PNAS RESEARCH ARTICLE | GENETICS

Synonymous edits in the *Escherichia coli* genome have substantial and condition-dependent effects on fitness

Dong-Dong Yang[a], Leo M. Rusch[a], Karl A. Widney[a], Andrew B. Morgenthaler[a,b], and Shelley D. Copley[a,1]

Edited by Andrew Murray, Harvard University, Cambridge, MA; received September 28, 2023; accepted December 19, 2023

CRISPR-Cas-based genome editing is widely used in bacteria at scales ranging from construction of individual mutants to massively parallel libraries. This procedure relies on guide RNA-directed cleavage of the genome followed by repair with a template that introduces a desired mutation along with synonymous "immunizing" mutations to prevent re-cleavage of the genome after editing. Because the immunizing mutations do not change the protein sequence, they are often assumed to be neutral. However, synonymous mutations can change mRNA structures in ways that alter levels of the encoded proteins. We have tested the assumption that immunizing mutations are neutral by constructing a library of over 50,000 edits that consist of only synonymous mutations in *Escherichia coli*. Thousands of edits had substantial effects on fitness during growth of *E. coli* on acetate, a poor carbon source that is toxic at high concentrations. The percentage of high-impact edits varied considerably between genes and at different positions within genes. We reconstructed clones with high-impact edits and found that 69% indeed had significant effects on growth in acetate. Interestingly, fewer edits affected fitness during growth in glucose, a preferred carbon source, suggesting that changes in protein expression caused by synonymous mutations may be most important when an organism encounters challenging conditions. Finally, we showed that synonymous edits can have widespread effects; a synonymous edit at the 5′ end of *ptsI* altered expression of hundreds of genes. Our results suggest that the synonymous immunizing edits introduced during CRISPR-Cas-based genome editing should not be assumed to be innocuous.

genome editing | synonymous mutations | CRISPR

Significance

Synonymous "immunizing" mutations in and near the protospacer-adjacent motif site are typically introduced during CRISPR-Cas9 genome editing in bacteria. Our work shows that immunizing mutations can have substantial fitness effects, particularly under strong selection conditions. Twenty-two percent of edits containing only immunizing mutations had high impacts on fitness during growth of *E. coli* in 50 mM acetate. This finding is profoundly important for the wide range of experimental investigations that utilize this genome editing protocol because unexpected effects of immunizing mutations may either augment or diminish the effects of target mutations. Further, our demonstration that clusters of two to six synonymous mutations can have strong fitness effects, especially under strong selection conditions, contributes to the growing recognition that synonymous mutations are not necessarily silent.

CRISPR/Cas-mediated genome editing has become a mainstay of the genetics tool kit for many bacteria. This technology takes advantage of guide RNA-directed nucleolytic cleavage at a specific site in the genome. In bacteria, precise genomic edits can be introduced when the double-strand break is repaired by homologous recombination with a homology repair template (HRT) containing a desired mutation. Synonymous mutations in or near the protospacer-adjacent motif (PAM) site are typically included in the HRT to prevent re-cleavage of the genome after successful editing. Editing efficiency is increased due to the death of cells that have not incorporated the desired mutation (1, 2).

The synonymous mutations introduced by CRISPR-Cas genome editing do not alter the sequence of encoded proteins and are often assumed to be innocuous. The assumption that synonymous mutations are "silent" dates back to the original elucidation of the genetic code (3). However, mounting evidence suggests that synonymous mutations can have important fitness effects. Biases in codon usage suggest that synonymous codons are not always interchangeable (4–6). Synonymous mutations can generate or disrupt promoters for downstream genes (7–9), impact sRNA binding (10, 11), alter splicing (12–15), and affect mRNA stability (16–18). They can impact the rates of translation initiation (16, 19–23) and elongation (24, 25) and even alter protein folding; substitution of rare codons with common ones can affect translational pauses that allow polypeptide folding without interference from the downstream sequence (24, 26–28). Synonymous mutations can affect bacterial growth in the presence of an antibiotic (29), a sub-optimal carbon source (30, 31) or an inefficient metabolic enzyme (16). Analyses of several organisms (32–34) suggest that 5 to 39% of synonymous sites are under purifying selection. Further, synonymous mutations are associated with a number of human diseases (10, 12, 15, 35, 36). The documented effects of synonymous mutations beg the question of whether the clusters of synonymous mutations used in HRTs for CRISPR-Cas-mediated genome editing have significant fitness effects.

We have investigated the fitness effects of clusters of synonymous mutations by introducing 50,102 synonymous edits into 30 *Escherichia coli* genes involved in central carbon metabolism. Each editing cassette introduced a synonymous "target" mutation as well as 1 to 5 "immunizing" synonymous mutations in or near the PAM site to prevent re-cleavage

Author contributions: D.-D.Y. and S.D.C. designed research; D.-D.Y., L.M.R., and A.B.M. performed research; D.-D.Y., L.M.R., K.A.W., and S.D.C. analyzed data; and D.-D.Y. and S.D.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: shelley.copley@colorado.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2316834121/-/DCSupplemental.

Published January 22, 2024.

Downloaded from https://www.pnas.org by 137.75.80.24 on March 18, 2025 from IP address 137.75.80.24.

of the genome after editing. A barcode in each editing cassette plasmid enabled quantification of variants. We followed the levels of each barcode during growth of the library in 0.2% glucose and in 50 and 200 mM sodium acetate (acetate hereafter), a less-preferred carbon source. We identified thousands of synonymous edits with substantial effects on fitness during growth in acetate, but far fewer during growth in glucose. Among nearly 1,700 sets of immunizing mutations used to edit the 30 genes, 22% had substantial fitness effects during growth in 50 mM acetate. A synonymous edit in a gene with regulatory functions had significant impacts on expression of hundreds of genes during growth in acetate. These results demonstrate that fitness effects due to synonymous immunizing mutations may contribute in unexpected ways to the fitness effects of target mutations of interest introduced during CRISPR-Cas genome editing.

Results

Synonymous Edits Have Striking Effects on Growth Rate in 50 mM Acetate. We used Onyx™ technology (Fig. 1), an improved implementation of CRISPR-enabled trackable genome engineering (2), to introduce 50,102 synonymous edits into 30 genes involved in central carbon metabolism in *E. coli*. These edits were designed to change each codon of the 30 genes to all other synonymous codons at that position with concomitant introduction of immunizing synonymous mutations in and near the PAM site to prevent re-cleavage of the genome after introduction of the target mutation. Editing was carried out in a strain of *E. coli* MG1655 with an 82 bp deletion upstream of *pyrE* that corrects a pyrimidine synthesis deficiency in K12 strains of *E. coli* (37) (designated MG1655*). Because this deletion occurs frequently during experimental evolution of *E. coli* K12 strains (38–40), we introduced it at the outset so that it would not occur during selection and lead to artifactual enrichment of clones. Genome editing was accomplished by MAD7™, a Type V CRISPR nuclease, and the lambda Red proteins in concert with a library of editing plasmids targeting every codon of the 30 genes. (Quality metrics of the edited clone library are provided in *SI Appendix*, Table S1.) A barcode in each editing plasmid enabled quantification of variants during selection.

We grew triplicate cultures of the edited clone library in 50 mM acetate for 15 to 20 population doublings. Carbenicillin was included to ensure retention of the barcode plasmid. We passaged the cultures at mid-log phase to maintain a constant selection pressure. We isolated plasmids from aliquots of each culture at the beginning of the experiment and at each passage (approximately four population doublings) and sequenced the barcodes. We quantified the change in abundance of each barcode by fitting the data to Eq. **1**, where *barcode* is the read count for the barcode and *total* is the total number of reads beginning at the first passage (T_1). We define the slope of the line relating the natural log of the fractional barcode abundance to the number of population doublings as the "competition coefficient" c (Dataset S1). (Barcodes with low read counts were removed from the analysis.) The initial period of growth from T_0 to T_1 allowed the cells to acclimate to the selection conditions. In a few cases, edits were so detrimental during acclimation that the barcode was depleted by T_1. In these cases, we estimated c based on the change between T_0 and T_1. These values may or may not accurately reflect the effect of the edits after acclimation. These edits are flagged in Dataset S1. Replicate data for competition coefficients were well-correlated (*SI Appendix*, Fig. S1).

$$\ln\left(\frac{barcode_t}{total_t}\right) = \ln\left(\frac{barcode_{T1}}{total_{T1}}\right) + c(population\ \ doublings) \quad [1]$$

Fig. 2*A* shows the distribution of competition coefficients for 47,218 edits in 30 genes measured during selection in 50 mM acetate. [Edits with low read counts (2,884) were removed from the set of 50,102 edits.] Each value of c depends upon the effect of the edit, the efficiency of genome editing (on average about 20% for introduction of complete edits), and the effects of any unintended mutations introduced during genome editing. Thus, c values should not be viewed as equivalent to selection coefficients.

Concerns have been raised about unintended mutations during CRISPR-Cas9 genome editing (41–43). To determine whether the MAD7™-based editing protocol used to generate our library also introduced unintended mutations, we sequenced the genomes of 30 clones from the library. Only two had a single unintended

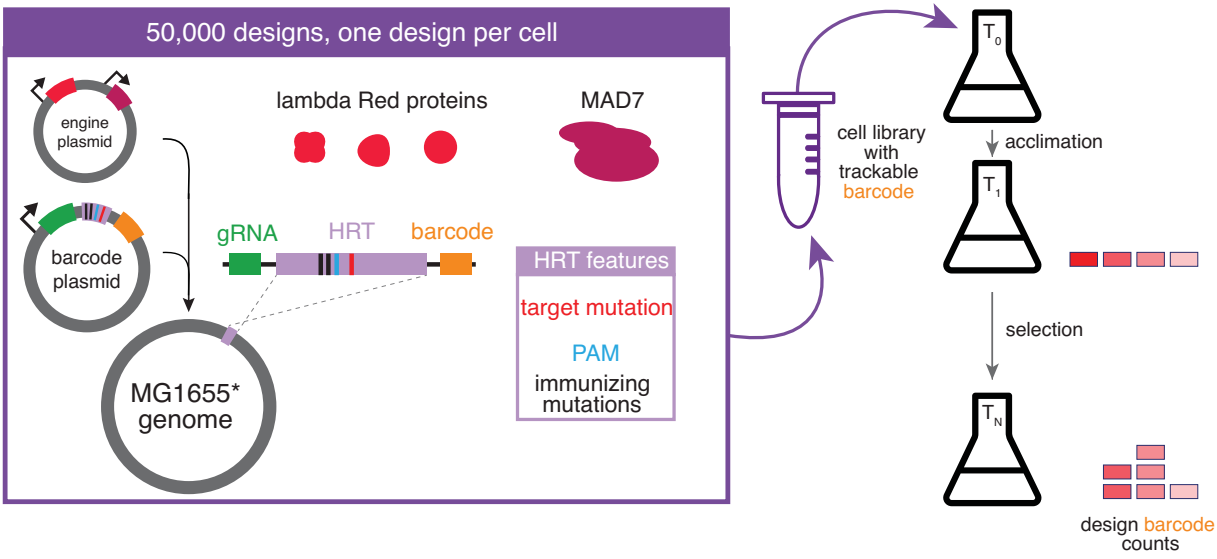

Fig. 1. CRISPR-based genome editing protocol. The engine plasmid encodes the lambda Red proteins and the MAD7™ CRISPR nuclease. The barcode plasmid encodes a gRNA, an HRT designed to introduce a genomic edit and a barcode. Immunizing edits introduced near the PAM site prevent cleavage of the genome after the desired edit is introduced.

Downloaded from https://www.pnas.org by 137.75.80.24 on March 18, 2025 from IP address 137.75.80.24.

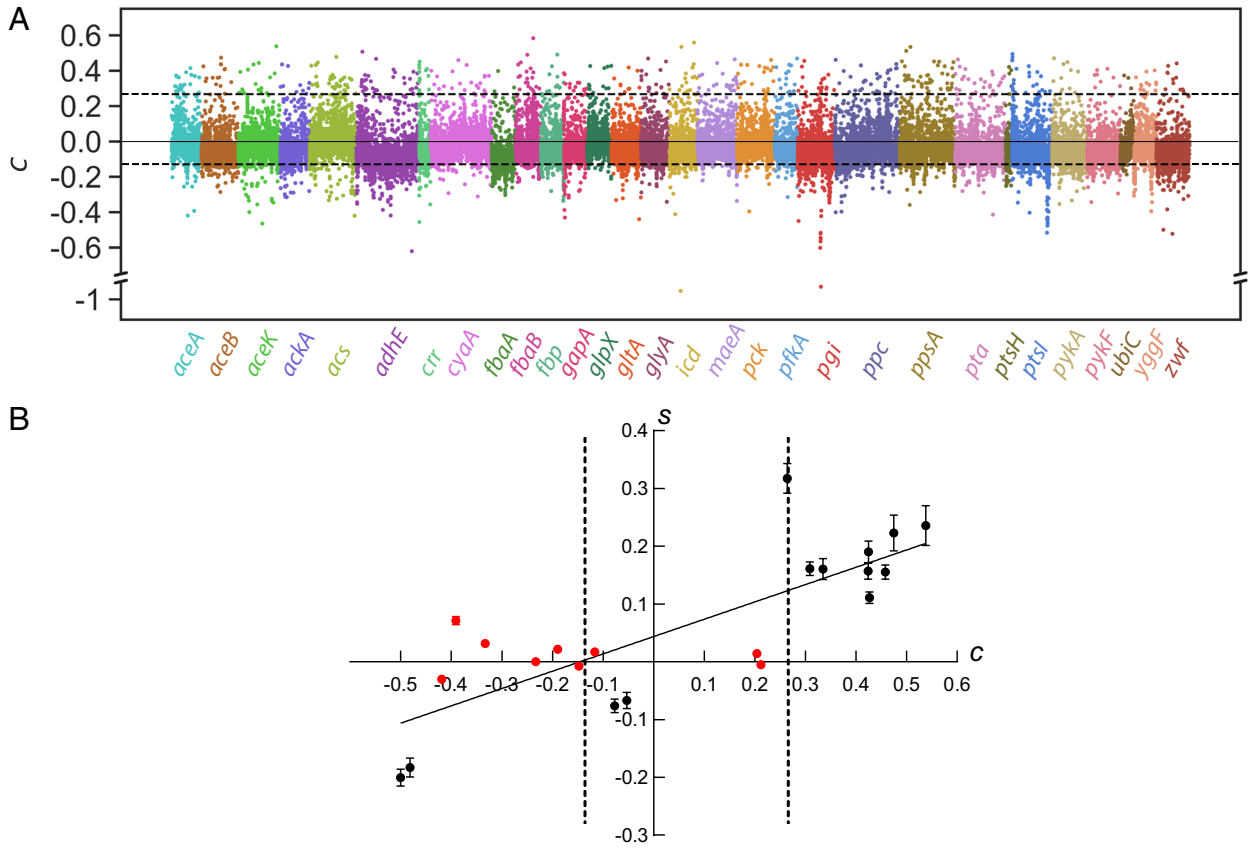

Fig. 2. Effects of 47,218 edits in 30 *E. coli* genes involved in central carbon metabolism. (*A*) Competition coefficients (*c*) determined by following barcode abundances during growth of the library in 50 mM acetate. (*B*) Correlation between competition coefficients and selection coefficients (*s*) for validated edits obtained by direct measurement of growth rates of the parental strain and reconstructed edited clones. Black and red indicate edits whose fitness effects were validated and not validated in reconstructed strains, respectively. The solid line is a fit to all data points ($y = 0.30x + 0.04$, $r^2 = 0.64$). Error bars represent SDs. Dashed lines in both panels encompass 98% of the values for the control gene *glpX*.

mutation. Thus, on the order of 7% of cells with each barcode will have unintended mutations elsewhere in the genome, and possibly different unintended mutations, because hundreds to thousands of cells containing the same barcode plasmid were edited independently. Although the incidence of unintended mutations is low, particularly high-impact mutations may contribute to the effects of genomic edits. An additional complexity arises because we expect the efficiency of editing to vary among designs due to variable gRNA efficiency (44–46). Consequently, we have not attempted a rigorous statistical analysis of the distribution of *c* values. Rather, we have taken a conservative approach to identifying the highest-impact edits by using *c* values for a control gene (*glpX*) as an indication of the range of effects due to unintended mutations and experimental noise. *glpX* is part of the glycerol utilization regulon and encodes fructose 1,6-bisphosphatase 2, one of three fructose 1,6-bisphosphatases in *E. coli*. This regulon is expressed in the presence of glycerol and the absence of glucose (47). GlpX represents only 0.003% of total detected protein in *E. coli* grown on 0.4% acetate (48). Edits in such a lowly expressed gene would be expected to have minimal fitness effects. Given that expectation, the wide range of *c* values for edits in *glpX* is surprising. *E. coli* grows poorly in 50 mM acetate supplemented with carbenicillin (included to maintain the barcode plasmid). Differences in the timing of the transition to stationary phase among clones may have contributed to experimental noise if the fitness effect of an edit differs between exponential growth and stationary phase. For example, *SI Appendix*, Fig. S2 shows an example of a reconstructed clone that enters stationary phase unexpectedly early relative to wild-type cells. Clones like this would have been entering stationary phase while the bulk population was still growing exponentially. In addition, interactions between the vast number of clones growing under suboptimal growth conditions may also have contributed to the experimental noise. Notably, much less noise was observed when the edited library was grown in glucose, a preferred carbon source (see further below).

We have focused on edits for which *c* is outside of 98% of the *c* values for the control gene *glpX*. Using these criteria, we detected 3,907 detrimental edits ($c \leq -0.130$, 8% of the total) and 484 beneficial edits ($c \geq 0.267$, 1% of the total) during selection in 50 mM acetate (*SI Appendix*, Table S2).

Two features of the data shown in Fig. 2*A* suggest that we have detected true biological effects of edits with high values of $|c|$. First, in many cases, high-impact edits (defined as those for which *c* is outside of 98% of the *c* values for *glpX*) are clustered in specific regions of genes. Clusters occur at the 5′ and/or 3′ ends of many genes. Codon choice at the 5′ end of genes can impact translation efficiency by altering secondary structure around the ribosome binding site (16, 19, 49–52) and synonymous mutations at the 3′ end of genes can alter promoters of downstream genes (7–9). Thus, perturbations due to synonymous edits in these regions are not unexpected. Second, the percentage of detrimental edits varies between 1 and 26% between genes (Fig. 3). If the observed effects were due primarily to unintended mutations or experimental noise, we would expect to see similar effects across genes.

A striking feature of the data in Fig. 3 is that synonymous edits in the 30 genes are more likely to be detrimental than beneficial. The preponderance of detrimental edits is consistent with the expectation that millions of years of evolution have optimized codon

Downloaded from https://www.pnas.org by 137.75.80.24 on March 18, 2025 from IP address 137.75.80.24.

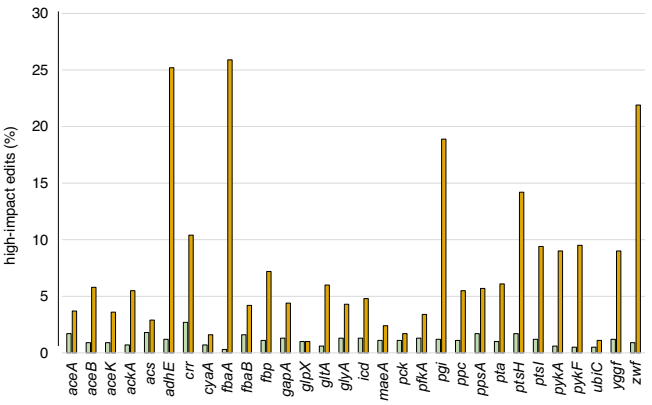

Fig. 3. Percentage of edits in 30 genes that have values of $c < -0.130$ (gold) or $c > 0.267$ (green) during growth of *E. coli* MG1655* in 50 mM acetate. These cutoffs encompass 98% of c values for the control gene *glpX*.

usage in many positions. The average percentage of detrimental edits across all 30 genes was 8.3 ± 6.8 but exceeded 20% in some genes. The differences between the percentage of detrimental edits across the 30 genes may be due to the importance of the encoded proteins during growth in acetate and/or to buffering of the effects of altered protein levels by regulatory mechanisms for some genes.

Twenty-five per cent of edits in *adhE* were detrimental. AdhE (a bifunctional acetaldehyde-CoA dehydrogenase/iron-dependent alcohol dehydrogenase) reduces acetyl CoA to acetaldehyde and then to ethanol. During growth in acetate, acetyl CoA is generated by acetyl CoA synthetase and phosphotransacetylase. When acetate concentrations are high, these reactions may diminish the supply of CoA that is needed for the TCA cycle. Under these circumstances, release of CoA from acetyl CoA by AdhE may help maintain cellular pools of CoA. Synonymous mutations that decrease the level of AdhE might compromise this function.

Twenty-two per cent of edits in *zwf* were detrimental. Zwf ($NADP^+$-dependent glucose-6-phosphate dehydrogenase) shunts glucose 6-phosphate produced by gluconeogenesis into the oxidative branch of the pentose phosphate pathway, the major source of NADPH for biosynthesis. Synonymous mutations that decrease the level of Zwf might hamper myriad biosynthetic pathways for the production of amino acids and fatty acids, as well as folate transformations.

We attempted to validate the fitness effects of 22 edits by measuring the growth rate of reconstructed clones after confirming by whole-genome sequencing that no unintended mutations were present. Growth rates of reconstructed strains (3 to 6 biological replicates) and the parental strain MG1655* (16 biological replicates) were measured during early log phase in a 16-vial turbidostat. When the OD_{600} reached a set value (0.18), a small amount of fresh medium was added to dilute the culture to an OD_{600} of 0.16 and an equivalent volume was diverted to waste. Growth rate (μ) was monitored during 20 to 100 successive cycles of growth and dilution, giving highly accurate values for μ. Selection coefficients (s) were calculated using Eq. **2**.

$$s = 1 - \frac{\mu_{parent}}{\mu_{mutant}} \qquad [2]$$

The effects of suspected high-impact edits were validated in 11 out of 16 cases (69%) for which $|c|$ exceeded the cutoffs for the control gene *glpX* (Fig. 2*B* and *SI Appendix*, Table S3). The effects of three out of six edits for which c fell within the *glpX* cutoffs were also validated, including two for which the values of c and s were very small (e.g., −0.06) and the growth rate of the reconstructed strain was not statistically significantly different from that of the parental strain. (Validated effects include those for high-impact edits for which a statistically significant effect on growth rate was observed in the reconstructed strain and those for low-impact edits when the values of c and s were similar.)

Although we observe a reasonably good correlation between c and s, the s values typically indicate fitness effects that are smaller than expected based upon the c values. This discrepancy is most likely due to the different conditions under which c and s values were measured. c values were determined from selection experiments carried out with a population of an enormous number of clones, some of which had unintended mutations, in flasks. s values were determined using pure cultures of reconstructed strains free of unintended mutations in a turbidostat (which was not available at the time the original selection experiment was carried out).

Immunizing Edits Alone Can Affect Fitness of *E. coli* in 50 mM Acetate. The previous analysis addresses a typical situation in which a large number of targeted mutations are introduced into the *E. coli* genome to enable high-throughput selection or screening, although in most cases, the targeted mutations will be nonsynonymous rather than synonymous mutations. The observation that thousands of edits composed of only synonymous mutations had substantial fitness effects prompted us to delve deeper into the effects of immunizing synonymous mutations alone. For each target codon in our library, we had included an edit design with the wild-type sequence at the target codon but the same set of immunizing mutations used to change the target codon to all other synonymous codons. Since most sets of immunizing mutations were used to edit multiple nearby positions, many edit designs had different barcodes but were actually identical. At the low end, 2 barcodes had identical mutations, providing 6 values of c from the three replicate selection cultures. At the high end, 34 barcodes had identical mutations, providing 102 values of c from the three replicate selection cultures. The additional replicates reduce the experimental noise considerably (Fig. 4). c values for 373 edits are outside of the range of c values for edits in the control gene *glpX* (*SI Appendix*, Table S4). Thus, 22% of the sets of immunizing edits have fitness effects that could either diminish or augment the fitness effects of target mutations. The increased percentage of high-impact edits among immunizing mutation-only edits (22%) relative to the percentage of high-impact edits (9%) in the full set of edits is likely due to the decreased noise, which makes it easier to discriminate between signal and noise.

The Impact of Synonymous Edits Depends On Growth Conditions.

We next investigated whether the fitness consequences of edits varied across different growth conditions. We assessed the impact of the 50,102 edits during growth in 200 mM acetate and 0.2% (11 mM) glucose (Fig. 5 *A* and *B*, Dataset S1, and *SI Appendix*, Tables S5 and S6).

Growth in 200 mM acetate is more challenging than growth in 50 mM acetate because high levels of acetate are toxic (*SI Appendix*, Fig. S3). Nevertheless, the percentages of high-impact edits are similar for most genes during growth in 50 and 200 mM acetate (Fig. 6). However, there are a few striking outliers. The percentage of detrimental edits in *pgi* increases from 19% in 50 mM acetate to 36% in 200 mM acetate (Fig. 6*A* and *SI Appendix*, Table S5). *pgi* encodes phosphoglucose isomerase, which interconverts fructose 6-phosphate and glucose 6-phosphate. Fructose 6-phosphate sits at an important branch point in the metabolic network (*SI Appendix*,

Downloaded from https://www.pnas.org by 137.75.80.24 on March 18, 2025 from IP address 137.75.80.24.

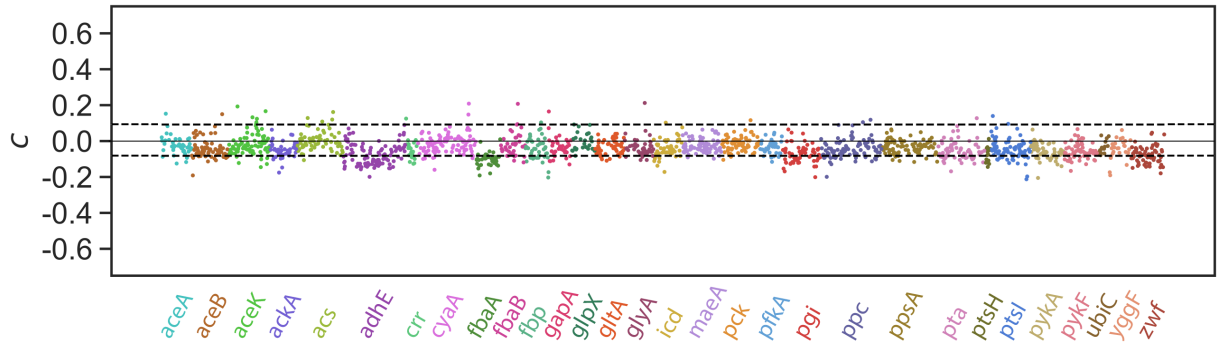

Fig. 4. Competition coefficients for 1,687 sets of immunizing mutations in 30 *E. coli* genes involved in central carbon metabolism.

Fig. S4). Fructose 6-phosphate must be converted to glucose 6-phosphate by Pgi during gluconeogenic growth on acetate to feed the oxidative pentose phosphate pathway and to provide precursors for synthesis of colanic acid, a negatively charged exopolysaccharide that forms a protective capsule around *E. coli* cells under stressful conditions. Fructose 6-phosphate is also a precursor for synthesis of peptidoglycan and other components of colanic acid. If appropriate partitioning of fructose 6-phosphate between these competing pathways is important for fitness, alterations in the level of Pgi caused by synonymous edits may result in suboptimal flux into one or the other of these pathways.

The percentage of high-impact beneficial edits was uniformly low (<2.7%) across all genes in 50 mM acetate but was increased by more than ninefold (and up to 23-fold) in five genes in 200 mM acetate (Fig. 6*B*). Relatively high percentages of beneficial edits were observed in genes involved in the phosphotransferase system that couples phosphorylation of glucose to its transport across the membrane (*crr, ptsH,* and *ptsI*; 7.4, 39.1, and 5.6%, respectively). PtsI donates a phosphate derived from PEP to the downstream protein PtsH (aka Hpr) (Fig. 7*A*). The phosphate is then transferred to Crr and ultimately to glucose imported across the membrane by EIIC and EIIB. Both PtsH and Crr serve regulatory functions that depend upon phosphorylation state; phosphorylation increases when phosphate cannot be delivered to glucose (e.g., during growth in acetate) (53). A high percentage (19%) of beneficial edits is also observed in *cyaA* during growth in 200 mM acetate. CyaA (adenylate cyclase) produces cAMP, which, in concert with the transcriptional regulator CRP, controls expression of hundreds of genes required for growth in the absence of sugars. The high percentages of beneficial edits in these genes suggest that the regulatory system of *E. coli* does not function optimally in unusual and toxic conditions like 200 mM acetate and, consequently, fitness can be improved by synonymous mutations that alter expression of key regulatory proteins.

The effects of synonymous edits are much smaller during growth in glucose than in acetate. The median *c* for all 30 genes was close to 0, so choosing data for a single control gene to approximate the impact of unintended mutations does not seem justified. Consequently, we have focused upon edits with *c* in the bottom and top 1% of all *c* values ($c \leq -0.052$ and $c \geq 0.043$, respectively).

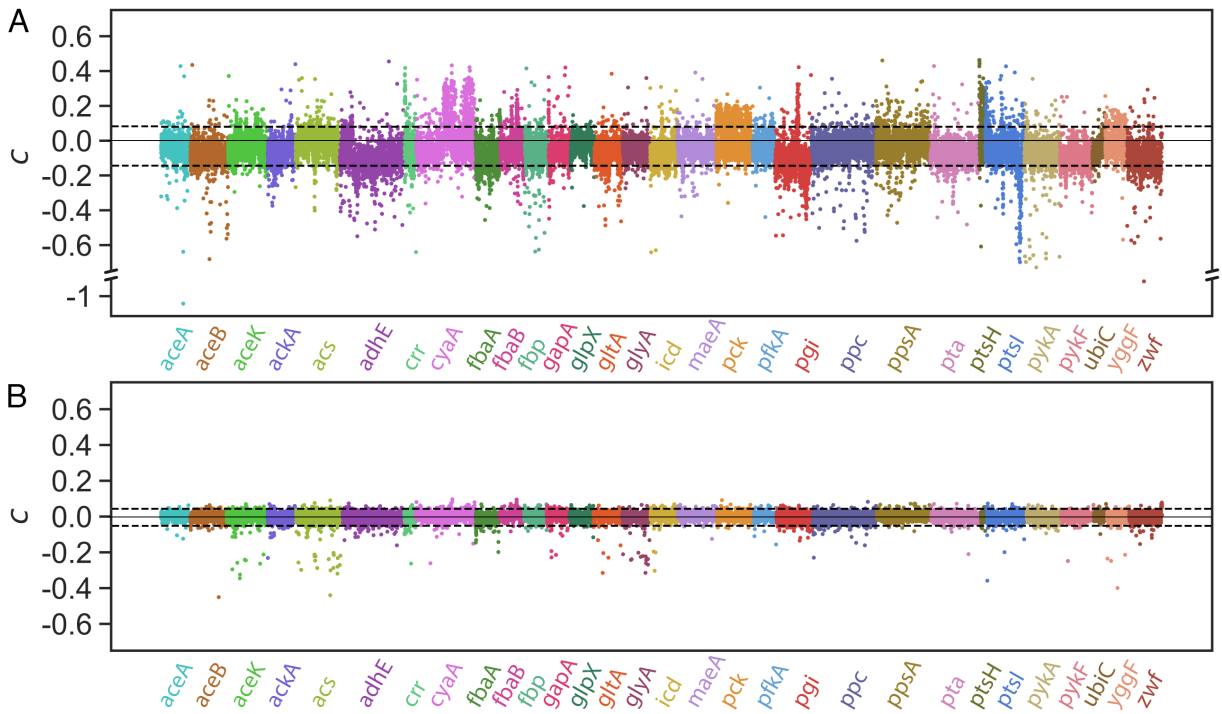

Fig. 5. The impact of synonymous edits depends upon growth conditions. (*A*) *c* values for 46,759 edits in 200 mM acetate. (*B*) *c* values for 47,393 edits in 11 mM glucose. (*c* values for edits with low read counts were removed from the data set in both cases.) Dashed lines in (*A*) encompass 98% of *c* values for the control gene *glpX* and in (*B*) encompass 98% of *c* values for the entire set of genes.

Downloaded from https://www.pnas.org by 137.75.80.24 on March 18, 2025 from IP address 137.75.80.24.

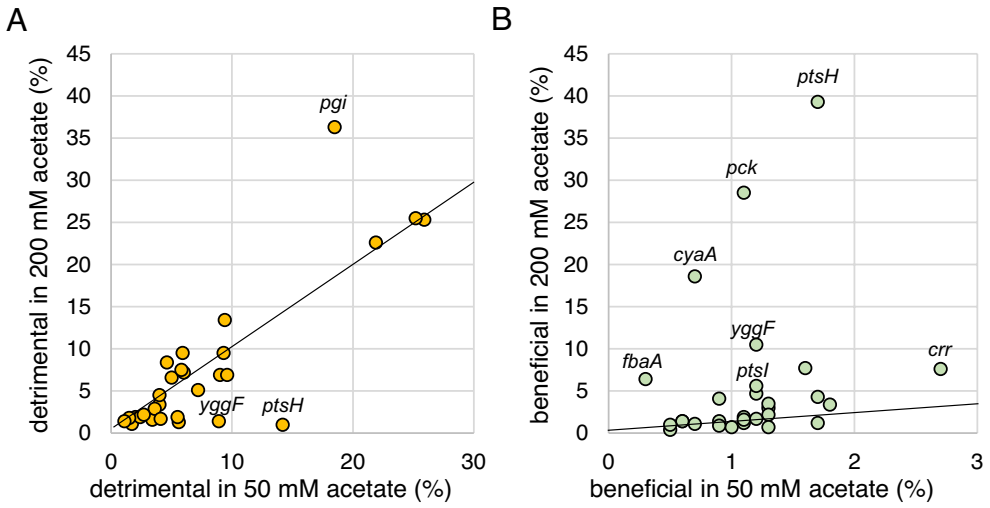

Fig. 6. Comparison of the percentages of high-impact detrimental (*A*) and beneficial (*B*) edits during growth in 50 and 200 mM acetate. (High-impact edits are defined as those with *c* values outside the range of those for 98% of edits in the control gene *glpX*.)

Although the number of high-impact edits is small (by definition 2%, or 946), we again observe that the numbers of beneficial and detrimental edits vary between genes (*SI Appendix*, Table S6). *fbaB* and *cyaA* have a relatively high number of beneficial edits (5% and 3%, respectively), whereas *acs, crr,* and *fbaA* have a relatively high number of detrimental edits (3%). Further, we observe clusters of high-impact edits in specific regions of genes (Fig. 5*B*). We assessed the effects of eight edits in reconstructed clones lacking unintended mutations. The effects of two of the four high-impact edits (defined as those for which $|c|$ was outside of the range of 98% of *c* values in the entire set) were validated (50%) (*SI Appendix*, Table S7). One of the four low-impact edits was also validated; *s* was similar to *c*, although the effect on growth rate was not significant (adj. *P* value = 0.67). Although we tested only a small number of reconstructed strains, these data suggest that as many as 1% of the synonymous edits may have bona fide effects on growth in glucose.

Synonymous Edits in *ptsI* that Decrease the Level of PtsI Impact the Global Transcriptome. We investigated the mechanistic basis of the fitness effects caused by three beneficial edits at the 5′ end of *ptsI* (Fig. 7*B*). We measured the effects of these edits on levels of protein and mRNA using fluorescent reporter assays. Edits 1, 2, and 3 decreased GFP fluorescence when a PtsI-GFP fusion protein was expressed from a plasmid by 96, 37, and 9%, respectively (Fig. 7*C*), likely due to impacts on translation initiation. Codon

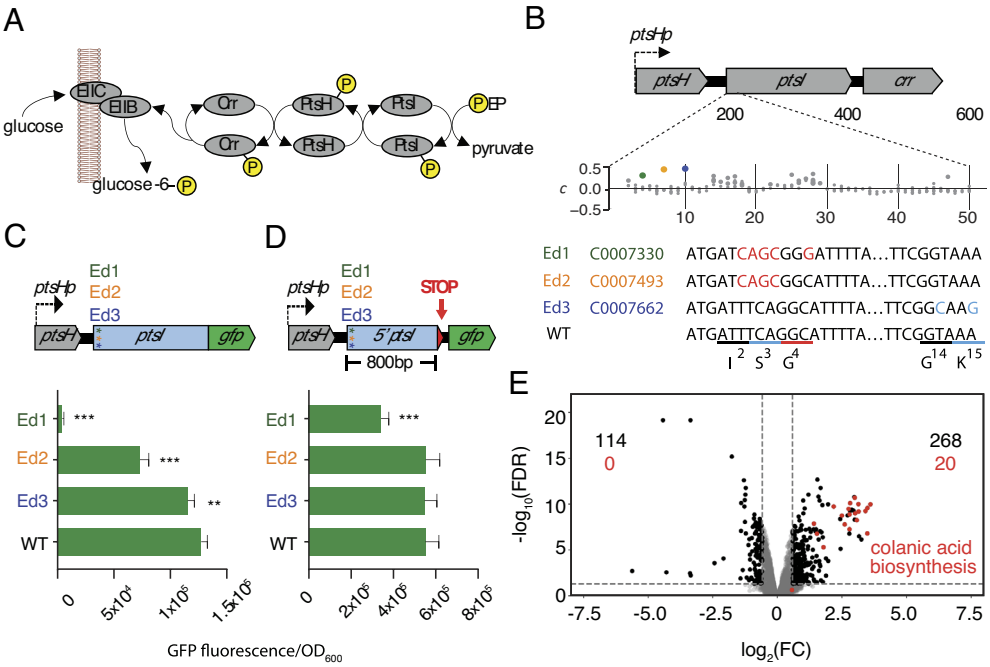

Fig. 7. Effects of synonymous edits near the 5′ end of *ptsI* during growth in 50 mM acetate. (*A*) PtsI is a component of the phosphotransferase system for glucose import. (*B*) The three edits at the 5′ end of *ptsI*. (*C*) Effects of edits 1 to 3 on levels of PtsI-GFP produced from a translational fusion reporter plasmid. (*D*) Effects of edits 1 to 3 on the level of a PtsI-GFP fusion protein produced from a plasmid in which *gfp* was cloned after *ptsHI*. (*E*) Global transcriptional changes caused by edit 1.

Downloaded from https://www.pnas.org by 137.75.80.24 on March 18, 2025 from IP address 137.75.80.24.

choice at the 5′ end of mRNAs is known to influence the efficiency of translation initiation by altering secondary structure around the Shine-Dalgarno sequence and start codon (16, 19, 49–52). Apparently lower levels of PtsI are advantageous during growth in acetate, possibly due to decreased phosphorylation of the downstream proteins PtsH and Crr, which are important for regulation of central carbon metabolism in the absence of glucose. Edit 1 also decreased GFP expression from an operon in which *gfp* was cloned after an 800 bp fragment of *ptsI* (Fig. 7*D*). We attribute the decreased GFP expression to a decrease in mRNA stability. This effect may be due to decreased ribosome occupancy due to impaired translation of *ptsI*, which may leave the mRNA more exposed to nucleases (18).

We examined the global effects of edit 1 on the transcriptome of cells grown in 50 mM acetate (Fig. 7*E*). Edit 1 caused significant changes (FDR ≤ 0.05, $\log_2$FC ≥ ±1.5) in expression of 382 genes (Dataset S2). The large number of transcriptional changes suggests that the beneficial effects of decreased PtsI expression could result from alteration of multiple cellular processes. The 8- to 12-fold increase in expression of genes encoding colanic acid synthesis proteins stands out. Colanic acid is associated with protection against a variety of stresses (54, 55) and may protect cells from toxicity caused by high concentrations of acetate and/or carbenicillin in the medium. The impact of Edit 1 on colanic acid biosynthesis is likely indirect. Transcription of colanic acid biosynthesis genes is activated by RcsA in conjunction with the response regulator RcsB of the two-component system RcsCB. Transcription of *rcsA* is activated by GadE, and transcription of *gadE* in turn is controlled by a host of negative and positive regulators, among them cAMP-CRP. Here lies the connection to PtsI. cAMP levels are impacted by the level of phosphorylated Crr. If the decreased level of PtsI caused by edit 1 decreases the level of phosphorylated Crr, then activation of adenylate cyclase will be reduced and cAMP levels will fall. The resulting decrease in cAMP-CRP should lead to derepression of *gadE,* increased expression of *rcsA,* and increased expression of colanic acid synthesis genes.

Discussion

Thousands of synonymous edits in 30 *E. coli* genes have substantial effects on growth in 50 mM acetate, a sub-optimal carbon source. Given the experimental noise in the data and the inevitability of unintended mutations in approximately 7% of the cells, it was imperative for us to validate the high-throughput selection data by reconstructing individual clones. The fitness effects of 11 out 16 high-impact edits (69%) were validated by measuring the growth rate of reconstructed strains with no unintended mutations.

The effects of synonymous edits during growth in 50 mM acetate are primarily detrimental. Several genes encoding enzymes involved in glycolysis, gluconeogenesis and the pentose phosphate pathway had surprisingly high percentages of high-impact detrimental edits, up to 26% in the case of *fbaA*. Genes encoding proteins that play important roles in regulating central carbon metabolism (*crr* and *ptsH*) also had high percentages of high-impact detrimental edits (10.4 and 14.2%, respectively). These findings suggest that perturbations of protein levels due to synonymous edits tend to affect fluxes in central carbon metabolism during growth on acetate in ways that slow growth. This finding is consistent with the general expectation that codon choice is fairly well optimized in *E. coli.* However, *E. coli* does not typically encounter 50 mM acetate as a sole carbon source, so codon choice has not evolved to be optimal at every position under these conditions. Indeed, we observed 484 high-impact beneficial edits for which one or more synonymous mutations may tune protein expression to a more appropriate level during growth in 50 mM acetate.

The impact of synonymous edits depended strongly upon the growth conditions. In contrast to the relatively few beneficial edits in 50 mM acetate, 5% of all edits were beneficial in 200 mM acetate (Fig. 4*A*). The effect of the more stressful high concentration of acetate was particularly marked for *ptsH,* in which 39% of edits had high-impact beneficial effects. In addition, nearly 19% of edits in *cyaA*, which encodes adenylate cyclase, had high-impact beneficial effects. Changes in the levels of regulatory proteins such as PtsH and adenylate cyclase due to synonymous mutations may be particularly important under stressful conditions. We also observed an increase in the percentage of high-impact beneficial edits for several genes that encode enzymes in glycolysis and gluconeogenesis. The most dramatic effect was observed for *pck*, which encodes pyruvate carboxykinase; 29% of edits had high-impact beneficial effects in 200 mM acetate, compared to only 1% in 50 mM acetate. These data suggest that synonymous mutations may play a particularly important role when organisms are exposed to highly stressful conditions.

In contrast to the impact of synonymous edits in acetate, relatively few edits had fitness impacts in glucose. The effects of synonymous mutations on mRNA structure and function are presumably similar in acetate and glucose, with the exception of mutations that affect binding of sRNAs whose abundance is condition-specific. Beneficial effects of synonymous edits may be uncommon in glucose because codon usage is well-optimized for this growth condition. Detrimental effects may also be uncommon because well-evolved regulatory mechanisms can compensate for codon changes that alter mRNA stability or translation efficiency by adjusting transcription to restore normal protein levels. These regulatory mechanisms may be less effective under conditions that *E. coli* does not normally encounter, such as growth on high concentrations of acetate.

The observation that high-impact edits occur at various positions in genes is consistent with the expectation that multiple mechanisms mediate their effects. Thus, our data provide a wealth of opportunities to probe the impact of sequence on mRNA structure and function. Fitness effects due to synonymous mutations near the 5′ end of coding sequences are expected; it is well-established that secondary structure in this region affects the efficiency of translation initiation (19, 50–52). However, the mechanisms responsible for the fitness effects of most of the thousands of high-impact edits are not obvious.

The impacts of synonymous edits can extend beyond just effects on the structure and function of the mRNA encoded by the edited gene. A beneficial edit at the 5′ end of *ptsI* altered transcription of hundreds of genes, suggesting that synonymous edits in genes encoding regulatory proteins may have particularly widespread effects.

From a practical point of view, our demonstration that synonymous edits can have substantial fitness effects suggests that researchers should carefully consider the potential effects of synonymous immunizing mutations introduced by CRISPR-based genome editing. An alternative, although more labor-intensive, approach to genome editing involves a two-step procedure in which an HRT in the first step introduces a target mutation along with immunizing mutations. In the second step, a gRNA directed at the introduced sequence cleaves the modified genome and promotes incorporation of a second HRT containing only the intended mutation. This procedure is not applicable to generation of massively parallel edited libraries, however. The recently developed prime editing procedure (56, 57), which has been adapted for use in *E. coli* (58), also avoids the introduction of immunizing edits. In this procedure, a prime editing gRNA (pegRNA) includes

Downloaded from https://www.pnas.org by 137.75.80.24 on March 18, 2025 from IP address 137.75.80.24.

a sequence encoding the desired mutation appended to a gRNA that directs a Cas9 nickase fused to reverse transcriptase to a target site. Following nicking of the genome, the 3′ end of the nicked strand binds to a primer binding site on the pegRNA that is used by reverse transcriptase to extend the 3′ end using the mutant sequence as a template. The resulting 3′ flap can displace the sequence on the other side of the nick, leading to the incorporation of the mutation after DNA repair. The risk of fitness effects due to immunizing mutations suggests that prime editing is preferable to classical CRISPR-Cas editing for the construction of a small number of edited strains. However, prime editing has not yet been adapted for the preparation of massively parallel edited libraries.

Selection and screening of large libraries are powerful approaches for identifying high-impact mutations that increase or decrease fitness. The Onyx™ technology enables the construction of massively parallel libraries and thus examination of enormous numbers of variants in the proper genomic context in a single experiment. Our results showed that a high percentage of high-impact edits in a library containing over 50,000 variants could be validated by the characterization of individual reconstructed clones. However, the fact that edits consisting of only synonymous mutations can have substantial fitness effects suggests that care must be taken in designing and interpreting the results of libraries generated using CRISPR-Cas genome editing. The uncertainty introduced by immunizing mutations might be ameliorated by restricting immunizing mutations to positions that are least likely to impact fitness. For example, synonymous mutations near the 5′ ends of genes, internal promoters, or sRNA binding sites should be avoided. Clusters of rare codons that may be required for proper protein folding (59) should also be avoided. Minimizing the number of immunizing mutations and using multiple sets of immunizing mutations may also be helpful. Finally, fitness effects should be confirmed by reconstructing and evaluating strains containing mutations of interest in the absence of synonymous immunizing mutations and unintended mutations elsewhere in the genome.

Materials and Methods

Detailed *Materials and Methods* are found in *SI Appendix*.

We used Inscripta's Designer software (development version) to design editing cassettes to introduce all possible synonymous mutations into 30 genes involved in central carbon metabolism (50,102 designs in total) in *E. coli* MG1655*. Each design consisted of a MAD7 gRNA, a homology repair template (HRT) encoding one of the possible synonymous codons at each target position (including the wild-type codon), and a unique DNA barcode. Each HRT also included one to five synonymous immunizing mutations to prevent recutting of the genome by MAD7 after editing. Oligonucleotides consisting of gRNA, HRT and barcode sequences were cloned into a vector backbone conferring carbenicillin resistance. Genome editing was accomplished by a modification of the method of Garst et al. (2) with Inscripta's developmental proprietary protocols and reagents. After editing, the libraries were cured of the temperature-sensitive engine plasmid.

Selection experiments were carried out in triplicate 50 mL cultures containing approximately 2×10^9 cells in M9 medium supplemented with carbenicillin (100 μg/mL) and either 50 mM sodium acetate, 200 mM sodium acetate, or 0.2% glucose (~11 mM). Cells were passaged at mid-log phase 5 to 7 times. Barcode plasmids were isolated at the beginning of each selection and at each passage. Barcodes were amplified, sequenced, and quantified.

Competition coefficients were calculated from barcode sequencing counts using Enrich2, a statistical framework for analysis of deep mutational scanning data (60). The Enrich2 competition coefficients, which are based on log ratios and determined from a plot of log (*edit/total*) against t/t_{max}, were converted to competition coefficients based upon ln ratios (rather than $\log_{10}$ ratios) and number of population doublings (rather than t/t_{max}) (Eq. **1**) to be consistent with existing evolutionary biology literature.

We reconstructed 28 edited clones as described in *SI Appendix* and sequenced their genomes to ensure that unintended mutations had not occurred during genome editing. We measured the growth rates of 3 to 6 replicate cultures of each reconstructed strain and 16 replicated cultures of the parental MG1655* strain in individual vials of a 16-vial FynchBio turbidostat. Each vial contained 20 mL of M9 containing 100 μg/mL carbenicillin and either 50 mM sodium acetate or 0.2% glucose. The cultures were grown with stirring at 37 °C with additional aeration provided by bubbling air into the medium. The turbidostat was set to maintain the cultures within an OD of 0.16 to 0.18 for a minimum of 20 generations or 100 cycles. The change in OD_{600} during each cycle was used to calculate μ. Once the growth rate stabilized, μ values from multiple cycles (typically 40 to 60, but as many as 100 in some cases) were used to calculate the median and SD of μ for each vial.

We cloned wild-type and edited versions of *ptsI* fused with *gfp* into a low-copy plasmid to generate translational reporter strains in which changes in fluorescence reflect the impact of edits on the level of PtsI (Fig. 7*C*). We created additional reporter strains by expressing *gfp* downstream of a truncated version of *ptsI* to *gfp* (Fig. 7*D*) to assess the impacts of edits on the stability of the mRNA. Effects of one edit at the 5′ end of *ptsI* on the global transcriptome were assessed by RNA Seq.

Data, Materials, and Software Availability. All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. This work was funded by National Institute of General Medical Sciences R01GM124365 to S.D.C. We gratefully acknowledge the contributions of Inscripta in library construction, barcode sequencing and data analysis.

Author affiliations: [a]Department of Molecular, Cellular and Developmental Biology and the Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO 80309; and [b]Amyris, Inc., Emeryville, CA 94608

1. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
2. A. D. Garst *et al.*, Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* **35**, 48–55 (2017).
3. T. Jukes, Silent nucleotide substitutions and the molecular evolutionary clock. *Science* **210**, 973–978 (1980).
4. T. Ikemura, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* **151**, 389–409 (1981).
5. R. Grantham, C. Gautier, M. Gouy, M. Jacobzone, R. Mercier, Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.* **9**, r43–74 (1981).
6. M. Conrad, C. Friedlander, M. Goodman, Evidence that natural selection acts on silent mutation. *Biosystems* **16**, 101–111 (1983).
7. J. P. Kershner *et al.*, A synonymous mutation upstream of the gene encoding a weak-link enzyme causes an ultrasensitive response in growth rate. *J. Bacteriol.* **198**, 2853–2863 (2016).
8. H. Ando, T. Miyoshi-Akiyama, S. Watanabe, T. Kirikae, A silent mutation in *mabA* confers isoniazid resistance on *Mycobacterium tuberculosis*. *Mol. Microbiol.* **91**, 538–547 (2014).
9. E. Lebeuf-Taylor, N. McCloskey, S. F. Bailey, A. Hinz, R. Kassen, The distribution of fitness effects among synonymous mutations in a gene under directional selection. *eLife* **8**, e45952 (2019).
10. P. Brest *et al.*, A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat. Genet.* **43**, 242–245 (2011).
11. J. J. Gartner *et al.*, Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13481–13486 (2013).
12. F. Pagani, M. Raponi, F. E. Baralle, Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6368–6372 (2005).
13. W. F. Mueller, L. S. Larsen, A. Garibaldi, G. W. Hatfield, K. J. Hertel, The silent sway of splicing by synonymous substitutions. *J. Biol. Chem.* **290**, 27700–27711 (2015).
14. L. Cartegni, S. L. Chew, A. R. Krainer, Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* **3**, 285–298 (2002).
15. L. B. Ahlborn *et al.*, Identification of a breast cancer family double heterozygote for RAD51C and BRCA2 gene mutations. *Fam. Cancer* **14**, 129–133 (2015).
16. J. Kristofich *et al.*, Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. *PLoS Genet.* **14**, e1007615 (2018).
17. L. R. Rapaport, G. A. Mackie, Influence of translational efficiency on the stability of the mRNA for ribosomal protein S20 in *Escherichia coli*. *J. Bacteriol.* **176**, 992–998 (1994).
18. A. Deana, J. G. Belasco, Lost in translation: The influence of ribosomes on bacterial mRNA decay. *Genes Dev.* **19**, 2526–2533 (2005).
19. G. Kudla, A. W. Murray, D. Tollervey, J. B. Plotkin, Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).
20. K. Bentele, P. Saffert, R. Rauscher, Z. Ignatova, N. Bluthgen, Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* **9**, 675 (2013).
21. W. Gu, T. Zhou, C. O. Wilke, A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput. Biol.* **6**, e1000664 (2010).

Downloaded from https://www.pnas.org by 137.75.80.24 on March 18, 2025 from IP address 137.75.80.24.

22. A. B. Morgenthaler *et al.*, Mutations that improve efficiency of a weak-link enzyme are rare compared to adaptive mutations elsewhere in the genome. *eLife* **8**, e53535 (2019).
23. A. Espah Borujeni, H. M. Salis, Translation initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism. *J. Am. Chem. Soc.* **138**, 7016–7023 (2016).
24. C. H. Yu *et al.*, Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol. Cell* **59**, 744–754 (2015).
25. A. R. Subramaniam *et al.*, A serine sensor for multicellularity in a bacterium. *eLife* **2**, e01501 (2013).
26. C. Kimchi-Sarfaty *et al.*, A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**, 525–528 (2007).
27. M. Zhou *et al.*, Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* **495**, 111–115 (2013).
28. I. M. Walsh, M. A. Bowman, I. F. Soto Santarriaga, A. Rodriguez, P. L. Clark, Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 3528–3534 (2020).
29. M. P. Zwart *et al.*, Unraveling the causes of adaptive benefits of synonymous mutations in TEM-1 beta-lactamase. *Heredity (Edinb)* **121**, 406–421 (2018).
30. S. F. Bailey, A. Hinz, R. Kassen, Adaptive synonymous mutations in an experimentally evolved *Pseudomonas fluorescens* population. *Nat. Commun.* **5**, 4076 (2014).
31. D. H. Lee, B. O. Palsson, Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a nonnative carbon source, L-1,2-propanediol. *Appl. Environ. Microbiol.* **76**, 4158–4168 (2010).
32. D. S. Lawrie, P. W. Messer, R. Hershberg, D. A. Petrov, Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet.* **9**, e1003527 (2013).
33. T. Zhou, W. Gu, C. O. Wilke, Detecting positive and purifying selection at synonymous sites in yeast and worm. *Mol. Biol. Evol.* **27**, 1912–1922 (2010).
34. I. Hellmann *et al.*, Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res.* **13**, 831–837 (2003).
35. Z. E. Sauna, C. Kimchi-Sarfaty, Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* **12**, 683–691 (2011).
36. R. Chen, E. V. Davydov, M. Sirota, A. J. Butte, Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS ONE* **5**, e13574 (2010).
37. K. F. Jensen, The *Escherichia coli* K-12 "wild types" W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. *J. Bacteriol.* **175**, 3401–3407 (1993).
38. D. Blank, L. Wolf, M. Ackermann, O. K. Silander, The predictability of molecular evolution during functional innovation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3044–3049 (2014).
39. T. M. Conrad *et al.*, Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biol.* **10**, R118 (2009).
40. A. Knoppel *et al.*, Genetic adaptation to growth under laboratory conditions in *Escherichia coli* and *Salmonella enterica*. *Front. Microbiol.* **9**, 756 (2018).
41. H. Manghwar *et al.*, CRISPR/Cas systems in genome editing: Methodologies and tools for sgRNA design, off-target evaluation, and strategies to mitigate off-target effects. *Adv. Sci. (Weinh)* **7**, 1902312 (2020).
42. C. Guo, X. Ma, F. Gao, Y. Guo, Off-target effects in CRISPR/Cas9 gene editing. *Front. Bioeng. Biotechnol.* **11**, 1143157 (2023).
43. L. Kruglyak *et al.*, Insufficient evidence for non-neutrality of synonymous mutations. *Nature* **616**, E8–E9 (2023).
44. A. Choudhury *et al.*, Determinants for efficient editing with Cas9-mediated recombineering in *Escherichia coli*. *ACS Synth. Biol.* **9**, 1083–1099 (2020).
45. H. Xu *et al.*, Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, 1147–1157 (2015).
46. S. B. Thyme, L. Akhmetova, T. G. Montague, E. Valen, A. F. Schier, Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat. Commun.* **7**, 11750 (2016).
47. E. C. C. Lin, Ed., *Dissimilatory Pathways for Sugars, Polyols, and Carbohydrates* (American Society for Microbiology, Washington, D.C., 1996), **vol. 1**.
48. M. Mori *et al.*, From coarse to fine: The absolute *Escherichia coli* proteome under diverse growth conditions. *Mol. Syst. Biol.* **17**, e9536 (2021).
49. D. B. Goodman, G. M. Church, S. Kosuri, Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342**, 475–479 (2013).
50. L. M. Munson, G. D. Stormo, R. L. Niece, W. S. Reznikoff, lacZ translation initiation mutations. *J. Mol. Biol.* **177**, 663–683 (1984).
51. M. N. Hall, J. Gabay, M. Debarbouille, M. Schwartz, A role for mRNA secondary structure in the control of translation initiation. *Nature* **295**, 616–618 (1982).
52. M. C. Ganoza, E. C. Kofoid, P. Marliere, B. G. Louis, Potential secondary structure at translation–initiation sites. *Nucleic Acids Res.* **15**, 345–360 (1987).
53. M. Madan Babu, S. A. Teichmann, Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.* **31**, 1234–1244 (2003).
54. Y. Mao, M. P. Doyle, J. Chen, Insertion mutagenesis of *wca* reduces acid and heat tolerance of enterohemorrhagic *Escherichia coli* O157:H7. *J. Bacteriol.* **183**, 3811–3815 (2001).
55. J. Chen, S. M. Lee, Y. Mao, Protective effect of exopolysaccharide colanic acid of *Escherichia coli* O157:H7 to osmotic and oxidative stress. *Int. J. Food Microbiol.* **93**, 281–286 (2004).
56. A. V. Anzalone *et al.*, Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
57. C. Lu *et al.*, Prime editing: An all-rounder for genome editing. *Int. J. Mol. Sci.* **23**, 9862 (2022).
58. Y. Tong, T. S. Jorgensen, C. M. Whitford, T. Weber, S. Y. Lee, A versatile genetic engineering toolkit for *E. coli* based on CRISPR-prime editing. *Nat. Commun.* **12**, 5206 (2021).
59. J. L. Chaney *et al.*, Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput. Biol.* **13**, e1005531 (2017).
60. A. F. Rubin *et al.*, A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* **18**, 150 (2017).

Downloaded from https://www.pnas.org by 137.75.80.24 on March 18, 2025 from IP address 137.75.80.24.