



Published in final edited form as:

Nature. 2015 February 5; 518(7537): 89–93. doi:10.1038/nature14095.

Recoded organisms engineered to depend on synthetic amino acids

Alexis J. Rovner^{1,2}, Adrian D. Haimovich^{1,2,*}, Spencer R. Katz^{1,2,*}, Zhe Li^{1,2}, Michael W. Grome^{1,2}, Brandon M. Gassaway^{2,3}, Miriam Amiram^{1,2}, Jaymin R. Patel^{1,2}, Ryan R. Gallagher^{1,2}, Jesse Rinehart^{2,3}, and Farren J. Isaacs^{1,2}

¹Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520, USA

²Systems Biology Institute, Yale University, West Haven, CT 06516, USA

³Department of Cellular and Molecular Physiology, Yale University, New Haven, CT 06520, USA

Abstract

Genetically modified organisms (GMOs) are increasingly used in research and industrial systems to produce high-value pharmaceuticals, fuels, and chemicals¹. Genetic isolation and intrinsic biocontainment would provide essential biosafety measures to secure these closed systems and enable safe applications of GMOs in open systems^{2,3}, which include bioremediation⁴ and probiotics⁵. Although safeguards have been designed to control cell growth by essential gene regulation⁶, inducible toxin switches⁷, and engineered auxotrophies⁸, these approaches are compromised by cross-feeding of essential metabolites, leaked expression of essential genes, or genetic mutations^{9,10}. Here, we describe the construction of a series of genomically recoded organisms (GROs)¹¹ whose growth is restricted by the expression of multiple essential genes that depend on exogenously supplied synthetic amino acids (sAAs). We introduced a *Methanocaldococcus jannaschii* tRNA:aminoacyl-tRNA synthetase (aaRS) pair into the chromosome of a GRO that lacks all TAG codons and release factor 1, endowing this organism with the orthogonal translational components to convert TAG into a dedicated sense codon for sAAs. Using multiplex automated genome engineering (MAGE)¹², we introduced in-frame TAG codons into 22 essential genes, linking their expression to the incorporation of synthetic phenylalanine-derived amino acids. Of the 60 sAA-dependent variants isolated, a notable strain harboring 3 TAG codons in conserved functional residues¹³ of MurG, DnaA and SerS and containing targeted tRNA deletions maintained robust growth and exhibited undetectable escape frequencies upon culturing $\sim 10^{11}$ cells on solid media for seven days or in liquid media for 20 days. This is a significant improvement over existing biocontainment approaches^{2,3,6-10}. We

Reprints and permissions information is available at www.nature.com/reprints

Correspondence and requests for materials should be addressed to F.J.I. (farren.isaacs@yale.edu).

*These authors contributed equally to this work

Supplementary Information: is linked to the online version of the paper at www.nature.com/nature.

Author Contributions A.J.R. and F.J.I. conceived the study, designed experiments and wrote the paper with assistance from A.D.H. and S.R.K.; A.J.R. conducted experiments with assistance from S.R.K., A.D.H., M.G., M.A., J.R.P. and R.R.G.; B.G. and J.R. conducted mass spectrometry; All authors commented on the paper and F.J.I. supervised all aspects of the study.

A.J.R. and F.J.I. have filed a provisional application with the US Patent and Trademark Office on this work.

constructed synthetic auxotrophs dependent on sAAs that were not rescued by cross-feeding in environmental growth assays. These auxotrophic GROs possess alternate genetic codes that impart genetic isolation by impeding horizontal gene transfer¹¹ and now depend on the use of synthetic biochemical building blocks, advancing orthogonal barriers between engineered organisms and the environment.

The advent of recombinant DNA technologies in the 1970s established genetic cloning methods¹⁴, ushering in the era of biotechnology. Over the past decade, synthetic biology has fueled the emergence of GMOs with increased sophistication as common and valued solutions in clinical, industrial, and environmental settings^{1,4,5}, necessitating the development of safety and security measures first outlined in the 1975 Asilomar conference on recombinant DNA¹⁵. While guidelines for physical containment and safe use of organisms have been widely adopted, intrinsic biocontainment – biological barriers limiting the spread and survival of microorganisms in natural environments – remains a defining challenge. Existing biocontainment strategies employ natural auxotrophies or conditional suicide switches where top safeguards meet the 10⁻⁸ NIH standard¹⁶ for escape frequencies (EFs), but can be compromised by metabolic cross-feeding or genetic mutation^{9,10}. We hypothesized that engineering dependencies on synthetic biochemical building blocks would enhance existing containment strategies by establishing orthogonal barriers not feasible in organisms with a standard genetic code.

Our approach to engineering biocontainment utilized a GRO lacking all instances of the TAG codon and release factor 1 (terminates translation at UAA and UAG), eliminating termination of translation at UAG and endowing the organism with increased viral resistance, a common form of horizontal gene transfer (HGT). The TAG codon was then converted to a sense codon through the introduction of an orthogonal translation system (OTS) containing an aaRS:tRNA pair, permitting site-specific incorporation of sAAs into proteins without impairing cellular fitness¹¹. Leveraging these unique properties of the GRO, we sought to reintroduce the TAG codon into essential genes to restrict growth to defined media containing sAAs. We also eliminated the use of multi-copy plasmids, which reduce viability and growth¹⁷, impose biosynthetic burden, persist poorly in host cells over time¹⁸, and increase the risk of acquiring genetic escape mutants (EMs)³, by manipulating native chromosomal essential genes and integrating the OTS into the genome. To engineer synthetic auxotrophies, we chose essential genes of varying expression levels (Methods), many of whose functions (*e.g.*, replication, translation) cannot be complemented by cross-feeding of metabolites. Genes dispersed throughout the genome were selected to prevent a single HGT event from compromising containment.

We pursued three strategies to engineer dependence on nontoxic, membrane permeable, and well-characterized sAAs through the introduction of TAG codons into essential genes: (1) insertion at the amino-terminus, (2) substitution of residues with computationally predicted tolerances¹⁹, and (3) substitution of conserved¹³ residues at functional sites (Fig. 1a). We initially pursued the first two strategies in a GRO containing an OTS optimized for the sAA p-acetyl-L-phenylalanine (pAcF, α). Using MAGE¹², we targeted 155 codons for TAG incorporation via four pools of oligonucleotides (Supplementary Table 1, Supplementary

Table 2) in permissive media containing pAcF and L-arabinose (aaRS induction) (Fig. 1b). After replica plating on nonpermissive media lacking pAcF and L-arabinose, we isolated eight pAcF auxotrophs with one strain containing two TAGs in essential genes (Fig. 1c and Supplementary Table 3). To determine whether our strategy was capable of creating synthetic auxotrophs dependent on other sAAs, MAGE was used to mutagenize annotated residues in the sAA binding pocket of the pAcF aaRS (Supplementary Table 4) to accommodate p-iodo-L-phenylalanine (pIF, β) or p-azido-L-phenylalanine (pAzF, γ) in two strains. After MAGE-based incorporation of TAGs and selections on permissive and nonpermissive solid media, we obtained eight pIF and 23 pAzF auxotrophs harboring one to four TAGs at 30 distinct loci across 20 essential genes (Supplementary Table 3, Supplementary Table 5). Together, these data demonstrate the modularity of our approach and that synthetic auxotrophs can be engineered across many essential genes using multiple sAAs (Extended Data Fig. 1).

Doubling time (DT) measurements in permissive media revealed minimal or no fitness impairment of synthetic auxotrophs relative to their non-contained ancestors with a genomically integrated OTS (Fig. 2a and Supplementary Table 6). To quantify the degree of containment, we measured the ratio of colony forming units (CFUs) on nonpermissive to permissive solid media and observed a range of EFs spanning 10^{-3} to 10^{-7} (Fig. 2b). One notable strain DnaX.Y113 α , preserved the DT of its non-contained ancestor (Fig. 2a) while maintaining an EF of 6.7×10^{-7} (Fig. 2b). We directly investigated pAcF incorporation in DnaX.Y113 α using mass spectrometry and identified peptides containing pAcF at Y113 (Fig. 2c).

To investigate escape mechanisms of EMs derived from synthetic auxotrophs with one essential TAG codon, we performed targeted sequencing and observed transition mutations (A•T→G•C and G•C→A•T) commonly observed in mismatch repair-deficient strains (*mutS*)²⁰. All isolated DnaX.Y113 α EMs incorporate tryptophan by mutation of the TAG codon to TGG. SecY.Y122 α EMs incorporate glutamine by mutation of *glnV* to form a glutamine amber suppressor or mutation of the *secY.Y122* TAG codon to CAG (Supplementary Table 7). One of three SecY.Y122 α EMs was wild type at the *secY.Y122* TAG codon and putative amber suppressor loci²¹, but whole genome sequencing (Supplementary Table 8) revealed a Q54D missense mutation in *rpsD* (30S ribosomal subunit S4). This site is implicated in ribosome fidelity^{22,23} and is the causal mutation leading to escape in this mutant (Extended Data Fig. 2).

These escape mechanisms informed two sets of experiments to engineer strains with lower EFs. First, we sought to create synthetic auxotrophs with increased numbers of TAGs (Fig. 2d) by combining TAGs from strains possessing the lowest EFs (*i.e.*, *dnaX.Y113*, *lspA.Y54*, and *secY.Y122*) into a single strain. In strains containing two TAGs, the EF was reduced to 1.4×10^{-7} (rEc. γ .dB.26) and 1.4×10^{-8} (rEc. β .dB.9) (strain annotations are listed in Supplementary Table 6). In strains containing three TAGs, EFs were further reduced to 5.0×10^{-9} (rEc. β .dC.11) and 4.7×10^{-9} (rEc. β .dC.12). We used MAGE to quantitatively assess the effects of non-synonymous mutations at individual TAG codons in strains incorporating pIF at SecY.Y122, DnaX.Y113, and LspA.Y54 by mutating the TAG site to sense codons for all 20 natural amino acids. Strains containing multiple TAGs were less likely to survive

when one TAG was compromised (Extended Data Fig. 3). In a second set of experiments, we restored *mutS* and observed a decreased EF in strains by 1.5- to 3.5-fold (Fig. 2e, Supplementary Table 6). EMs derived from *mutS*⁺ higher-order TAG strains exhibited impaired fitness with 1.14- to 1.28-fold greater DTs than their contained ancestors. Whole genome sequencing was performed on these EMs and revealed mutations of tyrosine tRNAs to form tyrosine amber (UAG) or ochre (UAA) suppressors (Supplementary Table 9).

To reduce EFs below $\sim 10^{-9}$ and eliminate rescue by natural amino acids, we pursued a third strategy to replace conserved and functional residues in essential proteins with sAAs (Fig. 1a). Using the Conserved Domain Database¹³, we searched all essential proteins for tyrosine, tryptophan, and phenylalanine residues involved in protein-protein interactions (e.g., dimerization) or located within active sites to identify candidates suitable for replacement with phenylalanine-derived sAAs. After targeted insertion of TAG codons using MAGE, we isolated four synthetic auxotrophs with pAzF incorporated at GlyQ.Y226 (glycyl-tRNA synthetase α subunit, dimer interface), Lnt.Y388 (apolipoprotein N-acyltransferase, active site), MurG.F243 (N-acetylglucosaminyl transferase, active site), and DnaA.W6 (chromosomal replication initiator protein, oligomerization site²⁴) in strains with minor fitness impairments (Fig. 3a) and EFs spanning 10^{-5} to 10^{-7} (Fig. 3b). Identical experiments to incorporate pAcF and pIF failed to generate synthetic auxotrophs, suggesting that the targeted residues are recalcitrant to replacement by pAcF and pIF. Since pAzF was able to replace conserved and functional tyrosine, phenylalanine and tryptophan residues across several essential proteins, we hypothesized that engineering strains to contain higher-order TAG combinations would limit escape by mutations that cause incorporation of natural amino acids at multiple TAG codons. EFs of 1.6×10^{-9} and 2.3×10^{-9} were observed for strains containing two TAG codons: rEc. γ .dB.41 (DnaA.W6 and MurG.F243) and rEc. γ .dB.43 (DnaA.W6 and SerS.F213), respectively. Upon restoring *mutS*, the EF of rEc. γ .dB.41' fell to 6.0×10^{-10} (Fig. 3b). Merging all three sites into one strain (rEc. γ .dC.46) and its *mutS*⁺ derivative (rEc. γ .dC.46') led to EFs of $< 7.9 \times 10^{-11}$ and $< 4.4 \times 10^{-11}$ (below the detection limit of our plate-based assay), respectively (Supplementary Table 6).

Temporal monitoring of rEc. γ .dC.46' revealed the emergence of growth-impaired EMs two days post-plating on nonpermissive solid media (Fig. 3d). Sequencing of EMs derived from strains rEc. γ .dC.41' and rEc. γ .dC.46' revealed amber suppressor-forming mutations at one of three tyrosine tRNAs (*tyrT*, *tyrV*, *tyrU*) with growth impairments spanning 1.61- to 2.10-fold increase in DT relative to their contained ancestors (Supplementary Table 9, Supplementary Table 10). Given that *E. coli* contains three tyrosine tRNAs, we hypothesized that deletion of *tyrT* and *tyrV*²⁵ would prevent acquisition of amber suppressor-forming mutations at *tyrU*, as preservation of this single remaining copy of tRNA^{Tyr} would be required to maintain fidelity of tyrosine incorporation during protein synthesis. We used λ -Red recombination to delete *tyrT* and *tyrV* in rEc. β .dC.12', rEc. β .dC.12'.E7 (EM of rEc. β .dC.12'), and rEc. γ .dC.46' with a chloramphenicol resistance gene. Deletion of *tyrT* and *tyrV* restored containment of the EM, establishing the causal escape mechanism (Extended Data Fig. 4). Moreover, *tyrT/V* deletions in rEc. β .dC.12'. tY and rEc. γ .dC.46'. tY decreased EFs below detectable levels ($< 4.9 \times 10^{-12}$ and $< 6.3 \times 10^{-12}$, respectively) over the seven-day observation period (Fig. 3d, Supplementary Table 11).

To challenge strains rEc.β.dC.12'. tY and rEc.γ.dC.46'. tY with natural amino acids and mimic a potential HGT event, we introduced constructs containing phenylalanine or tryptophan amber suppressor tRNAs. While growth of suppressor-containing strains was equivalent to the cognate contained ancestor in permissive liquid media, severely impaired or no growth was observed in nonpermissive media (Extended Data Fig. 5). Such findings are further supported in experiments where a large ($\sim 10^{11}$) inoculum of cells challenged on solid or in liquid (see below) nonpermissive media do not yield EMs, providing ample opportunity for natural formation of a phenylalanine amber suppressor via mutation of one of two native copies of tRNA^{Phe}. These data support our hypothesis that synthetic auxotrophs containing higher-order TAG combinations depend on the sAA and limit growth from natural amino acids.

To interrogate the long-term stability of synthetic auxotrophs where EM formation is not limited by a colony growth environment, temporal monitoring of rEc.γ.dC.46'. tY was performed on large cell populations in liquid culture (1L of LB media) for seven days with frequent OD₆₀₀ measurements to track cell growth (Fig. 3e). Inoculation of $\sim 10^{11}$ cells in permissive media led to a confluent culture of contained cells within 24 hours. Inoculation of $\sim 10^9$, $\sim 10^{10}$, and $\sim 10^{11}$ cells into nonpermissive media revealed transient growth, which we hypothesize is due to residual pAzF and L-arabinose from large inoculums, followed by a sustained decrease in cell density and growth termination. Cell survival and escape from liquid cultures was monitored by quantifying CFUs on permissive and nonpermissive solid media, respectively (Fig. 3f). Plating on permissive solid media revealed a drop in CFUs to below the limit of detection within one day from the nonpermissive flask inoculated with $\sim 10^9$ cells and three days from nonpermissive flasks inoculated with $\sim 10^{10}$ or $\sim 10^{11}$ cells (Fig. 3f). No CFUs were observed from any culture plated on nonpermissive solid media. To confirm the absence of a single EM following an extended 20-day growth period (Extended Data Fig. 6), the nonpermissive and permissive cultures inoculated with $\sim 10^{11}$ cells were plated across 30 nonpermissive plates. EMs were not observed and EFs remained below the detection limit after seven days, which is comparable to the solid media results. These results demonstrate that rEc.γ.dC.46'. tY depends on pAzF, maintains long-term stability of biocontainment in permissive liquid media and exhibits termination of growth in nonpermissive media.

To further investigate the dependency to sAAs, liquid growth profiles were collected for synthetic auxotrophs across sAA and L-arabinose concentration gradients. Growth of rEc.γ.dC.46'. tY was not observed below 0.002% L-arabinose and 0.5 mM pAzF (Fig. 3g, Extended Data Fig. 7). Growth increased in a dose-dependent manner with increasing concentrations of pAzF and L-arabinose, where 5 mM pAzF and 0.2% L-arabinose was optimal for fitness (*i.e.*, maximum OD₆₀₀ and minimum DT). In an equivalent experiment with rEc.β.dC.12'. tY, 1 mM pIF and 0.2% L-arabinose was optimal for fitness (Extended Data Fig. 8). Since growth was not observed in media lacking either L-arabinose or the sAA, these data further support the dependency of synthetic auxotrophs on sAAs.

To determine whether a synthetic auxotroph could be rescued by metabolic cross-feeding, we evaluated the viability of strains on diverse media types. We grew wild-type *MG1655 E. coli*, a biotin auxotroph (EcNR2¹²), a non-contained GRO (rEc.γ), and the pAzF synthetic

auxotroph (rEc.γ.dC.46') on solid media containing both pAzF/L-arabinose and biotin, either pAzF/L-arabinose or biotin, and on plates lacking small molecules (Fig. 4). Despite biotin auxotrophy, growth of EcNR2 on rich defined media without biotin was rescued in close proximity to wild-type *E. coli*, suggesting cross-feeding of essential metabolites (Extended Data Fig. 9). Blood agar and soil extracts without biotin or pAzF/L-arabinose supplementation supported growth of all strains except the synthetic auxotroph, which only grew on media supplemented with pAzF and L-arabinose. These data suggest that synthetic auxotrophies could lead to a more viable containment strategy for clinical (*e.g.*, blood) and environmental (*e.g.*, soil) settings, where metabolic auxotrophies can be overcome by proximal, metabolically competent strains.

Synthetic auxotrophs utilize unnatural biochemical building blocks necessary for essential proteins with activities that cannot be complemented by naturally occurring small molecules. We have previously shown that genomic recoding interferes with HGT from viruses¹¹ and have now extended orthogonal barriers by engineering two synthetic auxotrophs using two unique sAAs that exhibit EFs below our detection limit ($<6.3 \times 10^{-12}$). These synthetic auxotrophs possess three essential TAGs at loci dispersed throughout the genome (0.84, 0.86, and 2.9 MB apart), thereby limiting the likelihood that a single HGT event could compromise containment (Extended Data Fig. 1). These orthogonal barriers can be expanded further by incorporation of additional TAG sense codons across more than three essential genes, but will likely require concurrent advances in OTS performance to maintain fitness and viability (*e.g.*, enhanced activity and specificity of aaRS:tRNA pairs²⁶). Our modular approach to biocontainment limits growth to synthetic environments containing unnatural biochemical building blocks with diverse chemistries. We anticipate that further genome recoding efforts^{11,27} will enable auxotrophies for multiple sAAs that could be enhanced by other synthetic components including unnatural nucleotides and extended genetic alphabets²⁸⁻³⁰. Orthogonal biological systems employing multi-level containment mechanisms are uniquely suited to provide safe GMOs for clinical, environmental, and industrial applications¹.

Despite the breadth of genomic diversity found in nature, all species utilize the same biochemical foundation to sustain life. The semantic architecture of the GRO employs orthogonal translational components, establishing the basis for a synthetic molecular language that relieves limitations on natural biological functions by depending on the incorporation of sAAs with exotic chemistries. This work sets the stage for future experiments to probe the optimality of the natural genetic code and to explore the plasticity of proteins and whole organisms capable of sampling new evolutionary landscapes.

Methods

Reagents

Oligonucleotide synthesis was performed by Integrated DNA Technologies (IDT) and Keck Foundation Biotechnology Resource Laboratory at Yale University (Supplementary Table 1). Unless otherwise stated, all cultures were grown in LB medium. The following selective agents and inducers were used at the specified concentrations: ampicillin (amp, 50 µg/mL), carbenicillin (carb, 50 µg/mL), zeocin (zeo, 10 µg/ml), spectinomycin (spec, 95 µg/mL) and

sodium dodecyl sulfate (SDS, 0.005% w/v), isopropyl β -D-1-thiogalactopyranoside (IPTG, 100 μ M), 5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside (X-Gal, 40 μ g/mL), and L-arabinose (ara, 0.2% w/v unless otherwise indicated). sAAs were used at 1 mM unless otherwise indicated and purchased from: PepTech (pAcF, AL624-2), BaChem (pIF, F-3075.0005), and Chem-Impex International (pAzF, 03376).

Plasmids

All tRNAs used to assess tolerance for tryptophan and phenylalanine at TAG codons were contained within the pTech plasmid backbone and driven by the *lpp* promoter³¹. Isothermal assembly³² was used to replace the chloramphenicol acetyltransferase (*cat*) gene with the *sh ble* gene for resistance to zeocin.

The *supU* amber suppressor tRNA³³ was used to assess tolerance for tryptophan and a phenylalanine amber suppressor³¹ was used to assess tolerance for phenylalanine. pTech-*supU* was graciously provided by the laboratory of Dieter Söll and *supPhe* was synthesized by IDT and isothermally assembled into the pTech plasmid backbone to obtain pTech-*supPhe*.

Conversion of aminoacyl-tRNA synthetase specificity

The pAcF OTS was integrated into the genome of the GRO linked to a counter-selectable gene *tolC*. Co-selection multiplex automated engineering (CoS-MAGE³⁴) was used as described previously to introduce annotated mutations³⁵ to the sAA binding pocket of the aaRS for specificity toward pAzF or pIF (Supplementary Table 4). Sanger sequencing was used to verify these mutations. Conversion of sAA-specificity was assessed in sequence-verified clones upon growth in the presence of sAA incorporation and episomally-expressed GFP containing an in-frame TAG codon at residue 151 within the protein product. OTS-mediated suppression of this codon with the sAA (*i.e.*, pAzF, pIF) generated a full-length fluorescent product, indicating that sAA incorporation had occurred and specificity was achieved.

TAG codon incorporation into essential genes

We applied three unique strategies to identify permissive sites in essential genes for TAG codon incorporation (Fig. 1). In our first strategy, a subset of essential genes³⁶ were chosen for the incorporation of one or more TAG codons immediately after the start codon to encode a sAA at the amino-terminus. To explore a diverse library of incorporation targets within the *E. coli* proteome, in our second strategy we applied the Sorting Intolerant From Tolerant (SIFT) algorithm¹⁹ (downloaded on the Yale Biomedical High Performance Computing Cluster) to the entire panel of essential *E. coli* proteins³⁶. SIFT is an algorithm that uses sequence-homology to predict the tolerance of amino acid substitutions at different indices. In our workflow, genes were first split into three categories based on wild-type expression level³⁷, and a further four-subgroups by genomic location with the goal of targeting essential genes dispersed throughout the *E. coli* chromosome. Next, genes shown to be essential by multiple studies³⁶ were passed through SIFT. For each essential gene, two high, medium, and low tolerance sites were targeted for TAG incorporation by MAGE. By

this approach, we were able to sample diverse residue types in proteins with varying wild-type expression levels.

In our third strategy, using the conserved domain database¹³, we searched within all annotated essential proteins for tyrosine, tryptophan and phenylalanine residues predicted to participate in 1) essential enzymatic reactions or 2) protein-protein interactions (*e.g.*, dimerization). To minimize the probability that the added functionality of the sAA would perturb protein function, we targeted sites that were observed to occur as tyrosine or tryptophan in different homologs.

GROs containing an OTS integrated into the chromosome were grown to mid-log phase in liquid permissive LB and four cycles of MAGE were performed per pool of mutagenic oligos (oligo concentration 15 μ M) as described previously^{12,38}. To isolate synthetic auxotrophs, mutagenized cultures were plated on solid media and replica plated onto nonpermissive media. To identify TAG incorporation loci, MASC-PCR was used to interrogate pools of up to eleven targeted loci as previously described³⁸, followed by verification using Sanger sequencing.

Genotyping

Sanger sequencing was performed by the Keck DNA Sequencing Facility at Yale or by GENEWIZ, Inc. Genomic DNA for whole genome sequencing was prepared using a Qiagen Genomic DNA purification kit. Illumina libraries were prepared by the Yale Center for Genomic Analysis or the Dana Farber Cancer Institute. Illumina HiSeq or MiSeq sequencing systems were used for whole genome sequencing to generate 50 or 150 bp paired-end reads, respectively.

Whole genome sequencing was used to analyze three EMs per background. In all cases, the direct ancestor to the EM was also analyzed. SNPs in EMs were identified relative to the reference genome *Escherichia coli* C321. A (CP006698.1, GI:54981157) using a previously described¹¹ software pipeline. SNPs listed in Supplementary Table 8 and Supplementary Table 10 were called by Freebayes in EMs.

Strains

All GROs used in this study are derived from C321. A (CP006698.1, GI:54981157)³⁸ which lacks all TAG codons and release factor 1. This strain is derived from strain EcNR2 (*mutS:cat (ybhB-bioAB):[c1857 (cro-ea59):tetR-bla]*), modified from *E. coli* K-12 substr. MG1655. In all synthetic auxotrophs, the *M. jannaschii*-derived OTS was genomically integrated into the GRO fused to the counter-selectable gene *tolC*. The OTS consists of an L-arabinose-inducible aaRS driven by the *araBAD* promoter, and a constitutively expressed cognate amber-decoding tRNA driven by the *proK* promoter. All genome modifications that required incorporation of dsDNA (*e.g.*, modifications to the *mutS* gene, incorporation of antibiotic selectable markers) were performed via λ -Red recombination³⁹.

Nomenclature of genomically recoded organisms and synthetic auxotrophs

Number of essential TAG codons—To succinctly name strains, we have introduced a new one-letter amino acid code for sAAs using Greek lettering (pAcF = α , pIF = β , and pAzF = γ). Non-contained GROs lacking essential TAG codons are named according to the one letter sAA code for the specific OTS present in the organism. For example, a TAG GRO with a genomically integrated pAcF OTS is rEc. α .

Biocontained GROs containing essential TAG codons are named according to two conventions based on the number of essential TAG codons in the auxotroph: 1) Strains with one essential TAG are named by the essential protein containing the sAA and the position and identity of the residue substituted therein (*e.g.*, a strain containing pAcF at residue 113 in DnaX is DnaX.Y113 α); 2) Strains containing more than one essential TAG are named using the one letter sAA code for which the organism is auxotrophic. This is followed by a dependency code, d, indicating the presence of two (dB), three (dC) or four (dD) essential TAG codons, and then by a TAG combination number that uniquely identifies the specific combination of TAGs in the strain. Combinations are numbered from one through 46 and are listed in Supplementary Table 6.

Mismatch repair—The presence of a prime (') following the TAG combination number indicates that *mutS* has been restored at its native locus, imparting functional mismatch repair to the organism.

tRNA redundancy—Following the TAG combination number, t indicates the amino acid for which tRNA redundancy has been eliminated and is followed by the relevant amino acid (*e.g.*, a strain in which two of three total tyrosine tRNAs were deleted is tY).

Escape mutant identity—At least three EMs were characterized per strain background that permitted an escape mutant. An EM is designated by a number following the letter “E” (*e.g.*, E1).

The summary of synthetic auxotrophs generated in this study illustrated in Extended Data Fig. 1 was constructed using the Circos⁴⁰ software.

Fitness analysis

Strains were grown at 34°C in flat-bottomed 96-well plates containing 150 μ L of LB medium permissive for sAA incorporation, unless otherwise indicated. Strains were washed twice with sterile dH₂O prior to assessing growth in nonpermissive media. Kinetic growth (OD₆₀₀) was monitored on a BioTek plate reader at ten-minute intervals in triplicate. Raw OD₆₀₀ data from the plate reader were normalized to standard absorbance (OD₆₀₀ at 1 cm path length) values using an empirically derived calibration curve ($y = 1.9704x - 0.1183$, where $y = \text{OD}_{600}$ at 1 cm path length and $x = \text{OD}_{600}$ from plate reader; $R^2 = 0.998$). DTs were calculated in MATLAB using custom code. Reported values are the average between three technical replicates where error bars represent \pm s.d. All reported results repeated at least three times in independent experiments. Maximum OD₆₀₀ values were obtained after

24 hours of growth and represent the average of three technical replicates. Reported results repeated at least three times in independent experiments.

Mass spectrometry

Histidine-tagged proteins were purified on NiNTA resin (Qiagen). Resolution of purity was assessed via SDS-PAGE. In gel digestion was performed similarly to previously described methods⁴¹. Proteins were stained and imaged within the gel using Coomassie Blue (R-250). A band corresponding to the molecular weight of DnaX was excised. Gel slices were processed into 1 mm cubes, washed in 1:1 (v/v) 50% CH₃CN/50 mM NH₄HCO₃, and then washed in 1:1 (v/v) 50% CH₃CN/10 mM NH₄CO₃. 13.33 ng/μl trypsin solution in 9:1 (v/v) 50 mM NH₄CO₃/50% CH₃CN was added and samples were incubated overnight at 37°C. Peptides were extracted with 1:2 (v/v) 5% formic acid/50% CH₃CN and dried. Peptides were desalted by reconstitution in 3:8 (v/v) 70% formic acid/0.1% TFA, followed by loading onto a custom-made stage tip (2 × 1.06 mm punches of Empore C18 extraction disks [3 M] in a 200μl pipette tip)⁴² activated with 80% CH₃CN and 0.1% TFA. Tips were washed twice with 0.1% TFA and peptides eluted with 80% CH₃CN and 0.1% TFA. Peptides were dried and reconstituted for LC/MS/MS analysis. Capillary LC/MS/MS was carried out using an LTQ Orbitrap Velos (Thermo Scientific) with a nanoAcquity uHPLC (Waters) system as described previously⁴³. The data were processed as described previously¹¹. MASCOT scores were above the identity or extensive homology threshold and representative spectra are illustrated in Fig. 2c.

Quantitative assessment of amino acid tolerance

The following workflow was performed to assess the tolerance for natural amino acid substitution at residues chosen for sAA incorporation. Strains were grown to mid-log phase in 1 mL of permissive LB, and MAGE was performed as described¹² with modifications described here. Post induction of λ-Red proteins, cells were transferred to individual wells of a 96-well, V-bottomed plate, and washed twice at 4°C with sterile dH₂O. Cells were re-suspended in 50 μl of water or 1 μM mutagenic ssDNA to convert a single in-frame essential TAG codon to one of 20 sense codons, and electroporated in a 96-well plate. Cells were electroporated using the Harvard BTX electroporation system (2.4 kV, 750 Ω, 25 μF). Electroporated cells were recovered in 1.5 mL of fresh permissive media in a 96-well plate for four hours at 34°C shaking. Cells were pelleted, washed twice with sterile dH₂O, and re-suspended in 200 μL of 1× PBS. Serial dilutions were made in 1× PBS and 50 μL each of non-diluted and 100-fold diluted samples were plated on solid nonpermissive LB media. 50 μL each of higher dilutions were plated on permissive solid media and all plates were incubated for 20 hours at 34°C.

CFUs counted on nonpermissive media were expressed as a ratio of total CFUs on permissive media. Since the frequency of MAGE-mediated recombination (~ 0.3)¹² exceeds the EFs of these background strains (10^{-5}), we directly correlated these ratios to amino acid tolerance. MATLAB was used to calculate the log₁₀ of this ratio. Where no CFUs were observed on nonpermissive media, indicative of a highly intolerant substitution, a ratio could not be calculated and these values were defaulted to NaN within MATLAB. A heat map was

used to compare representative data for one experiment, where blue indicates a tolerated substitution and yellow, a non-tolerated substitution.

Twenty-one separate MAGE experiments were performed as described above for each strain, per essential genomic TAG, to assess tolerance for each of the 20 natural amino acids at each TAG site, plus a negative control (water). Strains with one TAG codon (SecY.Y122 β , DnaX.Y113 β , LspA.Y54 β , DnaA.W6 γ , SerS.F213 γ , and MurG.F243 γ) were assessed across 21 (including the negative control) experiments per strain, strains with two TAG codons (rEc. β .dB.9) were assessed across 42 (including two negative control) experiments per strain, and strains with three TAG codons (rEc. β .dC.12 and rEc. γ .dC.46) were assessed across 63 (including three negative control) experiments per strain. Reported results repeated at least three times in independent experiments.

Escape assays

Strains were grown in triplicate to late-log phase in 2 mL of permissive LB, pelleted, washed twice with sterile dH₂O, and re-suspended in 200 μ L 1 \times PBS. To obtain total and EM CFUs, serial dilutions were made and equal volumes were plated on permissive and nonpermissive solid media plates (100 \times 15 mm). Plates were incubated at 34 $^{\circ}$ C and EF was calculated as the total number of EM CFUs observed per total cells plated. Reported EFs are means of three technical replicates where error bars represent \pm s.d. To isolate EMs in strains with lower EFs, $\sim 10^{10}$ - 10^{11} cells were plated and the resulting EF from a representative escape assay is reported. When EMs were not detected upon plating $\sim 10^{10}$ - 10^{11} cells, the EF is described to be below the limit of detection and reported as less than a frequency of one over the total number of cells plated. In all cases, reported results repeated at least three times in independent experiments. Where temporal monitoring of EFs on solid media is reported (Fig. 3d), representative escape assays are plotted and results repeated at least three times in independent experiments.

Liquid escape assays

Long term liquid growth was assessed for two strain backgrounds: the pAzF-dependent strain, rEc. γ .dC.46'. tY and its non-contained ancestor, rEc. γ . Growth of rEc. γ .dC.46'. tY was separately assessed in permissive (+sAA/+L-arabinose) and nonpermissive (-sAA/-L-arabinose) media and growth of rEc. γ was assessed in nonpermissive media. In all cases, flasks contained carbenicillin to prevent contamination.

Strains were grown in 100 ml of LB media overnight. Cultures were then pelleted and washed twice with the same volume of sterile dH₂O. Washed pellets were re-suspended in LB media plus or minus small molecules and this slurry was then added to shake flasks containing 1 L of LB media plus or minus small molecules, at time zero. At this first time point, a 1 ml sample was obtained from each flask, from which the OD₆₀₀ was measured and 50 μ l was plated on both permissive (+sAA/+L-arabinose/+carbenicillin) and nonpermissive (-sAA/-L-arabinose/+carbenicillin) solid LB media, in three technical replicates. Average CFU counts are reported and error bars represent \pm s.d. In all cases, CFUs on solid media were counted after 24 hours of incubation at 34 $^{\circ}$ C. Time points were collected in this manner for all subsequent time points for the following 20 days.

After 20 days of growth in liquid media, the two 1 L cultures of rEc.γ.dC.46'. tY grown in nonpermissive and permissive media were interrogated for the presence of a single EM. The entire culture was pelleted, re-suspended in 7 ml of 1× PBS, and plated across 30 large nonpermissive solid media plates that were subsequently monitored for CFU formation over the following seven-day period. All reported results repeated at least three times in independent experiments.

Environmental challenges

Wild-type *E. coli* K-12 substr. MG1655, and additional *E. coli* strains EcNR2¹², rEc.γ.dC.46' (a pAzF auxotroph), and rEc.γ (non-contained ancestor to rEc.γ.dC.46') were grown to mid-log phase in 1 mL of LB media supplemented with small molecules, where necessary. Cultures were washed three times with sterile dH₂O and re-suspended in 1 mL 1× PBS. A total of 16, two-fold serial dilutions were made and spotted on the following solid media types: LB, EZ Rich Defined⁴⁴ (with modifications by Teknova) containing 100× carbon source: 40% glycerol, Blood Agar (Teknova), or Soil Extract Agar (HiMedia). Prior to spotting, plates were topically supplemented with pAzF, L-arabinose, and/or biotin, and dried for at least one hour. Spotted plates were incubated for one day at 34°C and photographed in a Gel Doc XR+ running ImageLab v4.0.1 (BioRad).

Selectable markers used in this study

>cat (1,015 bp)

```

cctgtgacggaagatcacttcgcagaataaataaactcctggtgtccctgttgataccgggaagcctgggccaacttttgcgaaaat
gagacgttgatcggcagcgaagaggttccaacttcaccataatgaaataagatcactaccggcgctatttttgagttgctgagatttc
aggagctaaggaagctaaaatggagaaaaaactcctggatataccaccgttgatataccaatggcagcgtaaagaacatttgag
gcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaagaccgtaagaaaaataagcac
aagttttaccggcctttattcacattcttccccctgatgaatgctcatccggaattacgtatggcaatgaagacggtagctggtga
tatgggatagttaccctgttacaccgtttccatgagcaaacgaaacgttttcacgctctggagtgatataccacgacatttccg
gcagttctcacacataattcgaagatggtgctgttacgggtaaacctggcctatttccctaaagggttattgagaatatttttctg
ctcagccaatcctgggtgagttcaccagttttgattaaactggccaatattggacaacttctcggccccgtttccaccatgggcaaa
tattatacgcaagcgacaaggtgctgatgccgctggcaggttcacatgccgtttgtgatggcttccatgctggcagaatgctt
aatgaattacaacagtactgcgatgagtgccagggcggggcgaatttttaaggcagttattggtgcccctaaacgctggtgctac
gcctgaataagtataaagcggatgaatggcagaattcgaagcaaatcgaccggctgctggttcagggcagggctggttaaa
tagccgcttatgtctattgctggtt

```

>kanR (1,165 bp)

```

cctgtgacggaagatcacttcgcagaataaataaactcctggtgtccctgttgataccgggaagcctgggccaacttttgcgaaaat
gagacgttgatcggcagcgaagaggttccaacttcaccataatgaaataagatcactaccggcgctatttttgagttgctgagatttc
aggagctaaggaagctaaaatgagccatattcaacgggaaacgctgagggccgcgattaaattccaacatggatgctgatttatgg
gtataaatgggctcgcgataatgctgggcaatcaggtgcaaatctatcgcttgatgggaagcccgatgccagagttgtttctg
aaacatggcaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaattatgcctctccgacct
caagcattttaccgtactcctgatgatgcatggttactaccactcgatccccgaaaaacagcattccaggtattagaagaatatcc
tgattcaggtgaaaatattgttgatgctgctggcaggtttcctgcccgttcattcattctgtttgtaattgcttttaaacagcagctg
cgtatttcgctcgcctcagggcgaatcacgaatgaataacggtttggttgatgagtgatgagcagcgtaatggctgcccctgt
tgaacaagctggaagaaatgcataaactttgccatttcaccggattcagctgctcactcatggtgatttctacttgataacctattttt

```

gacgaggggaaattaataggtgtattgatgttgacgagtcggaatcgacaccgataccaggatcttccatcctatggaactgcc
 tcggtgagtttctcctcattacagaacggcttttcaaaaatattggtattgataatcctgatataaataaagtcagtttcatttgatgctc
 gatgatgttttctaatttttaaggcagttattggtgcccttaacgcctggtgtacgctgaataagtgataataagcggatgaatgg
 cagaaattcgaagcaaaatcgaccggctgctggttcaggggcagggctgtaaatagcccttatgtctattgctggt

>spec^R (1,201 bp)

cagccaggacagaaatgcctcgacttcgctgctgcccaagggtgccgggtgacgcacaccgtggaacggatgaaggcacgaac
 ccagtgacataagcctgttcggtcgtaatgcaagtagcgtatgcgctcacgcaactggtccagaacctgaccgaacg
 cagcgggtgtaacggcgagtgccggttttcatggctgttatgactgtttttggggtacagctctatgcctcgggcatcaagcagca
 agcgcgttacccgctgggtcgtgtttgatgttatggagcagcaacgatgttacgcagcagggcagtcgcctaaaacaaagttaa
 catcatgagggaaagcgggtgacgccgaagatcactcaactatcagaggtatgtgctcgcgagccatctcgaaccagctt
 gctggccgtacattgttacgctccgagtggtgagcggcctgaagccacacagtgatattgattgctggttacggtgaccgtaagg
 ctgtatgaaacaacgcggcgagctttgatcaacgacctttggaacttcggcttcccctggagagagcgagattcctcgcgctgtag
 aagtcaccattgtgtgcacgacgacatcctcgtggcgttatccagctaaagcgcgaactgcaattggagaatggcagcgcaatga
 cattctgcaggtatcttcgagccagccacgatcgacattgatctggctatcttctgacaaaagcaagagaacatagcgttgccttgg
 aggtccagcggcgaggaaactcttgatccggttcgtaacaggatctattgagggcctaataaaccctaacgctatggaactcg
 ccgccgactgggctggcgatgagcgaatgtagtcttacggtgtcccatttggtagcgcagtaaccggcaaatcgcgcc
 gaaggatgctgctgccgactgggcaatggagcgcctgccgcccagatcagcccgcatactgaagctagacagccttatcttgg
 acaagaagaagatcgttggcctcgcgcgagatcagttggaagaattgtccactacgtgaaaggcgagatcaccaaggtagtgcg
 caaataaagctttactgagctaataacaggactgctgtaatcgacggcctttttattctgca

>tolC (1,746 bp)

ttgaggcacattaacgccctatggcacgtaacccaaccttttgcggtagcggcttctgtagaatccgcaataatttaccagtttgatc
 gcgctaaactgcttaccacaaggaatgcaaatgaagaaattgctcccattctatcggcctgagccttctgggtcagttcgttga
 gccagggccgagaacctgatgcaagttatcagcaagcagccttagtaaccggaaftgcgtaagctgcccgcgatcgtgatgctg
 cctttgaaaaaataatgaagcgcgagtcctactgccacagctaggtttaggtgcagattacacctatagcaacggctaccgga
 cgcgaacggcatcaactaagcgcgaccagtgcgctccttcagttactcaatcattttgatatgcaaatggcgtgctgtaacgct
 gcaggaaaaagcagcagggattcaggacgtcacgtatcagaccgatcgaacacctgatcctcaacaccgcgaccgcttattca
 acgtgtgaaftgctattgacgttcttctatacacaggcacaanaaagcagatcaccgtcaattagatcaaacaccaccaacgttta
 acgtgggctgtagcgtgacaccagctgcagaacccccgcacagtagataccgtgctggcgaacgaagtgaccgcagta
 ataacttgataacgcggtagagcagctgcgccagatcaccggtaactactatccggaactggctgctgctgaatgcaaaacttaa
 aaccgacaaaccacagccggttaacgcgctgctgaaagaagccgaaaacgcaacctgctgctgttacaggcacgcttgagccag
 gacctggcgcgagcaaatcgccagggcaggatggtcacttaccgactctggtttaacggcttaccgggatttctgacacct
 ctatagcggttcgaiaaccggtgctgctggtaccagatagcaatattggccagaacaaagttggcctgagcttctc
 gctgccgattatcaggggcaatggttaactgcaggtgaaacaggcacagtagaactttgctgggtgccagcgagcaactggaaa
 gtgccatcgtagcgtgagaccgtgcttctcctcaacaacatfaatgcatctatcagtagcattaacgcctacaacaagcc
 gtatftccgctcaaagctcattagacgcgatggaagcgggctactcggctggtacgcgtaccattgtgatgttggatgcgacca
 ccacgtgtacaacgccaagcagagctggcgaatgcgctgtataactacctgattaatcagctgaatattaagtcagctctgggtac
 gttgaacgagcagatctgctgacactgaacaatgcgctgagcaaacgggttccactaatccgaaaacgttgaccgcaaacgc
 cggaacagaatgctattgtgatgttatgcgctgatagcccggcaccagctgttcagcaaacatccgcagcactaccaccagta
 acggtcataacccttccgtaactgatgacgacgaggggaagcttaattagctgatctagagcatcaataaacgaaagctca
 gtcgaaaagactgggcttctgtttatctgtttgtcggtaacgctcctcctgtagtaggacaaatccgccccctaga

Author Manuscript

Author Manuscript

Author Manuscript

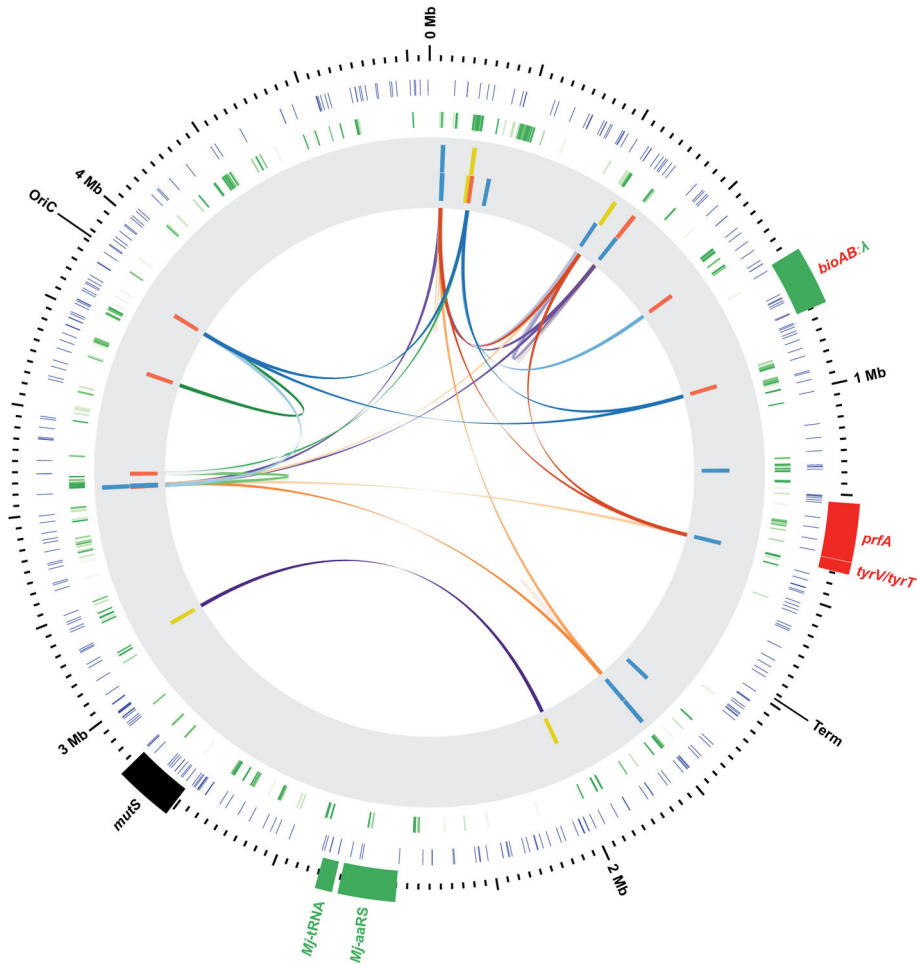
Author Manuscript

>zeo^R (762 bp)

```

gggtgtgacaattaatcatcggcatagatatcgccatagataatacagacaaggtgaggaactaaacatggccaagttgaccagtg
ccgttccgggtgctcaccgcgcgcgacgtcgcggagcggctcgagttctggaccaccggctcgggttctcccgggacttcgtgga
ggacgacttcgccggtgtggtccgggacgacgtgaccctgttcatcagcgcgggtccaggaccaggtggtgccggacaacacctg
gcctgggtgtgggtgctcggcctggacgagctgtacgccagtggtcggaggtcgtgtccacgaactccgggacgctccggg
ccggccatgaccgagatcggcgagcagccgtgggggagggttcgacctgcgcgaccggccggcaactgcgtgcacttcgt
ggccgaggagcaggactgacacgtccgacggcggccacgggtcccaggcctcggagatccgtccccctttctcttgcgatac
atgtaattagttatgacagcttacattcacgcctccccccacatccgtctaaccgaaaaggaaggagttagacaacctgaagtcta
ggtccctatttttttatagttatgttagtattaagaacgtatttatattcaattttctttttctgtacagacgcgtgtacgatgtaac
attatactgaaaacctgcttgagaaggtttgggacgctcgaaggcttaattgcaagct
    
```

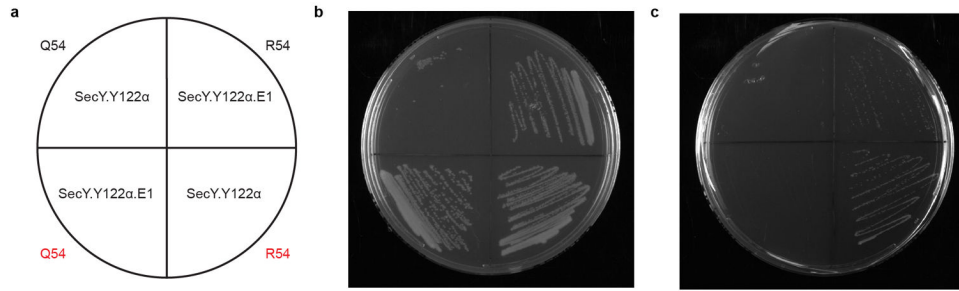
Extended Data



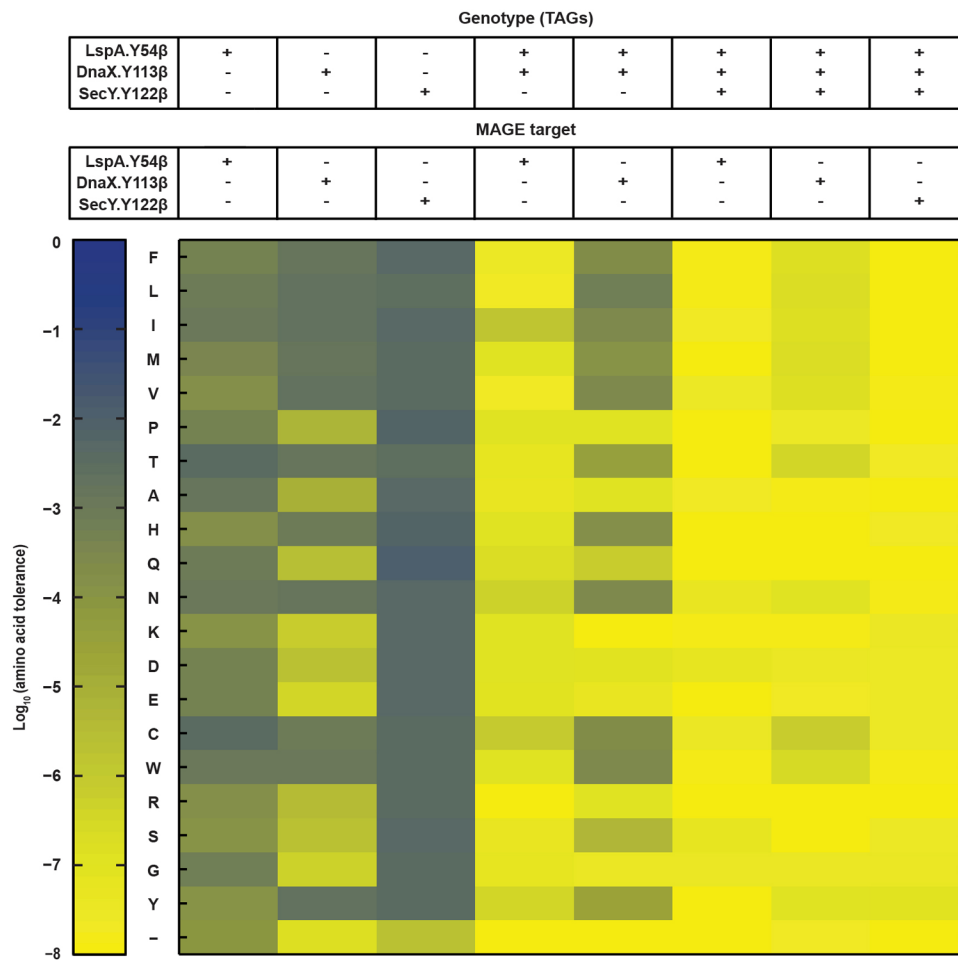
Extended Data Figure 1. Comprehensive map of synthetic auxotrophs

Circos plot summarizing synthetic auxotrophs generated in this study. Red and green genes reflect knockouts and insertions, respectively. Outermost ticks indicate genomic location, inner blue ticks indicate locations where TAG codons were converted to TAA in the GRO, and green ticks reflect locations of 303 *E. coli* essential genes. The shaded grey inner circle contains essential TAG loci in synthetic auxotrophs, where yellow ticks represent amino-

terminal insertions, blue ticks represent tolerant substitutions, and red ticks represent functional site substitutions. Innermost links represent unique combinations of TAGs in higher-order synthetic auxotrophs. Links of a single color correspond to a single strain.

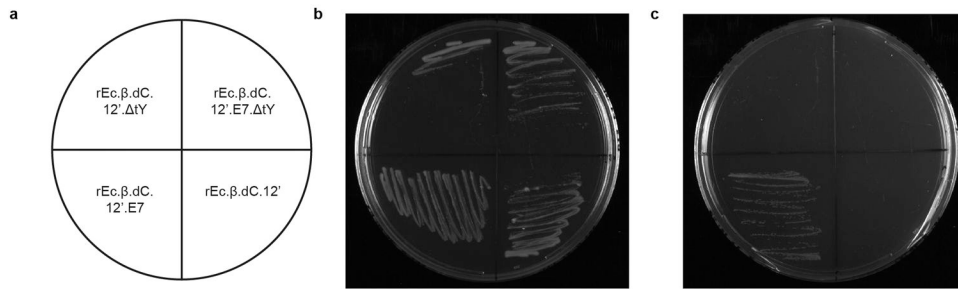


Extended Data Figure 2. *rpsD.Q54R* is sufficient for loss of pAcF-dependence in SecY.Y122α
a, Plate map with genotypes of strains shown in **b** and **c**. Top half of plate: SecY.Y122α.E1 (upper right quadrant) contains the *rpsD.Q54R* mutation and is an EM of pAcF-auxotroph, SecY.Y122α (upper left quadrant). Lower half of plate: The *rpsD.Q54R* mutation was introduced into SecY.Y122α (lower right quadrant), resulting in a loss of pAcF-dependence, and reverted to wild type in SecY.Y122α.E1 (lower left quadrant), restoring pAcF-dependence. The amino acid present at residue 54 within RpsD is indicated at the perimeter of the plate, where red signifies that the given mutation was introduced into the genotype by MAGE to demonstrate the causal mechanism of escape. **b**, Growth on solid permissive media demonstrates growth of all strains. **c**, Growth on solid nonpermissive media. Introduction of the *rpsD.Q54R* mutation into the synthetic auxotroph SecY.Y122α results in loss of containment (lower right quadrant). Reverting the mutation to wild type in SecY.Y122α.E1 results in restoration of containment (lower left quadrant). Together, these data demonstrate that the *rpsD.Q54R* mutation is sufficient for loss of pAcF-dependence in SecY.Y122α.



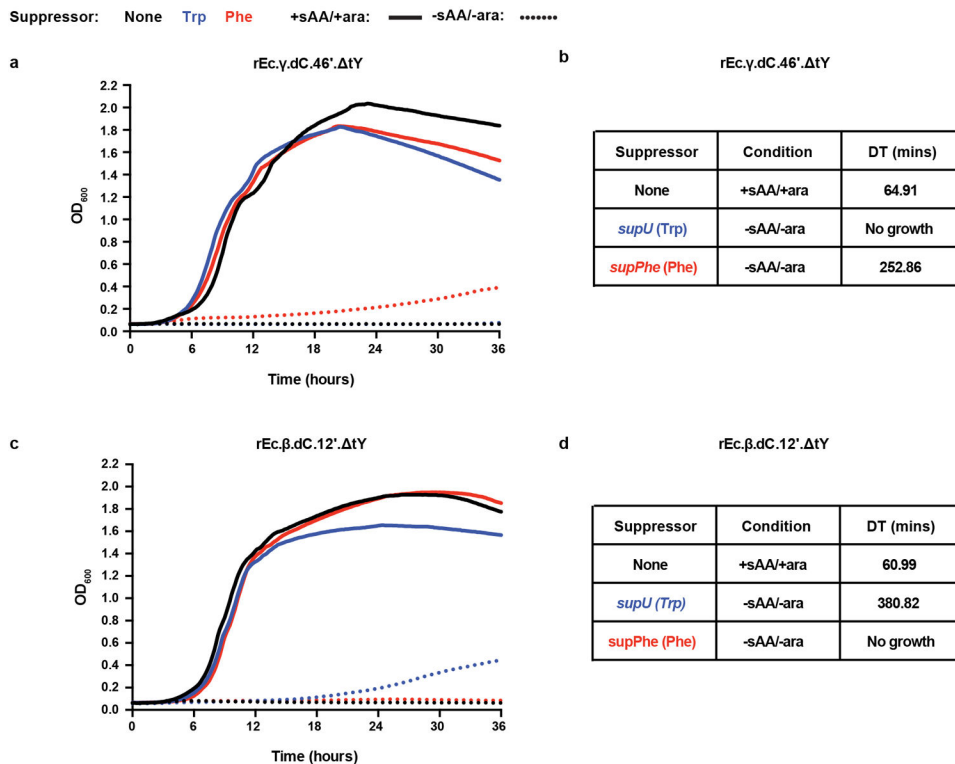
Extended Data Figure 3. Quantitative assessment of amino acid tolerance in higher-order pIF auxotrophs

Representative assay surveying tolerance of one of three essential TAG loci to the twenty amino acids in different synthetic auxotrophs and expressed as \log_{10} of total cell survival. The + symbol indicates the presence of a TAG codon at the specified locus in the background strain and - indicates the wild type codon. Blue and yellow indicate high and low tolerance to substitution, respectively. Substitutions DnaX.Y113W and SecY.Y122Q are tolerated but yielded a lower percentage of survival on nonpermissive media in a background with two TAGs, an effect that was pronounced in a background with three TAGs. While DnaX.Y113, SecY.Y122 and LspA.Y54 are permissive for most natural amino acids, strains with more than one of these essential TAGs are less prone to survive in the event that any one TAG is compromised. SecY.Y122Q and DnaX.Y113W were tolerated substitutions also observed in real EMs of these strains (Supplementary Table 7). Refer to the methods for a complete description of this experiment.



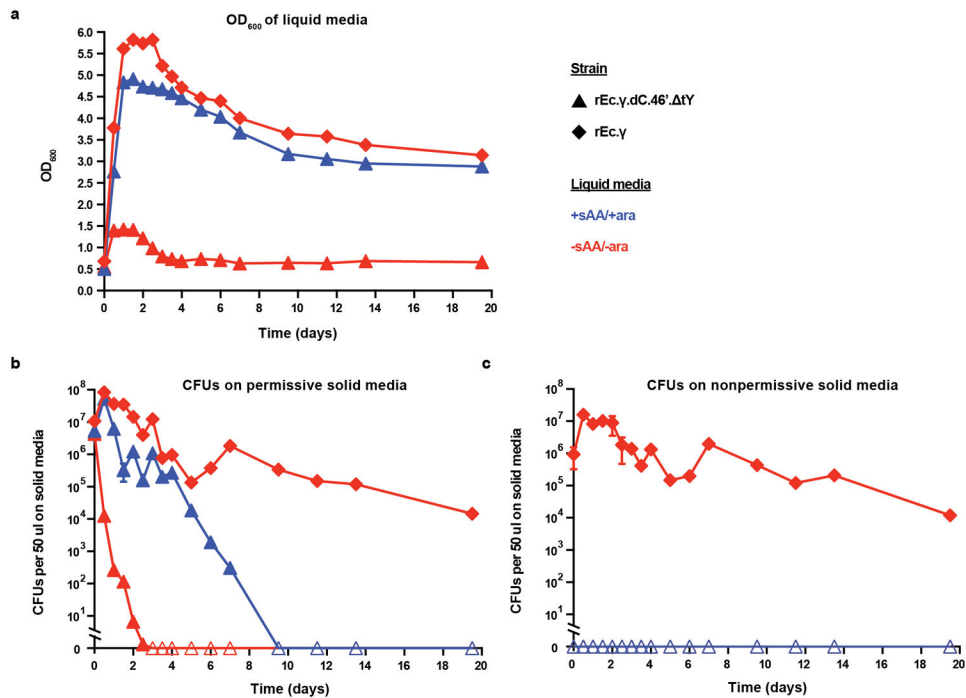
Extended Data Figure 4. Deletion of *tyrT* and *tyrV* restores pIF-dependence and fitness of rEc.β.dC.12'.E7

a, Plate map with genotypes of strains in parts **b** and **c**. rEc.β.dC.12'.E7 is an EM of its sAA-dependent ancestor (rEc.β.dC.12') and contains a *tyrT* ochre suppressor mutation (*supC*). The fitness of rEc.β.dC.12'.E7 in permissive media is impaired relative to rEc.β.dC.12', with DTs of 91.74 (\pm 1.49) and 61.81 (\pm 0.65) minutes, respectively. Tyrosine tRNA redundancy was eliminated (Δ tY) in both strains by λ -Red mediated replacement of *tyrT* and *tyrV* with chloramphenicol acetyltransferase (*cat*), rendering the resulting strains (rEc.β.dC.12'. Δ tY and rEc.β.dC.12'.E7. Δ tY) dependent on *tyrU* for tyrosine incorporation during normal protein synthesis. Elimination of tyrosine redundancy reduced the EF of rEc.β.dC.12' from 2.17×10^{-9} (Fig. 2e) to $< 4.85 \times 10^{-12}$ (no EMs were observed upon plating 2.06×10^{11} cells) and restored pIF-dependence in rEc.β.dC.12'.E7 to $< 4.73 \times 10^{-12}$ (no EMs were observed upon plating 2.12×10^{11} cells). EMs were not detected for either strain up to seven days after plating on nonpermissive media (Fig. 3d and Supplementary Table 11). Tyrosine tRNA deletion also restored the fitness of the EM to approximately that of its sAA-dependent ancestor (60.66 \pm 0.12 minutes). Taken together, these results establish *tyrT* as the causal mechanism of escape in rEc.β.dC.12'.E7. **b**, Growth on solid permissive LB media. **c**, Growth on solid nonpermissive LB media. All reported DTs are averages, where n=3 technical replicates, and error bars are \pm s.d. Refer to the methods for a complete description of EFs.



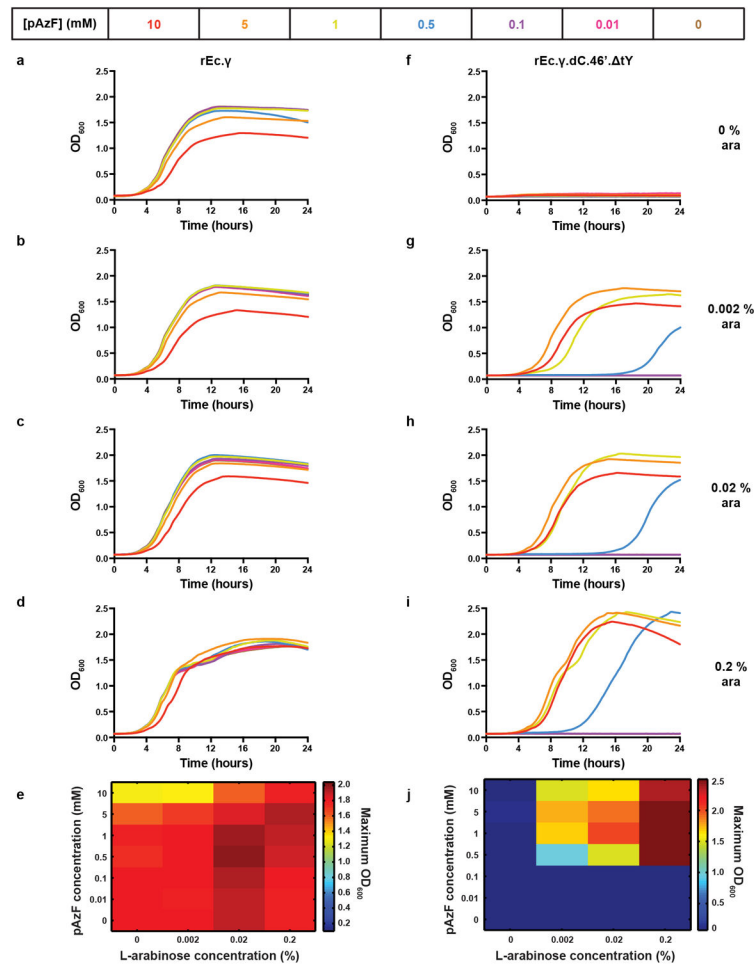
Extended Data Figure 5. Growth profiles of strains expressing phenylalanine or tryptophan amber suppressor tRNAs

Growth was assessed for rEc.γ.dC.46'. tY and rEc.β.dC.12'. tY in the presence of amber suppression by either pTech-*supU* (blue), pTech-*supPhe* (red), or in the absence of plasmid-based amber suppression (black). Cells were washed twice with dH₂O and re-suspended in the same volume of 1× PBS. Washed cells were normalized by OD₆₀₀ to inoculate roughly equal numbers of cells per well. Growth profiles are shown for **a**, **b**, rEc.γ.dC.46'. tY and **c**, **d**, rEc.β.dC.12'. tY in permissive (+sAA/+l-arabinose, solid lines) and nonpermissive (-sAA/-l-arabinose, dashed lines) LB liquid media. DTs are shown for the ancestral strain (black) in permissive media and suppressor-containing strains (red and blue) in nonpermissive media where growth was observed. Plasmid containing strains were always grown in the presence of zeocin for plasmid maintenance. Growth was never observed for the contained ancestors in nonpermissive media (black, dashed lines). In the presence of tryptophan suppression, growth of rEc.γ.dC.46'. tY was not observed and growth of rEc.β.dC.12'. tY was severely impaired (380 minute DT) with a 6.24-fold increase in DT relative to the contained ancestor grown in permissive media. In the presence of phenylalanine suppression, growth of rEc.β.dC.12'. tY was not observed and growth of rEc.γ.dC.46'. tY was severely impaired (252 minute DT) with a 3.90-fold increase in DT relative to the contained ancestor grown in permissive media. Representative growth profiles and DTs are reported. These results repeated at least three times in individual experiments.



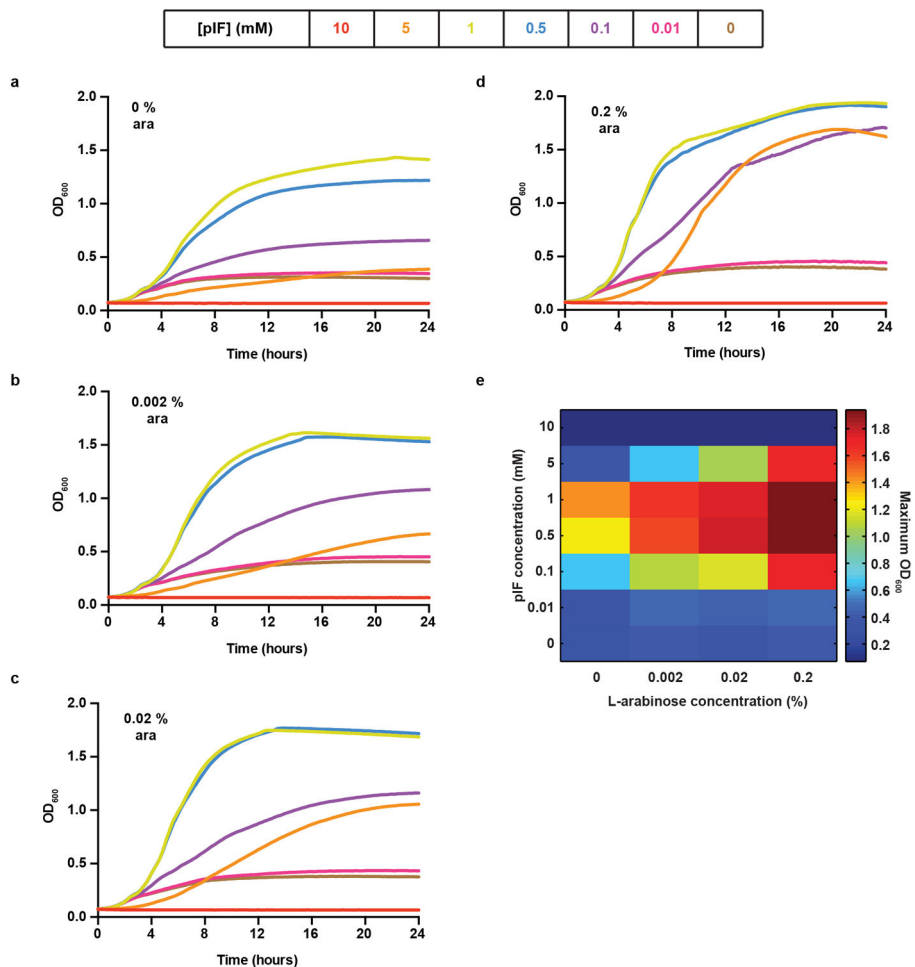
Extended Data Figure 6. Long term growth of rEc.γ.dC.46'. tY in liquid LB media relative to rEc.γ

Approximately 10^{11} cells of strain rEc.γ.dC.46'. tY (▲) was inoculated into 1 L of permissive (+sAA/+L-arabinose, blue) or nonpermissive (-sAA/-L-arabinose, red) LB media and incubated with agitation at 34°C for 20 days. Results from the equivalent experiment with the non-contained ancestor rEc.γ (◆) are also shown. Cultures were frequently monitored by a, OD₆₀₀ and quantification of CFUs on solid b, permissive (+sAA/+L-arabinose) and c, nonpermissive (-sAA/-L-arabinose) LB media. CFUs are plotted as the average of three replicates. Open symbols indicate that no CFUs were observed. Symbols for rEc.γ.dC.46'. tY are not visible because CFUs were never observed from either permissive or nonpermissive liquid cultures plated on nonpermissive solid media. At the end of the 20-day growth period, both cultures containing rEc.γ.dC.46'. tY were interrogated for the presence of a single EM by plating each 1 L culture across 30 nonpermissive solid media plates. CFUs were not observed and remained below the limit of detection for the following seven-day observation period. We hypothesize that the decrease in CFU counts obtained on permissive solid media for the permissive culture of rEc.γ.dC.46'. tY reflects pAzF degradation at 6 days. Reported CFUs are averages, where n=3 technical replicates, and error bars are ±s.d. Refer to the methods for a complete description of this experiment.

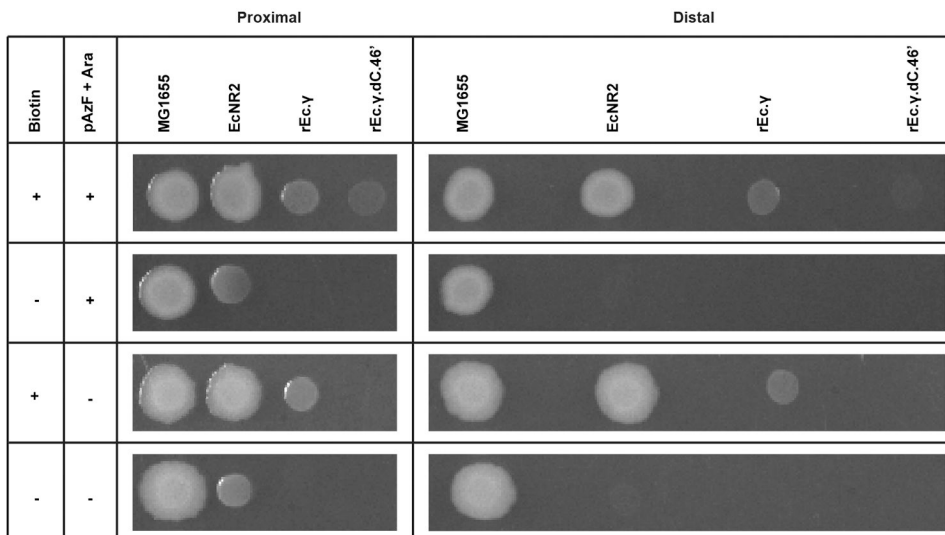


Extended Data Figure 7. Dose-dependent growth of rEc.γ.dC.46' tY in pAzF and L-arabinose compared to the non-contained ancestor

Growth in LB media supplemented with different concentrations of pAzF and L-arabinose. Growth profiles for rEc.γ across a gradient of pAzF concentrations in the presence of **a**, 0%, **b**, 0.002%, **c**, 0.02%, and **d**, 0.2% L-arabinose. Growth profiles for rEc.γ.dC.46' tY across a gradient of pAzF concentrations in the presence of **f**, 0%, **g**, 0.002%, **h**, 0.02%, and **i**, 0.2% L-arabinose. **e** and **j**, Growth profiles illustrated in parts a-d and f-i are depicted as heat maps in parts e and j, respectively, where the maximum OD₆₀₀ was obtained from the average of three replicates and plotted in MATLAB. Reported growth profiles and heat map values are averages, where n=3 technical replicates, and error bars are ±s.d.



Extended Data Figure 8. Dose-dependent growth of rEc.β.dC.12'. tY in pIF and L-arabinose
 Growth in LB media supplemented with different concentrations of pIF and L-arabinose. Growth profiles for rEc.β.dC.12'. tY across a gradient of pIF concentrations in the presence of **a**, 0%, **b**, 0.002%, **c**, 0.02%, and **d**, 0.2% L-arabinose. **e**, Growth profiles illustrated in parts a-d are depicted as a heat map, where the maximum OD₆₀₀ was obtained from the average of three replicates and plotted in MATLAB. Reported growth profiles and heat map values are averages, where n=3 technical replicates, and error bars are ±s.d.



Extended Data Figure 9. Proximity-dependent complementation of biotin auxotrophy
Wild type *E. coli* K-12 substr. MG1655 and three strains auxotrophic for biotin, EcNR2, rEc.γ (a non-contained GRO with an integrated pAzF OTS), and rEc.γ.dC.46' (also a synthetic auxotroph) were grown either adjacent or separately on rich-defined solid media. EcNR2 grew on biotin-deficient media when plated in close proximity to wild type *E. coli*, suggesting cross-feeding of the essential metabolite. The pAzF auxotroph only grew on media supplemented with biotin, pAzF, and L-arabinose.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to D. Söll and Y.S. Wang for inspired discussion and for providing pTech-*supU*. We thank N. Carriero and R. Bjornson at the Yale Biomedical High Performance Computing Cluster for assistance with SIFT. Funding was from Defense Advanced Research Projects Agency (N66001-12-C-4020, N66001-12-C-4211); NIH-MSTP-TG-T32GM07205 (A.D.H.); DuPont Inc. and the Arnold and Mabel Beckman Foundation (F.J.I.).

References

- Way JC, Collins JJ, Keasling JD, Silver PA. Integrating Biological Redesign: Where Synthetic Biology Came From and Where It Needs to Go. *Cell*. 2014; 157:151–161.10.1016/j.cell.2014.02.039 [PubMed: 24679533]
- Schmidt M, de Lorenzo V. Synthetic constructs in/for the environment: managing the interplay between natural and engineered Biology. *FEBS letters*. 2012; 586:2199–2206.10.1016/j.febslet.2012.02.022 [PubMed: 22710182]
- Moe-Behrens GH, Davis R, Haynes KA. Preparing synthetic biology for the world. *Frontiers in microbiology*. 2013; 4:5.10.3389/fmicb.2013.00005 [PubMed: 23355834]
- Pieper DH, Reineke W. Engineering bacteria for bioremediation. *Current opinion in biotechnology*. 2000; 11:262–270. [PubMed: 10851148]
- Steidler L. Genetically engineered probiotics. *Best practice & research Clinical gastroenterology*. 2003; 17:861–876. [PubMed: 14507594]
- Kong W, et al. Regulated programmed lysis of recombinant Salmonella in host tissues to release protective antigens and confer biological containment. *Proceedings of the National Academy of*

- Sciences of the United States of America. 2008; 105:9361–9366.10.1073/pnas.0803801105 [PubMed: 18607005]
7. Szafranski P, et al. A new approach for containment of microorganisms: dual control of streptavidin expression by antisense RNA and the T7 transcription system. *Proceedings of the National Academy of Sciences of the United States of America*. 1997; 94:1059–1063. [PubMed: 9037005]
 8. Steidler L, et al. Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin 10. *Nature biotechnology*. 2003; 21:785–789.10.1038/nbt840
 9. Ronchel MC, Ramos JL. Dual system to reinforce biological containment of recombinant bacteria designed for rhizoremediation. *Applied and environmental microbiology*. 2001; 67:2649–2656.10.1128/AEM.67.6.2649-2656.2001 [PubMed: 11375176]
 10. Jensen LB, Ramos JL, Kaneva Z, Molin S. A substrate-dependent biological containment system for *Pseudomonas putida* based on the *Escherichia coli* *gef* gene. *Applied and environmental microbiology*. 1993; 59:3713–3717. [PubMed: 8285679]
 11. Lajoie MJ, et al. Genomically recoded organisms expand biological functions. *Science*. 2013; 342:357–360.10.1126/science.1241459 [PubMed: 24136966]
 12. Wang HH, et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. 2009; 460:894–898.10.1038/nature08187 [PubMed: 19633652]
 13. Marchler-Bauer A, et al. CDD: a Conserved Domain Database for protein classification. *Nucleic acids research*. 2005; 33:D192–196.10.1093/nar/gki069 [PubMed: 15608175]
 14. Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America*. 1973; 70:3240–3244. [PubMed: 4594039]
 15. Berg P, Baltimore D, Brenner S, Roblin RO, Singer MF. Summary statement of the Asilomar conference on recombinant DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America*. 1975; 72:1981–1984. [PubMed: 806076]
 16. NIH Guidelines for research involving recombinant or synthetic DNA molecules. Office of Biotechnology Activities; Bethesda, MD: 2013.
 17. Betenbaugh MJ, Beaty C, Dhurjati P. Effects of plasmid amplification and recombinant gene expression on the growth kinetics of recombinant *E. coli*. *Biotechnology and bioengineering*. 1989; 33:1425–1436.10.1002/bit.260331110 [PubMed: 18587883]
 18. Sorensen SJ, Bailey M, Hansen LH, Kroer N, Wuertz S. Studying plasmid horizontal transfer in situ: a critical review. *Nature reviews Microbiology*. 2005; 3:700–710.10.1038/nrmicro1232 [PubMed: 16138098]
 19. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols*. 2009; 4:1073–1081.10.1038/nprot.2009.86 [PubMed: 19561590]
 20. Schaaper RM, Dunn RL. Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: the nature of in vivo DNA replication errors. *Proceedings of the National Academy of Sciences of the United States of America*. 1987; 84:6220–6224. [PubMed: 3306672]
 21. Eggertsson G, Soll D. Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*. *Microbiological reviews*. 1988; 52:354–374. [PubMed: 3054467]
 22. Maisnier-Patin S, Berg OG, Liljas L, Andersson DI. Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. *Molecular microbiology*. 2002; 46:355–366. [PubMed: 12406214]
 23. Holberger LE, Hayes CS. Ribosomal protein S12 and aminoglycoside antibiotics modulate A-site mRNA cleavage and transfer-messenger RNA activity in *Escherichia coli*. *The Journal of biological chemistry*. 2009; 284:32188–32200.10.1074/jbc.M109.062745 [PubMed: 19776006]
 24. Abe Y, et al. Structure and function of DnaA N-terminal domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on *oriC*. *The Journal of biological chemistry*. 2007; 282:17816–17827.10.1074/jbc.M701841200 [PubMed: 17420252]
 25. Bosl M, Kersten H. A novel RNA product of the *tyrT* operon of *Escherichia coli*. *Nucleic acids research*. 1991; 19:5863–5870. [PubMed: 1840671]
 26. O'Donoghue P, Ling J, Wang YS, Soll D. Upgrading protein synthesis for synthetic biology. *Nature chemical biology*. 2013; 9:594–598.10.1038/nchembio.1339 [PubMed: 24045798]

27. Lajoie MJ, et al. Probing the limits of genetic recoding in essential genes. *Science*. 2013; 342:361–363.10.1126/science.1241460 [PubMed: 24136967]
28. Bain JD, Switzer C, Chamberlin AR, Benner SA. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature*. 1992; 356:537–539.10.1038/356537a0 [PubMed: 1560827]
29. Pinheiro VB, et al. Synthetic genetic polymers capable of heredity and evolution. *Science*. 2012; 336:341–344.10.1126/science.1217622 [PubMed: 22517858]
30. Hammerling MJ, et al. Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. *Nature chemical biology*. 2014; 10:178–180.10.1038/nchembio.1450 [PubMed: 24487692]
31. Normanly J, Masson JM, Kleina LG, Abelson J, Miller JH. Construction of two *Escherichia coli* amber suppressor genes: tRNAPheCUA and tRNACysCUA. *Proceedings of the National Academy of Sciences of the United States of America*. 1986; 83:6548–6552. [PubMed: 3529087]
32. Gibson DG, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*. 2009; 6:343–345.10.1038/nmeth.1318 [PubMed: 19363495]
33. Fan C, Ho JM, Chirathivat N, Soll D, Wang YS. Exploring the substrate range of wild-type aminoacyl-tRNA synthetases. *Chembiochem: a European journal of chemical biology*. 2014; 15:1805–1809.10.1002/cbic.201402083 [PubMed: 24890918]
34. Carr PA, et al. Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic acids research*. 2012; 40:e132.10.1093/nar/gks455 [PubMed: 22638574]
35. Young TS, Ahmad I, Yin JA, Schultz PG. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *Journal of molecular biology*. 2010; 395:361–374.10.1016/j.jmb.2009.10.030 [PubMed: 19852970]
36. Baba T, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology*. 2006; 2:20060008.10.1038/msb4100050
37. Richmond CS, Glasner JD, Mau R, Jin H, Blattner FR. Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic acids research*. 1999; 27:3821–3835. [PubMed: 10481021]
38. Isaacs FJ, et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science*. 2011; 333:348–353.10.1126/science.1205822 [PubMed: 21764749]
39. Sharan SK, Thomason LC, Kuznetsov SG, Court DL. Recombineering: a homologous recombination-based method of genetic engineering. *Nature protocols*. 2009; 4:206–223.10.1038/nprot.2008.227 [PubMed: 19180090]
40. Krzywinski M, et al. Circos: an information aesthetic for comparative genomics. *Genome research*. 2009; 19:1639–1645.10.1101/gr.092759.109 [PubMed: 19541911]
41. Rinehart J, et al. WNK2 kinase is a novel regulator of essential neuronal cation-chloride cotransporters. *The Journal of biological chemistry*. 2011; 286:30171–30180.10.1074/jbc.M111.222893 [PubMed: 21733846]
42. Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical chemistry*. 2003; 75:663–670. [PubMed: 12585499]
43. Heinemann IU, et al. Enhanced phosphoserine insertion during *Escherichia coli* protein synthesis via partial UAG codon reassignment and release factor 1 deletion. *FEBS letters*. 2012; 586:3716–3722.10.1016/j.febslet.2012.08.031 [PubMed: 22982858]
44. Neidhardt FC, Bloch PL, Smith DF. Culture medium for enterobacteria. *Journal of bacteriology*. 1974; 119:736–747. [PubMed: 4604283]

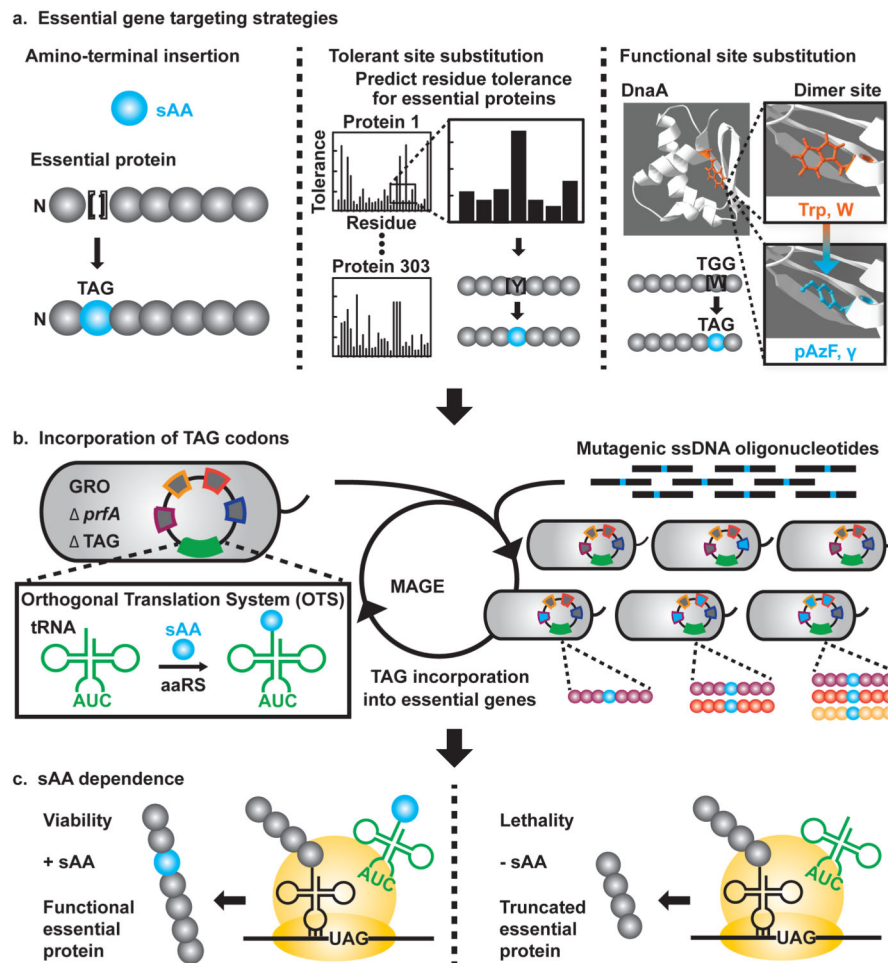


Figure 1. Strategy used to engineer GROs to require sAAs for growth

a, Approaches used to identify suitable loci within essential proteins for synthetic amino acid (sAA, blue) incorporation. **b**, MAGE was used for site-specific incorporation of TAG codons into essential genes of a genomically recoded organism (GRO) lacking all natural TAG codons (ΔTAG) and release factor 1 ($\Delta prfA$), and containing an orthogonal translation system (OTS, green) consisting of the *M. jannaschii* aminoacyl-tRNA synthetase (aaRS) and cognate UAG-decoding tRNA. **c**, Synthetic auxotrophs that depend on sAAs for growth were isolated.

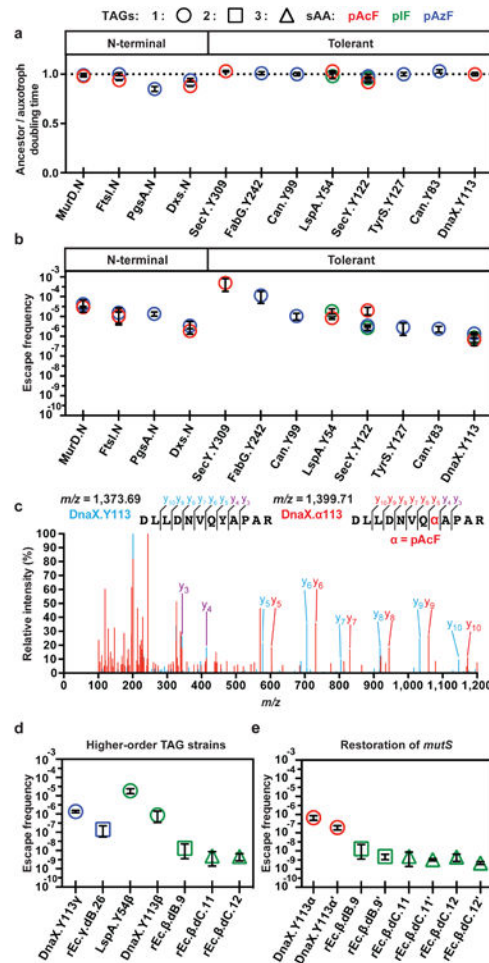


Figure 2. Characterization of strains dependent on sAA incorporation in essential proteins
a, Doubling time (DT) ratios for the non-contained ancestor to synthetic auxotroph containing one TAG. **b**, Escape frequencies (EFs) of strains from **a**. **c**, Superimposed MS/MS spectra for DnaX peptides from DnaX.Y113 α (red) and the non-contained ancestor rEc. α (blue). Overlapping peaks are purple and a mass shift relative to rEc. α identifies Y113 as the pAcF incorporation site in DnaX.Y113 α ; see methods. **d**, EFs for strains with multiple TAG codons **e**, and/or functional mismatch repair (prime, *mutS*⁺). For all plots, average values of 3 technical replicates are plotted with error bars, \pm s.d.

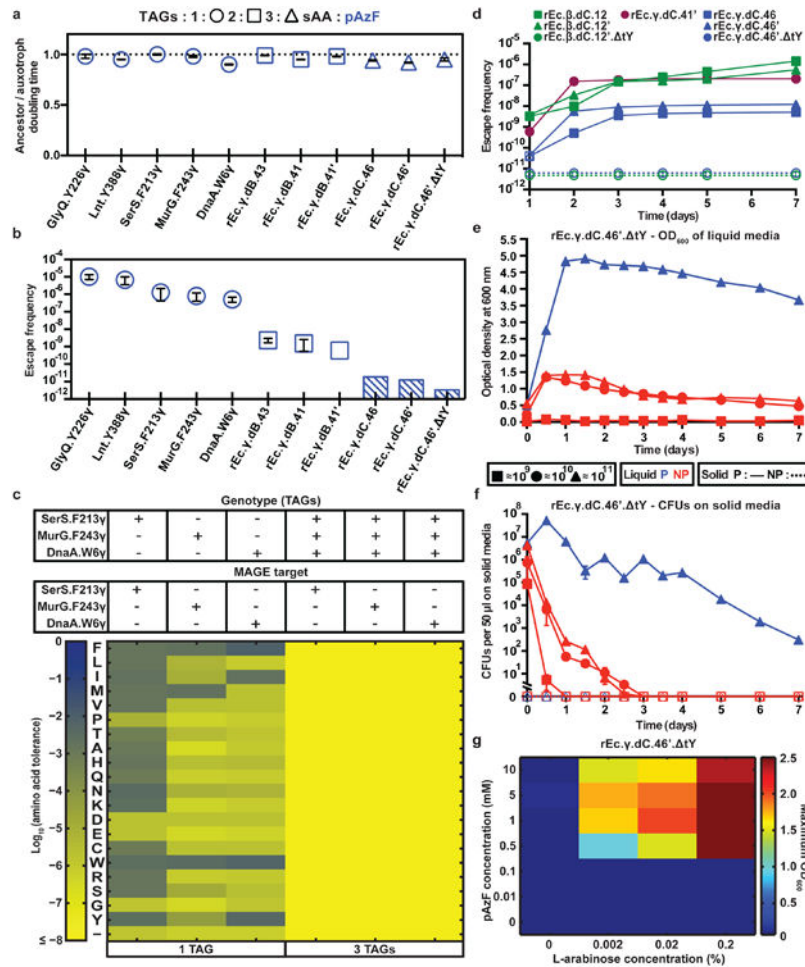


Figure 3. Characterization of strains dependent on sAA incorporation at active and dimerization sites in essential proteins

a, Doubling time (DT) ratios for the non-contained ancestor to pAzF auxotroph with one or more TAGs at functional loci calculated from growth in 5 mM pAzF and 0.2% L-arabinose. **b**, Escape frequencies (EFs) of strains in part a; bars represent EFs below the detection limit; average EFs are plotted. **c**, Representative assay surveying tolerance of TAG loci to 20 amino acids in different synthetic auxotrophs and expressed as log₁₀ of total cell survival; + indicates a TAG codon at the locus in the background strain and - indicates the wild type codon; blue and yellow indicate high and low tolerance to substitution, respectively; see methods. **d**, Representative escape assay monitoring EFs up to seven days after plating on solid nonpermissive media; hollow symbols/dashed lines, no observed EMs; see methods. **e**, Temporal monitoring of permissive (P, blue) and nonpermissive (NP, red) cultures inoculated with ~10⁹, 10¹⁰, or 10¹¹ cells of rEc.γ.dC.46'. tY by OD₆₀₀ and **f**, associated CFUs as sampled on P (solid lines) or NP (dashed lines) solid media; CFUs were never observed on NP solid media; hollow data points indicate no observed CFUs. **g**, Maximum OD₆₀₀ values during growth in LB across a concentration gradient of pAzF and L-arabinose. For all plots, average values of 3 technical replicates are plotted with error bars, ±s.d.

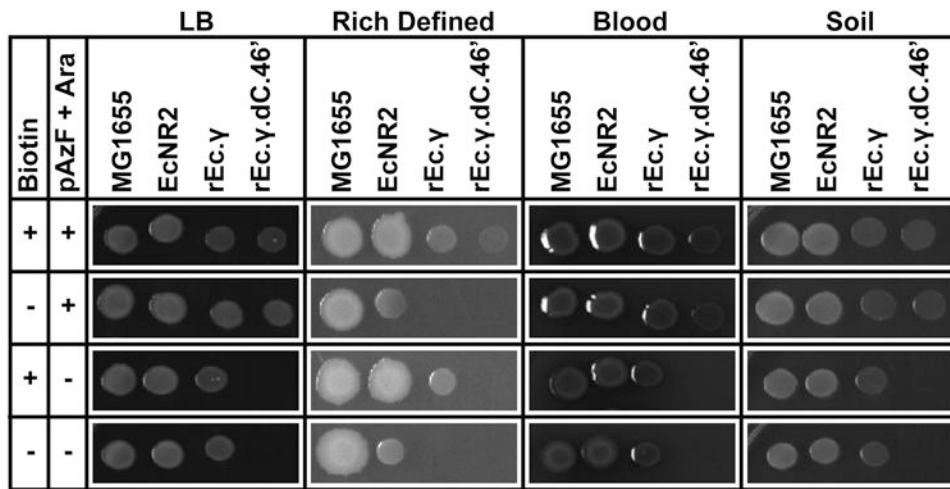


Figure 4. Investigating the viability of synthetic auxotrophs on diverse media types
Rescue by cross-feeding shown through spotting on diverse media types +/- pAzF/-arabinose and biotin supplementation; EcNR2, rEc.γ, and rEc.γ.dC.46' are auxotrophic for biotin (*bioA/B*) and rEc.γ.dC.46' is also a pAzF auxotroph.