the phenotypes we observed for natural genes, namely, an increased value of 1,316 AU for YFP-CGU and a reduced YFP-CGG value of 1,328 AU. Thus, the YFP-CGG/YFP-CGU ratio was significantly reduced in the recoded strain and was measured to be 0.99 (Fig. 5*B*). This result further supports the view that the increased translational demand for CGG in the recoded strain hampers the production of proteins that utilize the CGG codon.

Since the anticodon-switching mutation alleviated the translation difficulty of the recoded strain, we next asked whether it would also restore the translation efficiency of the YFP-CGG reporter. Hence, we generated three more anticodon-switched strains, each with a different $tRNA^{ACG}$ copy which we mutated, and measured their YFP-CGG/YFP-CGU expression ratio. Indeed, all four strains with the anticodon-switching mutation showed YFP-CGG/YFP-CGU ratios closer to that of the wild-type strain and above the ratio of the

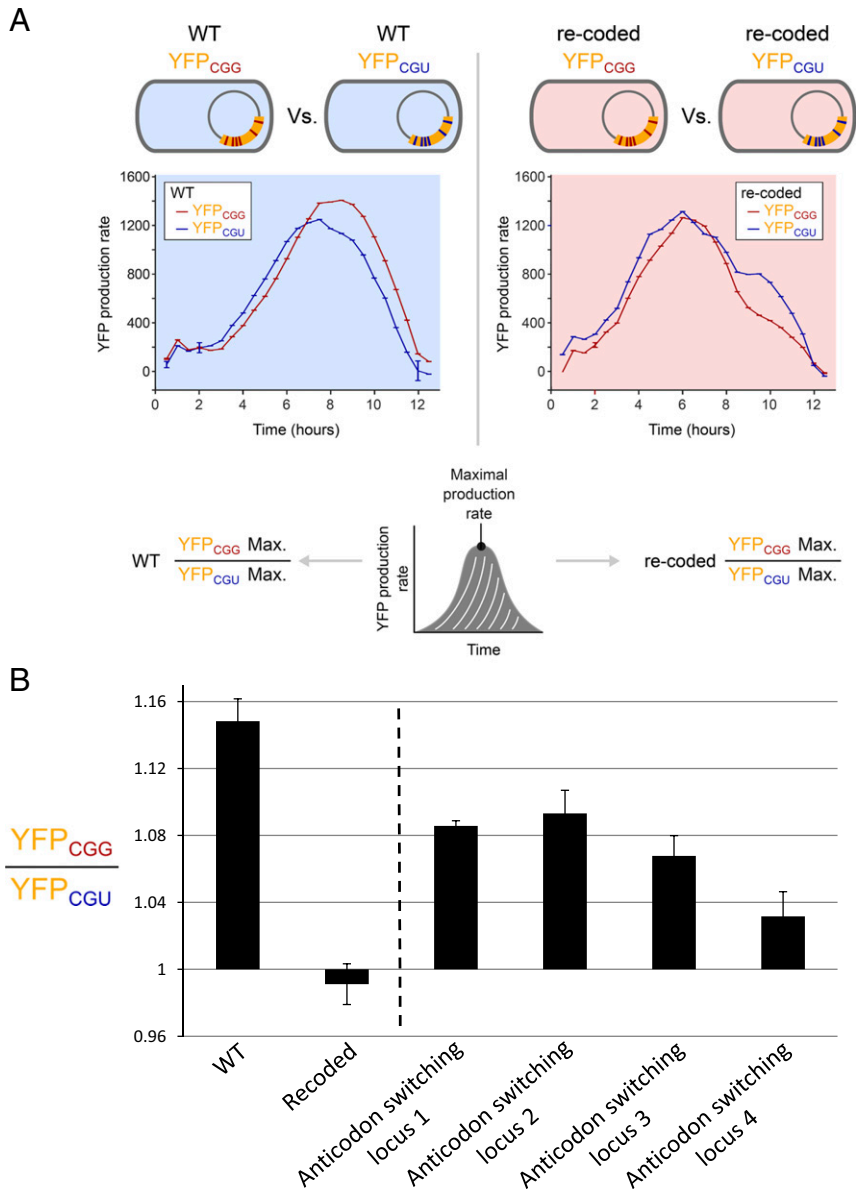


**Fig. 5.** Increased translational demand for CGG hampers the protein synthesis of a reporter gene. (*A*) To directly link frequency manipulation of CGG with the protein synthesis of other genes, we utilized two versions of a YFP-reporter gene with six occurrences of either CGU or CGG. These YFP reporters were introduced separately to either the wild-type or the recoded strain. Following the production of YFP vs. time along the growth cycle allowed us to derive the maximal YFP production for each combination of strain and YFP version. (*B*) For each strain, a YFP-CGG/YFP-CGU ratio is shown for maximal YFP production. The recoded strain demonstrates lower ratios for both these parameters compared with the wild-type strain ($P$ value $= 5.6 \times 10^{-5}$), supporting our observation that changing the codon usage of a small subset of genes hampers the production of other genes that contain the CGG codon. Upon anticodon switching on the background of the recoded strain, the maximal YFP rate is restored to values similar to those in the wild-type strain.

recoded strain (Fig. 5*B*). These observations further support our conclusion that the recoded strain suffers from low availability of tRNA$^{CCG}$ due to the codon manipulation of CGG, which hampers protein production in a codon-specific manner. This perturbation could be alleviated with increased tRNA supply in the cell.

**Increased Codon Usage of a Rare Codon Reduces Cellular Fitness Due to Excessive Use of tRNA Molecules.** The physiological effects between the wild-type and recoded strains encouraged us to ask whether these global translation-efficiency changes disturb cellular growth and reduce fitness. We thus tested whether introducing the rare codon CGG on highly expressed genes is deleterious to the cell. We compared the growth of the wild-type and recoded strains (*Materials and Methods*) and observed that the recoded strain suffers from a growth defect (Fig. 6*A*). We used a recent logistic growth model (34) that calculates relative fitness from growth curves and observed that the relative fitness of the recoded strain is 0.87 compared with the wild-type strain.

We next hypothesized that the growth reduction of the recoded strain is the result of a lack of a sufficient tRNA supply that leads to changes in the translation efficiency of many genes. However, cellular fitness could also be affected by the off-target mutations that the recoded strain accumulated following our genome-engineering efforts. To test our hypothesis, we compared the growth of all four anticodon-switched strains, in which tRNA$^{CCG}$ levels are increased, and observed that they all demonstrated increased relative fitness in comparison with the recoded strain

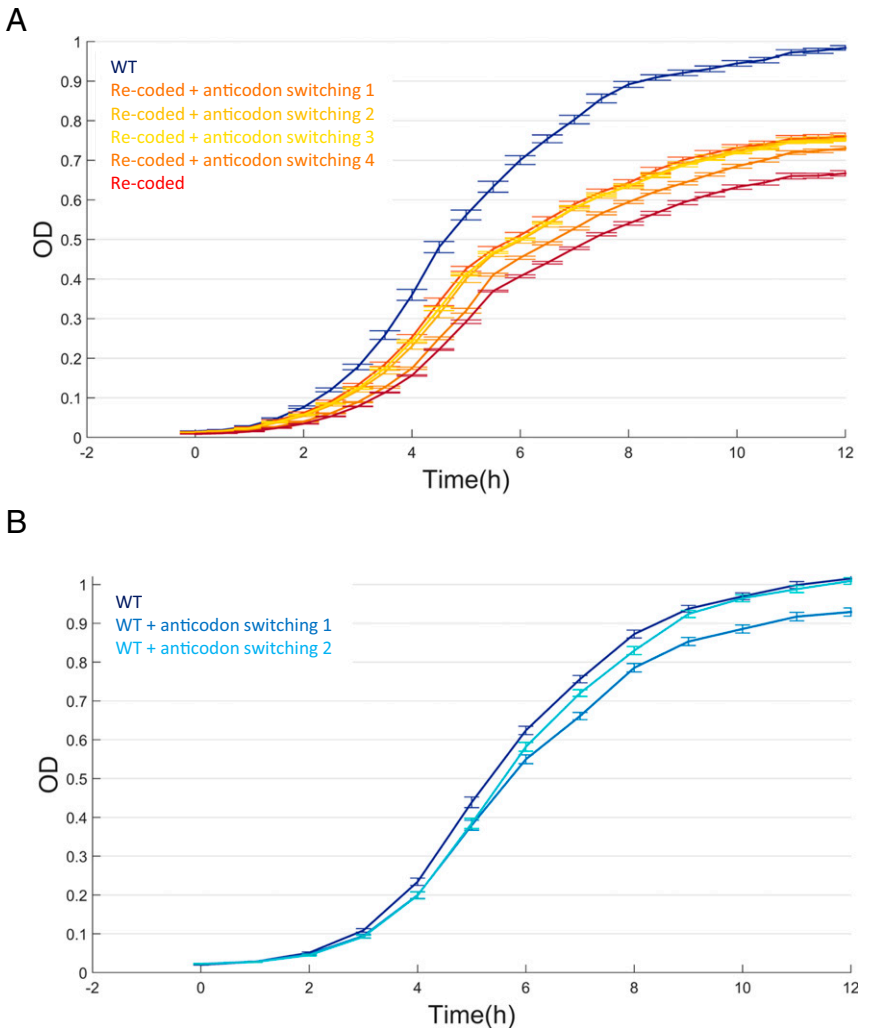


**Fig. 6.** A change in global translation efficiency patterns is deleterious. (*A*) Growth experiment (OD vs. time) of the wild-type strain (blue), the recoded strain (red), and the four anticodon-switched strains ($tRNA^{ACG}$ argQ, dark orange; $tRNA^{ACG}$ argZ, dark yellow; $tRNA^{ACG}$ argY, bright yellow; and $tRNA^{ACG}$ argV, bright orange). The recoded strain demonstrates a reduction in relative fitness to 0.87 compared with the wild-type strain ($P$ value $< 10^{-10}$). The four strains with anticodon switching (increased $tRNA^{CCG}$ supply) on the background of the recoded strain demonstrate higher fitness compared with the recoded strain itself, demonstrating that restored translation efficiency patterns also alleviated the growth defect (relative fitness compared with recoded strain of switched argQ = 1.06, argZ = 1.08, argY = 1.02, and argV = 1.04). (*B*) Switching the anticodon of $tRNA^{ACG}$ from ACG to CCG on the background of the wild-type strain reduces fitness (relative fitness of switched argQ = 0.95 and of argZ = 0.96 compared with the wild-type strain).

(Fig. 6*A*). Importantly, when the same anticodon mutation was inserted on the background of the wild-type strain, a reduction in relative fitness was observed (Fig. 6*B*). These results suggest that introducing a rare codon on highly expressed genes reduces cellular fitness not because of its effects on the manipulated genes themselves but because it hampers the translation of other genes due to an excessive use of tRNA molecules and results in global physiological perturbations.

**Changes in Codon Usage Lead to Mistranslation at Modified Positions.** We did not observe changes in translation efficiency for the recoded genes (Fig. 3*A*), suggesting that the manipulation of CGG demand leads to stronger *trans* effects on global cellular patterns of translation efficiency. However, we hypothesized that our recoding might have other *cis* effects in the form of mistranslation. To test this idea, we used a methodology we recently developed that uses mass spectrometry proteomics data to identify peptides that harbor mistranslation events that result in the replacement of the correct amino acid with a different one (35). Strikingly, we identified such events for two of the recoded genes, *ompC* and *ompA*, exactly at the position at which CGU or CGC, respectively, was mutated into CGG (Fig. 7). The mistranslation event for *ompC* was found in the recoded strain and replaced the coded arginine with glutamine (which has a codon, CAG, that is a near-cognate to CGG) at position 238 of the protein. The mistranslation event for *ompA* was found in the anticodon-switched strain, and it replaced the coded arginine with lysine at position 329 of the protein, suggesting that the additional $tRNA^{ACG}$ supply in this strain did not fully alleviate the mistranslation phenotype, similar to other phenotypes we observed in this study.

## Discussion

Often in biology a correlation between two factors can be explained either by a physiological causal link or by an evolutionary one (36). This is particularly relevant to the correlation that is broadly observed between codon usage and expression level (1, 2, 37–40). On the one hand, optimal codon usage could lead to higher translation speeds (1, 20, 28), suggesting that some proteins enjoy higher expression levels because of their codon

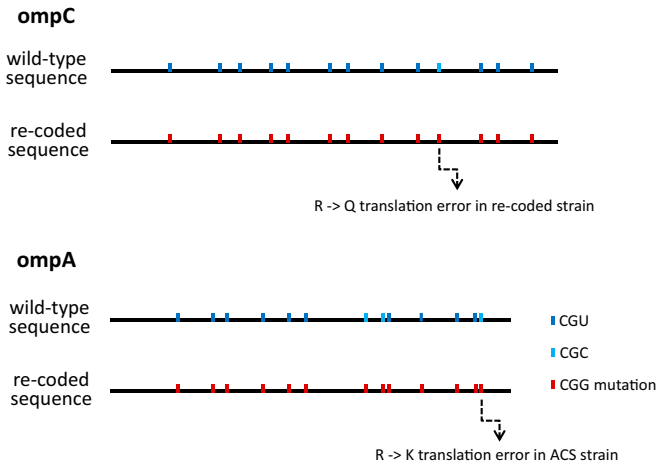


**Fig. 7.** Introducing CGG on highly expressed genes results in mistranslation events. Using a methodology to identify translation errors from mass spectrometry data, we identified such events for two of the recoded genes, *ompC* and *ompA*, exactly at the positions at which CGU or CGC, respectively, was mutated into CGG. While we did not find errors in the wild-type strain, we did observe them in the recoded and anticodon-switched strains. The mistranslation event for *ompC* was found in the recoded strain, and it replaced the coded arginine with glutamine (which has a near-cognate anticodon, CUG) at position 238 of the protein. The mistranslation event for *ompA* was found in the anticodon-switched strain, and it replaced the coded arginine with lysine at position 329 of the protein.

usage. On the other hand, highly expressed genes could be strongly evolutionarily selected for codon optimization compared with lowly expressed genes because the fitness cost of not optimizing them is greater, and hence they force the genome to optimize their codon usage (13, 16, 18). Furthermore, selection for or against specific codons may occur for reasons other than their effect on translation itself, for example to maintain mRNA structure (21), splicing signals (4), or the degradation rate of the transcript (41) or to minimize the cost of gene expression (42).

An additional reason for the strong codon bias of highly expressed genes could be their massive representation in the transcriptome and overall impact on the translation machinery. Thus, a nonoptimal codon present on a gene with a high mRNA level could disturb the translation of other genes that utilize this codon (27).

Here we study this possibility, and our results provide experimental evidence that introducing a rare codon into highly expressed genes indeed hampers the protein production of other genes, especially those that are encoded with that codon. In our experimental system, we introduced 60 new occurrences of a rare arginine codon, CGG, on highly expressed genes, and this manipulation led to a reduction in the translation efficiency of CGG-containing genes. Importantly, we confirmed the protein-synthesis difficulties of such genes by using two versions of a reporter gene that either uses the CGG codon or avoids it. Thus, our results demonstrate that the translation of a certain gene is influenced not only by its own regulation in the form of codon usage or mRNA level but also by the translation efficiency of all the other genes in the cellular genetic network.

One limitation of the genome-engineering approach we took here is the accumulation of off-target mutations in addition to the planned mutations (*Materials and Methods*). However, we argue that our observed phenotypes are mostly due to the on-target mutations for the following reasons. First, the off-target mutations in the recoded strain did not manipulate CGG specifically (in direct contrast to the on-target mutations), and they are diverse in nature: Half are synonymous mutations, seven are intergenic, 20% occurred inside unvalidated or uncharacterized proteins, and none occurred in genes that are part of the translation machinery. It is extremely unlikely that the off-target mutations, which are diverse and do not show any pattern, would lead to the CGG-specific phenotype we observed. Second, of the 1,021 genes with increased or decreased translation efficiency, only eight had off-target mutations in them. Importantly, these genes were either enriched in or depleted from the CGG codon in agreement with the overall increased translational demand in the recoded strain. This phenotype, together with the observation for YFP production discussed above, is extremely unlikely to occur as a result of the off-target mutations, especially given that we planned the on-target changes to directly manipulate CGG. Third, and most importantly, we could cure most of the phenotypic and molecular defects of the recoded strain by tRNA manipulation in the form of anticodon switching. This is a very clear indication that the recoded strain suffers mainly from a direct effect of recoding and less from off-target mutations. In particular, we observed the following phenotypes for the anticodon-switched strain that support the direct effects of the on-target mutations: (*i*) The anticodon-switching strain has an additional copy of $tRNA^{CCG}$, and therefore the translation efficiency of the perfectly matching codon CGG-containing genes is increased. (*ii*) The anticodon-switching stain has one less copy of $tRNA^{ACG}$, and therefore the translation efficiency of CGU/CGC-containing genes is decreased. (*iii*) The translation efficiency pattern of the anticodon-switched strain clusters with that of the wild type, although it is genetically closer to the recoded strain. This result shows that the recoded strain indeed suffers from an imbalance between CGG demand and $tRNA^{CCG}$ supply that is rectified by anticodon switching. If off-target mutations were dominant in determining the phenotypes we observe, we would have obtained exactly the opposite result: The anticodon-switched recoded strain would have clustered with the recoded strain since they share the off-target mutations.

Our observations are also relevant to the context of heterologous gene expression. The codon usage of a gene is most relevant to its successful expression in a foreign system (1, 22). However, the effects on cell physiology of artificially expressing a gene that is not native to the genome, usually to high levels, have not been explored thoroughly. Our results allow us to measure and appreciate the proteome-wide changes under such conditions. Although in our systems the recoded genes are natural to the genome, it is likely that they apply to heterologous proteins, which are thus predicted here to affect the translation apparatus in a similar manner.

Finally, this work raises the question of whether changes in global translation efficiencies could pose a challenge to the translation machinery both physiologically and evolutionarily. Previous work has demonstrated how the codon-to-tRNA balance reacts to changes in the environment (32, 43, 44), to the formation of cancerous tumor (45), or to an evolutionary challenge (30, 46). In agreement with these works, we observed that the recoded strain suffers from a growth defect, providing a need for selection to optimize the translation economy in the cell. Interestingly, we could alleviate these translation and growth phenotypes by increasing the tRNA supply to meet the new CGG demand. Thus our work demonstrates that codons and tRNA genes may coevolve not only to tune the expression level of individual (highly expressed) genes but also to maintain the efficiency of global protein translation in the cell.

## Materials and Methods

**Genome Engineering.** To introduce synonymous mutations that replace the origin codons (CGU and CGC) with the destination codon (CGG), we used Co-Selection Multiplex Automated Genome Engineering (CoS-MAGE) as previously described (47–49). The background *E. coli* strain was EcM2.1, a strain especially designed for high MAGE efficiency (50). Each day one

CoS-MAGE cycle was performed with 10 90-mer oligos selecting either for or against the tolC marker. Briefly, cells were grown overnight at 34 °C. Then, 30 μL of the saturated culture was transferred into 3 mL of fresh LBL medium (as described in ref. 50) until OD = 0.4 was reached and then was moved to a shaking water bath (350 rpm) at 42 °C for 15 min, after which it was moved immediately to ice. Next, 1 mL was transferred to an Eppendorf tube, and cells were washed twice with sterile water at a centrifuge speed of 13,000 × *g* for 30 s. Next, the bacterial pellet was dissolved in 50 μL of double-distilled water containing 4 μM of 10 SS-DNA oligos and a tolC dedicated oligo and was transferred into a cuvette. Electroporation was performed at 1.78 kV, 200 ohms, 25 μF. After electroporation, the bacteria were transferred into 1 mL of fresh LBL medium for recovery and then were moved to selection medium. Selection for tolC was performed on liquid LBL + 0.005% SDS, and selection against tolC was performed with LBLCCoV plates that contain 50 μg/mL carbenicillin + 64 μg/mL vancomycin + purified Colicin E1 (51). Every four CoS-MAGE cycles, random colonies were screened for on-target mutations via multiplex allele-specific colony PCR (mascPCR), and colonies with highest number of mutations were sequenced for further verification. Then the best colony was picked for successive engineering via additional CoS-MAGE cycles.

To facilitate recoding efforts, we split the eight targeted genes into two groups according to their genomic loci (Fig. S1). Strain A was recoded for the genes *ahpC*, *cspE*, *pal*, *ompX*, *ompF*, and *ompA*. Strain B was recoded for the genes *atpE* and *ompC*. After engineering was completed for both strains, we merged their genome by following the conjugative assembly genome engineering (CAGE) protocol (52, 53). Strain A was the donor and thus was transformed with the pRK29 plasmid; strain B was the recipient. Selection for final strain was done on LBL plates with 0.005% SDS + 100 μg/mL spectinomycin + 5 μg/mL gentamycin. To maintain as similar a genetic background as possible between the recoded and the wild-type strains, we also transformed the resistance markers for SDS, spectinomycin, and gentamycin to the same loci as in the recoded strain.

We confirmed the successful introduction of all 60 planned genomic changes by whole-genome sequencing. We also revealed 58 additional off-target mutations, as typically happens with this genome-editing technology (54). Any off-target gene that was mutated unintentionally was excluded from all our down-stream analyses. Importantly, the off-target mutations in the recoded strain did not manipulate CGG specifically (in direct contrast to the on-target mutations), and they were of diverse nature: Half were synonymous mutations, seven were intergenic and did not affect any gene, 20% occurred inside unvalidated or uncharacterized proteins, and none occurred in genes that are part of the translation machinery. Thus, it is unlikely that the off-target mutations, which were diverse and did not show any pattern, would lead to the CGG-specific phenotype we observed. See the full discussion on off-target mutations in *Discussion*.

Off-target mutations are listed in Dataset S1, and strains, CoS-MAGE oligos, mascPCR primers, and tolC information are given in Dataset S2.

**Liquid Growth Measurements.** Cultures were grown for 48 h in LB medium at 30 °C back-diluted in a 1:100 ratio and dispensed on 96-well plates in a checkerboard manner. Wells were measured for optical density at $OD_{600}$, and measurements were taken during growth at 30-min intervals until stationary phase was reached. For each strain, a growth curve was obtained by averaging all wells. Then we converted these curves to relative fitness using the Curveball approach (34).

**YFP Production Measurements.** Each strain was transformed with the plasmid pZS*11-YFP-Kan harboring a Kan-resistance cassette and a YFP gene with six occurrences of either CGG or CGU. Growth was measured as described above, but only YFP measurements (excitation = 500 ± 25 nm, emission = 540 ± 25 nm) were taken in addition to $OD_{600}$. The YFP production rate was measured as previously described (55) by subtracting the YFP value at time *t* from the YFP value in time $t - 1$ and dividing the result by the $OD_{600}$ value at time *t*. The maximal production rate was defined as the highest value on this curve. We followed the YFP production along the entire growth curve (from lag to saturation), as it includes all the different physiological states the cells experience under these growth conditions. Then, a YFP-CGG/YFP-CGU ratio was calculated for each strain.

**Harvesting Cells for Transcriptome and Proteome Analyses.** To compare the transcriptome of the wild-type, recoded, and anticodon-switched strains, we grew each strain with three independent repetitions in LB at 30 °C overnight. Then, for each repetition, 400 μL of culture was diluted in 50 mL of LB and was grown until cells reached an $OD_{600}$ of 0.4. Cells were flash-frozen in liquid nitrogen, and the pellets were used for either RNA-sequencing or mass spectrometry.

**Transcriptome Analysis.** RNA-seq was performed as described by Dar et al. (56). RNA was extracted with the standard protocol. Then samples were treated with DNase using the TURBO DNA-free Kit (Ambion), and rRNA was depleted by using the Ribo-Zero rRNA Removal Kit (Epicenter). Next, strand-specific RNA-seq was performed with the NEBNext Ultra Directional RNA Library Prep Kit (New England Biolabs). Libraries were sequenced by using the NextSeq system (Illumina) with a read length of 50 nucleotides.

**Proteome Analysis.** See full description in *Supporting Information*.

**ACKNOWLEDGMENTS.** We thank Arvind R. Subramaniam for suppling plasmids and helpful discussions; Nir Fluman for help with ribosome-profiling analysis; Tslil Ast, Raz Bar-Ziv, Hila Gingold, Daniel B. Goodman, Gleb Kuznetsov, Michael Napolitano, Dan Bar-Yaacov, Avihu Yona, and Emmanuel Levy for helpful discussions and critical reading of the manuscript; Maya Schuldiner and Ron Milo for many supportive discussions; Rotem Sorek, Maya Shamir, and Shany Doron for help with the RNA-seq protocol; Ernest Mordret and Avia Yehonadav for help with the mistranslation pipeline; Shlomit Gilad, Sima Benjamin, Barak Markus, Alon Savidor, and Yishai Levin from the Nancy and Stephen Grand Israel National Center for Personalized Medicine for assistance with high-throughput data; Yoav Ram for help with Curveball; and Nitai Steinberg for figure design. I.F. was supported by a European Molecular Biology Organization short-term fellowship and an Azrieli PhD Fellowship award from the Azrieli Foundation. Grant support was provided by the Minerva Center for Live Emulation of Evolution in the Lab of the Weizmann Institute of Science and the Israel Science Foundation.

1. Plotkin JB, Kudla G (2011) Synonymous but not the same: The causes and consequences of codon bias. *Nat Rev Genet* 12:32–42.
2. Gingold H, Pilpel Y (2011) Determinants of translation efficiency and accuracy. *Mol Syst Biol* 7:481.
3. Chamary JV, Parmley JL, Hurst LD (2006) Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet* 7:98–108.
4. Hershberg R, Petrov DA (2008) Selection on codon bias. *Annu Rev Genet* 42:287–299.
5. Shah P, Gilchrist MA (2011) Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc Natl Acad Sci USA* 108:10231–10236.
6. Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129:897–907.
7. Duret L (2002) Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev* 12:640–649.
8. Pechmann S, Frydman J (2013) Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol* 20:237–243.
9. Dixon JR, et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–380.
10. Tuller T, et al. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141:344–354.
11. Tuller T, Waldman YY, Kupiec M, Ruppin E (2010) Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci USA* 107:3645–3650.
12. Subramaniam AR, Zid BM, O'Shea EK (2014) An integrated approach reveals regulatory controls on bacterial translation elongation. *Cell* 159:1200–1211.
13. Drummond DA, Wilke COC (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341–352.
14. Zhou T, Weems M, Wilke CO (2009) Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol Biol Evol* 26:1571–1580.
15. Wilke CO, Drummond DA (2010) Signatures of protein biophysics in coding sequence evolution. *Curr Opin Struct Biol* 20:385–389.
16. Akashi H (1994) Synonymous codon usage in Drosophila melanogaster: Natural selection and translational accuracy. *Genetics* 136:927–935.
17. Stoletzki N, Eyre-Walker A (2007) Synonymous codon usage in Escherichia coli: Selection for translational accuracy. *Mol Biol Evol* 24:374–381.
18. Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science* 324:255–258.
19. Zhou M, et al. (2013) Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 495:111–115.
20. Navon S, Pilpel Y (2011) The role of codon selection in regulation of translation efficiency deduced from synthetic libraries. *Genome Biol* 12:R12.
21. Goodman DB, Church GM, Kosuri S (2013) Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342:475–479.
22. Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. *Trends Biotechnol* 22:346–353.
23. Qian W, Yang JR, Pearson NM, Maclean C, Zhang J (2012) Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet* 8:e1002603.
24. Akashi H, Eyre-Walker A (1998) Translational selection and molecular evolution. *Curr Opin Genet Dev* 8:688–693.

25. Akashi H (2003) Translational selection and yeast proteome evolution. *Genetics* 164: 1291–1303.
26. Andersson SG, Kurland CG (1990) Codon preferences in free-living microorganisms. *Microbiol Rev* 54:198–210.
27. Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB (2013) Rate-limiting steps in yeast protein translation. *Cell* 153:1589–1601.
28. dos Reis M, Savva R, Wernisch L (2004) Solving the riddle of codon usage preferences: A test for translational selection. *Nucleic Acids Res* 32:5036–5044.
29. Salis HM (2011) The ribosome binding site calculator. *Methods Enzymol* 498:19–42.
30. Yona AH, et al. (2013) tRNA genes rapidly change in evolution to meet novel translational demands. *eLife* 2:e01339.
31. Rogers HH, Griffiths-Jones S (2014) tRNA anticodon shifts in eukaryotic genomes. *RNA* 20:269–281.
32. Subramaniam AR, Pan T, Cluzel P (2013) Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. *Proc Natl Acad Sci USA* 110:2419–2424.
33. Chevance FFV, Le Guyon S, Hughes KT (2014) The effects of codon context on in vivo translation speed. *PLoS Genet* 10:e1004392.
34. Ram Y, et al. (2015) Predicting microbial relative growth in a mixed culture from growth curve data. bioRxiv:10.1101/022640. Preprint, posted August 3, 2016.
35. Mordret E, et al. (2018) Systematic detection of amino acid substitutions in proteome reveals mechanistic basis of ribosome errors. bioRxiv:10.1101/255943. Preprint, posted January 29, 2018.
36. Karmon A, Pilpel Y (2016) Biological causal links on physiological and evolutionary time scales. *Elife* 5:e14424.
37. Grosjean H, Fiers W (1982) Preferential codon usage in prokaryotic genes: The optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* 18:199–209.
38. Sharp PM, Tuohy TM, Mosurski KR (1986) Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* 14:5125–5143.
39. Sharp PM, Li W-HH (1987) The codon adaptation index–A measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15: 1281–1295.
40. Man O, Pilpel Y (2007) Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. *Nat Genet* 39:415–421.
41. Presnyak V, et al. (2015) Codon optimality is a major determinant of mRNA stability. *Cell* 160:1111–1124.
42. Frumkin I, et al. (2017) Gene architectures that minimize cost of gene expression. *Mol Cell* 65:142–153.
43. Dittmar KA, Sørensen MA, Elf J, Ehrenberg M, Pan T (2005) Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep* 6:151–157.
44. Wiltrout E, Goodenbour JM, Fréchin M, Pan T (2012) Misacylation of tRNA with methionine in Saccharomyces cerevisiae. *Nucleic Acids Res* 40:10494–10506.
45. Gingold H, et al. (2014) A dual program for translation regulation in cellular proliferation and differentiation. *Cell* 158:1281–1292.
46. Yona AH, Frumkin I, Pilpel Y (2015) A relay race on the evolutionary adaptation spectrum. *Cell* 163:549–559.
47. Wang HH, et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460:894–898.
48. Gallagher RR, Li Z, Lewis AO, Isaacs FJ (2014) Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat Protoc* 9:2301–2316.
49. Carr PA, et al. (2012) Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res* 40:e132.
50. Gregg CJ, et al. (2014) Rational optimization of tolC as a powerful dual selectable marker for genome engineering. *Nucleic Acids Res* 42:4779–4790.
51. Schwartz SA, Helinski DR (1971) Purification and characterization of colicin E1. *J Biol Chem* 246:6318–6327.
52. Ma NJ, Moonan DW, Isaacs FJ (2014) Precise manipulation of bacterial chromosomes by conjugative assembly genome engineering. *Nat Protoc* 9:2285–2300.
53. Isaacs FJ, et al. (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* 333:348–353.
54. Lajoie MJ, et al. (2013) Genomically recoded organisms expand biological functions. *Science* 342:357–360.
55. Zeevi D, et al. (2011) Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Res* 21:2114–2128.
56. Dar D, et al. (2016) Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* 352:aad9822.