ARTICLE

doi:10.1038/nature14121

Biocontainment of genetically modified organisms by synthetic protein design

Daniel J. Mandell[1]*, Marc J. Lajoie[1,2]*, Michael T. Mee[1,3], Ryo Takeuchi[4], Gleb Kuznetsov[1], Julie E. Norville[1], Christopher J. Gregg[1], Barry L. Stoddard[4] & George M. Church[1,5]

Genetically modified organisms (GMOs) are increasingly deployed at large scales and in open environments. Genetic biocontainment strategies are needed to prevent unintended proliferation of GMOs in natural ecosystems. Existing biocontainment methods are insufficient because they impose evolutionary pressure on the organism to eject the safeguard by spontaneous mutagenesis or horizontal gene transfer, or because they can be circumvented by environmentally available compounds. Here we computationally redesign essential enzymes in the first organism possessing an altered genetic code (*Escherichia coli* strain C321.ΔA) to confer metabolic dependence on non-standard amino acids for survival. The resulting GMOs cannot metabolically bypass their biocontainment mechanisms using known environmental compounds, and they exhibit unprecedented resistance to evolutionary escape through mutagenesis and horizontal gene transfer. This work provides a foundation for safer GMOs that are isolated from natural ecosystems by a reliance on synthetic metabolites.

GMOs are rapidly being deployed for large-scale use in bioremediation, agriculture, bioenergy and therapeutics[1]. In order to protect natural ecosystems and address public concern it is critical that the scientific community implements robust biocontainment mechanisms to prevent unintended proliferation of GMOs. Current strategies rely on integrating toxin/antitoxin 'kill switches'[2,3], establishing auxotrophies for essential compounds[4], or both[5,6]. Toxin/antitoxin systems suffer from selective pressure to improve fitness through deactivation of the toxic product[7,8], while metabolic auxotrophies can be circumvented by scavenging essential metabolites from nearby decayed cells or cross-feeding from established ecological niches. Effective biocontainment strategies must protect against three possible escape mechanisms: mutagenic drift, environmental supplementation and horizontal gene transfer (HGT). Here we introduce 'synthetic auxotrophy' for non-natural compounds as a means to biological containment that is robust against all three mechanisms. Using the first genomically recoded organism (GRO)[9] we assigned the UAG stop codon to incorporate a non-standard amino acid (NSAA) and computationally redesigned the cores of essential enzymes to require the NSAA for proper translation, folding and function. X-ray crystallography of a redesigned enzyme shows atomic-level agreement with the predicted structure. Combining multiple redesigned enzymes resulted in GROs that exhibit markedly reduced escape frequencies and readily succumb to competition by unmodified organisms in non-permissive conditions. Whole-genome sequencing of viable escapees revealed escape mutations in a redesigned enzyme and also disruption of cellular protein degradation machinery. Accordingly, reducing the activity of the NSAA aminoacyl-tRNA synthetase in non-permissive conditions produced double- and triple-enzyme synthetic auxotrophs with undetectable escape when monitored for 14 days (detection limit 2.2×10^{-12} escapees per colony forming unit (c.f.u.)). We additionally show that while bacterial lysate supports the growth of common metabolic auxotrophs, the environmental absence of NSAAs prevents such natural products from sustaining synthetic auxotrophs. Furthermore, distributing redesigned enzymes throughout the genome reduces susceptibility to HGT. When our GROs incorporate sufficient foreign DNA to overwrite the NSAA-dependent enzymes, they also revert UAG function, thereby preserving biocontainment by deactivating recoded genes. The general strategy developed here provides a critical advance in biocontainment as GMOs are considered for broader deployment in open environments.

Computational design of synthetic auxotrophs

We focused on the NSAA L-4,4′-biphenylalanine (bipA), which has a size and geometry unlike any standard amino acid, and a hydrophobic chemistry expected to be compatible with protein cores. We introduced a plasmid containing a codon-optimized version of the bipA aminoacyl-tRNA synthetase (*bipARS*)/$tRNA_{bipA}$ system[10] into a GRO (genomically recoded *E. coli* strain C321.ΔA (ref. 9)), thereby assigning UAG as a dedicated codon for bipA incorporation. Using a model of bipA in the Rosetta software for macromolecular modelling[11] we applied our computational second-site suppressor design protocol to 13,564 core positions in 112 essential proteins[12] with X-ray structures (Methods). We refined designs for cores that tightly pack bipA while maximizing neighbouring compensatory mutations predicted to destabilize the proteins in the presence of standard amino acid suppressors at UAG positions (Fig. 1a). We further required that candidate enzymes produce products that cannot be supplemented by environmentally available compounds. For example, we rejected *glmS* designs because glucosamine supplementation rescues growth of *glmS* mutants[13]. We selected designs of six essential genes for experimental characterization: adenylate kinase (*adk*), alanyl-tRNA synthetase (*alaS*), DNA polymerase III subunit delta (*holB*), methionyl-tRNA synthetase (*metG*), phosphoglycerate kinase (*pgk*) and tyrosyl-tRNA synthetase (*tyrS*). For all cases we designed oligonucleotides (Supplementary Table 1) encoding small libraries suggested by the computational models (Supplementary Table 2) and used them to directly edit the target essential gene in C321.ΔA using co-selection multiplex automated genome engineering (CoS–MAGE)[14]. Since *tyrS* featured the greatest number of compensatory mutations, we additionally

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Program in Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA. [4]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [5]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

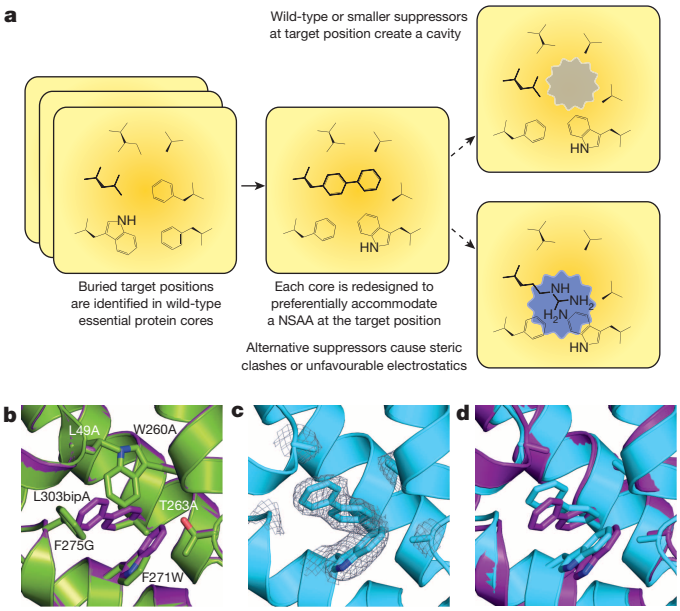


Figure 1 | Computational design of NSAA-dependent essential proteins. **a**, Overview of the computational second-site suppressor strategy. **b**, Computational design of a NSAA-dependent tyrosyl-tRNA synthetase (purple) overlaid on the wild-type structure (green; PDB code 2YXN). Six substituted residues are shown in stick representation. **c**, X-ray crystallography of the redesigned synthetase with an electron density map ($2F_o - F_c$ contoured at 1.0σ) for substituted residues; substitution F236A is on a disordered loop and is not observed. **d**, The crystal structure of the redesigned enzyme (cyan) superimposed onto the computationally predicted model (purple).

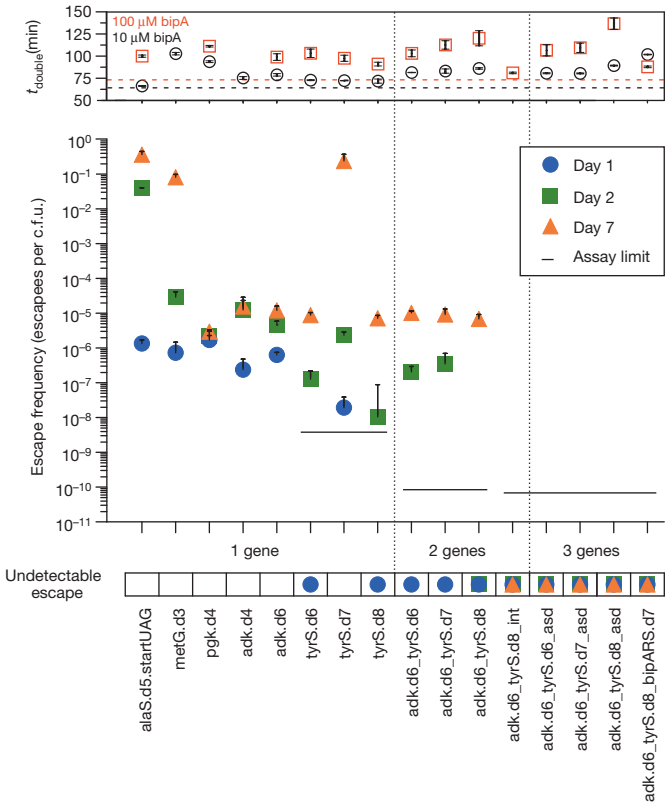


Figure 2 | Escape frequencies and doubling times of auxotrophic strains. Escape frequencies are shown for engineered auxotrophic strains calculated as colonies observed per c.f.u. plated over three technical replicates on solid media lacking arabinose and bipA. Assay limit is calculated as 1/(total c.f.u. plated) for the most conservative detection limit of a cohort, with a single-enzyme auxotroph limit of 3.5×10^{-9} escapees per c.f.u., a double-enzyme auxotroph limit of 8.3×10^{-11} escapees per c.f.u. and a triple-enzyme auxotroph limit of 6.41×10^{-11} escapees per c.f.u. Positive error bars represent the s.e.m. of the escape frequency over three technical replicates (Methods). The top panel presents the doubling times for each strain in the presence of 10 μM or 100 μM bipA, with the parental strain doubling times represented by the dashed horizontal lines. No marker indicates undetectable growth. Positive and negative error bars represent the s.e.m.

synthesized eight computational *tyrS* designs and used them to replace the endogenous *tyrS* gene (Supplementary Table 3). We screened our CoS–MAGE populations for bipA-dependent clones by replica plating from permissive media (containing bipA and arabinose for *bipARS* induction) to non-permissive media (lacking bipA and arabinose) and validated candidates by monitoring kinetic growth in the presence and absence of bipA (Methods and Extended Data Fig. 1). Mass spectrometry confirmed the specific incorporation of bipA in redesigned enzymes (Methods and Extended Data Fig. 2). X-ray crystallography of a redesigned enzyme at 2.65 Å resolution (Protein Data Bank (PDB)[15] code 4OUD, Extended Data Table 1) shows atomic-level agreement with computational predictions (Fig. 1b–d, Extended Data Fig. 3 and Supplementary Discussion). Selectivity for bipA in a redesigned core was further confirmed by measuring soluble protein content when bipA is mutated to leucine (wild-type residue) or tryptophan (most similar natural residue to bipA by mass) (Methods and Extended Data Fig. 4).

Characterization of synthetic auxotrophs

We characterized the escape frequencies of eight strains by plating on non-permissive media with and without *bipARS* inducer arabinose (Fig. 2, Supplementary Tables 4, 5 and Methods). Escapees exhibiting varying fitness were detected by the emergence of colonies in the absence of bipA. Two *tyrS* variants (tyrS.d6 and tyrS.d7) and two *adk* variants (adk.d4 and adk.d6) showed robust growth in permissive conditions and low escape frequencies in the absence of bipA. Strain alaS.d5 showed only minor impairment in the absence of bipA, suggesting that near-cognate suppression of the UAG codon by endogenous tRNA or mischarging of natural amino acids by BipARS is adequate to support growth. Consistent with this hypothesis, inserting a UAG immediately after the start codon (strain alaS.d5.startUAG) further impairs growth in the absence of bipA, although bipA dependence is readily overcome by mutational escape. *HolB* recombinants presented only the designed bipA mutation (holB.d1) and none of the compensatory mutations, suggesting that the intended compensation may be too destabilizing, or that the native amino acids at those positions may be critical for function. The lack of compensation for bipA results in a strong and continuous selective pressure to incorporate standard amino acids at the bipA position, so holB.d1 was not carried forward.

We hypothesized that since the designed proteins have structurally distinct cores, each variant may favour different standard amino acids at the bipA position. Therefore, viable UAG suppressors for one enzyme may be deleterious for another. We sought to determine the distribution of standard amino acids accommodated at the UAG position in each variant to identify combinations of redesigned enzymes that could drive escape frequencies even lower. We cultured the top seven dependent strains in permissive media and used MAGE[16] to introduce all 64 codons at the UAG positions (Methods). For tyrS.d7 we collaterally introduced the V307A mutation observed in tyrS.d6, since the same oligonucleotide containing V307A was used to encode degeneracy for both strains at the UAG position, producing the eighth strain tyrS.d8. Immediately following electroporation, cells were shifted to non-permissive media so that recombinants with canonical amino acids replacing bipA would overtake the population according to their relative fitness.

We sampled the eight populations at 1-h and 4-h time points, at confluence, and at two subsequent passages to confluence (100-fold dilution in each passage), by which point the preferred genotypes emerged. Using next-generation sequencing we determined the relative abundance of all standard amino acid codons at the UAG positions for each time point (Extended Data Fig. 5 and Supplementary Table 6) and computed

the 'flatness' (Shannon entropy[17]) of each amino acid frequency distribution (Fig. 3). The two strains showing the greatest escape frequencies, alaS.d5 and metG.d3, also have the flattest amino acid frequency distributions. Correspondingly, the strains with the lowest escape frequencies exhibit peaked amino acid frequency distributions. These amino acid preference profiles show a strong relationship between structural selectivity for bipA and escape frequency, supporting the rationale underlying our computational strategy. Furthermore, they confirm our hypothesis that different redesigned protein cores will favour different standard amino acids. In particular, phenylalanine and tryptophan (aromatics) dominate tyrS.d7 and tyrS.d8 populations, whereas the other recombinants tend towards valine, leucine, isoleucine and methionine (aliphatics) (Fig. 3a). In agreement with these observations, we were able to isolate viable recombinants of adk.d6 containing leucine but not tryptophan at the bipA position, while also isolating viable recombinants of tyrS.d8 containing tryptophan but not leucine at the bipA position (Supplementary Table 7). In considering candidates for combination, we omitted *alaS* and *metG* due to their susceptibility for near-cognate suppression. We also determined that *pgk* mutants can grow robustly in the presence of pyruvate and/or succinate (Extended Data Fig. 6) even though they do not grow in lysogeny broth Lennox (LB^L)[12]. Since these carbon sources are environmentally available, *pgk* violates our definition of essentiality and we removed pgk.d4 from consideration. Finally, we excluded adk.d4 due to its poor survival at stationary phase (Supplementary Table 8, Supplementary Discussion). We therefore focused on combinations of tyrS.d6, tyrS.d7 and tyrS.d8 with adk.d6, all of which maintain robust growth in permissive media, show strong dependence for bipA, and are metabolically isolated from environmental compounds.

Combining *tyrS* designs with adk.d6 yielded three strains with no detectable escapees after 24 h, including adk.d6_tyrS.d8, which has undetectable growth after >72 h (detection limit 7.44×10^{-11} escapees per c.f.u., Fig. 2 and Supplementary Table 4). Colonies bearing the adk.d6_tyrS.d8 genotype were observed between 4 and 7 days of incubation, but showed severely impaired fitness when grown in non-permissive liquid culture and were readily outcompeted by prototrophic *E. coli* (Fig. 4 and Supplementary Discussion). The relative reductions in escape frequencies support the hypothesis that combining variants with distinct amino acid preferences at the UAG position decreases fitness of escapees. Even though tyrS.d6 and tyrS.d8 exhibit similar escape frequencies as single-enzyme auxotrophs, strain adk.d6_tyrS.d6 produces faster growing escapees (adk.d6 and tyrS.d6 share a preference for leucine) than strain adk.d6_tyrS.d8 (tyrS.d8 prefers tryptophan). Although tyrS.d7 and tyrS.d8 both prefer aromatic residues, tyrS.d7 exhibits a broader amino acid preference profile (Fig. 3a) and produces faster-growing escapees than tyrS.d8 (Fig. 2). Accordingly, adk.d6_tyrS.d8 yields the lowest escape frequency of the combined *tyrS* and *adk* variants.

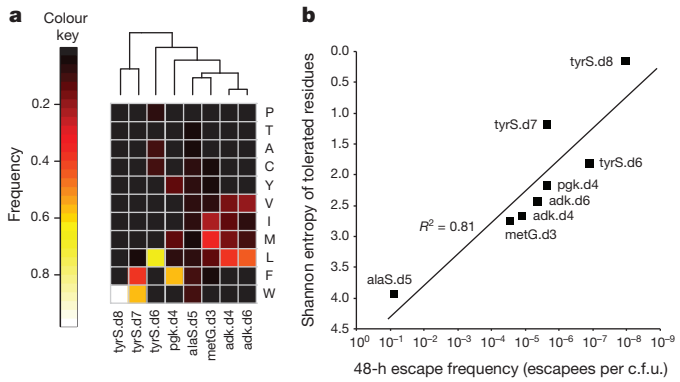


Figure 3 | Structural specificity at designed UAG positions in eight NSAA-dependent enzymes correlates with escape frequencies. **a**, Amino acid preferences at UAG positions in eight synthetic auxotrophs were determined by replacing the UAG codon with full NNN degeneracy and then sequencing the resulting populations with an Illumina MiSeq. Frequencies of each amino acid as a fraction of total sequences observed after three 1:100 passages to confluence are shown (top 11 most frequent amino acids only). Samples are clustered by Euclidean distance between amino acid frequencies. The frequency of an amino acid reports on the fitness conferred by the corresponding natural amino acid suppressor at the UAG position relative to all other amino acids. **b**, Shannon entropy was computed over the distributions of amino acids preferred at the UAG positions of the eight single-enzyme auxotrophs and plotted against the 48-h escape frequency for each strain. Entropy correlates log-linearly with escape frequency, suggesting that enzyme cores with high structural specificity for bipA at the UAG position have fewer fit evolutionary routes to escape. Strains alaS.d5 and metG.d3 have a deactivated *mutS* gene.

Prevention of mutagenic escape

The appearance of escapee colonies from adk.d6_tyrS.d8 after >72 h suggests the emergence of rare genotypes conferring weak viability (doubling time ≥348 min) in the absence of bipA. To uncover mutagenic routes to escape we performed whole-genome sequencing on escapees of adk.d6, tyrS.d8 and adk.d6_tyrS.d8 (Methods, four escapees for each single-enzyme auxotroph and three escapees for the double-enzyme auxotroph were sequenced). We observed no mutations in the ribosome or tRNAs that could account for UAG translation in the absence of bipA, nor did we observe mutations to any designed amino acid positions. However, we identified a point mutation (A70V) to *tyrS* in all four tyrS.d8 escapees sequenced (Supplementary Table 9). The A70V mutation may improve packing of the tyrS.d8 catalytic domain in the context of a destabilized neighbouring helical bundle lacking bipA (Extended Data Fig. 7a). To validate this escape mechanism we produced strain tyrS.d8.A70V and performed an escape assay on non-permissive media. Within 5 days of plating, we observed colony formation from all plated cells (Extended Data Fig. 7b), confirming that A70V is an escape mechanism for tyrS.d8. The A70V mutant of tyrS.d8 does not impair fitness in permissive conditions (Supplementary Table 10), so the genotype spontaneously arises as a neutral mutation within the fitness landscape by replication errors. However, targeted sequencing of the *tyrS* gene in eight additional tyrS.d8 escapees did not reveal the A70V mutation, suggesting that A70V is not the only escape mechanism for tyrS.d8.

Whole-genome sequencing of adk.d6 and adk.d6_tyrS.d8 escapees revealed disruptive mutations to Lon protease in all seven cases. One clone contained a frame shift and another contained a non-synonymous substitution (L611P) within the *lon* gene. The remaining five cases had a transposable element inserted within the promoter of *lon*. Targeted

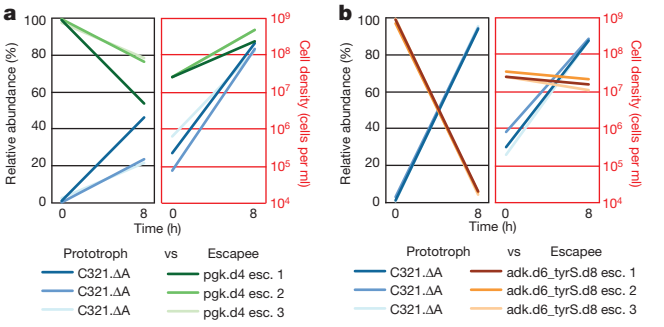


Figure 4 | Competition between synthetic auxotroph escapees and prototrophic *E. coli*. C321.ΔA was competed in the absence of bipA against escapees from a single-enzyme bipA auxotroph (pgk.d4, moderate NSAA dependence), or from a double-enzyme bipA auxotroph (adk.d6_tyrS.d8, strong bipA-dependence). Populations were seeded with 100-fold excess escapees and grown for 8 h in non-permissive conditions. The populations were evaluated using flow cytometry for episomally expressed fluorescent proteins at $t = 0$ and $t = 8$ h. Results from separate competition experiments against three different escapees are shown for each synthetic auxotroph. **a**, pgk.d4 escapees continue to expand in a mixed population with C321.ΔA after 8 h. **b**, adk.d6_tyrS.d8 escapees are rapidly outcompeted by C321.ΔA, which overtakes the population after 8 h.

sequencing characterized the insertion sequence in at least one clone as IS186, exactly recapitulating the Lon protease deficiency of *E. coli* BL21[18]. We validated Lon disruption as an escape mechanism using λ Red-mediated recombination to replace *lon* with a kanamycin resistance gene (kan^R) in adk.d6, tyrS.d8 and adk.d6_tyrS.d8. Recombinants were replica plated from permissive to non-permissive media containing kanamycin. Colony PCR confirmed that 27 of 27 non-bipA-dependent colonies screened (9 escapees per dependent strain) had Lon deleted by kan^R.

Since the Lon protease is the primary apparatus for bulk degradation of misfolded proteins in the *E. coli* cytoplasm[19], we hypothesized that its disruption would allow the persistence of poorly folded adk.d6 and tyrS.d8 proteins when standard amino acids are incorporated in place of bipA. We further hypothesized that basal UAG suppression from the promiscuous activity of pEVOL-BipARS produced sufficient full-length protein to support viability in the absence of Lon-mediated degradation. To test this hypothesis and safeguard against Lon-mediated escape we pursued two independent strategies to reduce the activity of BipARS in non-permissive conditions. First, we reduced the gene copy number approximately tenfold by moving *bipARS* and $tRNA_{bipA}$ from the p15A pEVOL plasmid to the genome of adk.d6, producing the strain adk.d6_int (Methods). Second, we applied our computational second-site suppressor strategy to residue V291 in *bipARS* (homologous to L303bipA in our *tyrS* designs) and reintroduced it into adk.d6 on the pEVOL vector, producing strain adk.d6_bipARS.d7 (Methods). This latter strategy produced a BipARS variant that requires bipA for folding and function, thereby abrogating residual activity towards standard amino acids in the absence of bipA. Both strategies resulted in a >200-fold reduction in 7-day escape frequency (Supplementary Table 4). Introducing tyrS.d8 to these strains produced double- and triple-enzyme synthetic auxotrophs adk.d6_tyrS.d8_int and adk.d6_tyrS.d8_bipARS.d7 that exhibited undetectable escape when monitored for 14 days (Fig. 2 and Supplementary Table 4, detection limit 2.2×10^{-12} escapees per c.f.u.). Both strains also showed undetectable escape in the presence of arabinose (Supplementary Table 5), and presented no fitness impairment relative to the parental adk.d6_tyrS.d8 synthetic auxotroph (Supplementary Table 4, doubling times of 57 and 55 min).

Protection from natural supplementation

To compare synthetic auxotrophy to current biocontainment practices we generated natural metabolic auxotrophs by knocking out *asd* and *thyA* genes from an MG1655-derived *E. coli* strain (EcNR1). The *asd* knockout renders the strain dependent on diaminopimelic acid (DAP) for cell-wall biosynthesis[4], while the *thyA* knockout deprives the cell of thymine, an essential nucleobase[20]. These well studied auxotrophies are commonly incorporated into biocontainment strategies[4,6]. In agreement with previous studies[4,6], the *asd* knockout shows strong dependence on its requisite metabolite, with a 7-day escape frequency of 8.97×10^{-9} escapees per c.f.u. (Supplementary Table 11). Knocking out *thyA* from this strain to produce a double-enzyme auxotroph did not reduce the 7-day escape frequency (8.79×10^{-9} escapees per c.f.u.). Nevertheless, metabolic strategies could complement synthetic auxotrophies to improve escape frequencies in defined ecological niches. To test this principle we knocked out *asd* from the double-enzyme synthetic auxotrophs of *adk* and *tyrS* resulting in three triple-enzyme auxotrophs (adk.d6_tyrS.d6_asd, adk.d6_tyrS.d7_asd and adk.d6_tyrS.d8_asd) that grew robustly in permissive conditions but showed undetectable escape after 7 days on media lacking bipA and DAP (Fig. 2 and Supplementary Tables 4 and 5, detection limit 6.4×10^{-11} escapees per c.f.u.).

While bacterial growth assays are often carried out in variations of media enriched with yeast extract, GMOs are increasingly deployed among a diversity of ecosystems that may provide opportunities for scavenging or cross-feeding essential metabolites. To compare metabolic and synthetic auxotroph strategies in an environment mimicking endogenous bacterial communities we grew engineered variants of both natural and synthetic auxotrophs in LB^L containing *E. coli* lysate (Methods). We hypothesized that since DAP is an essential component of the bacterial cell wall, the *Δasd* strains may scavenge sufficient DAP from *E. coli* lysate to complement the auxotrophy. As anticipated, metabolic auxotrophs obtained sufficient nutrients from the yeast/tryptone (LB^L) and the bacterial remnants (lysate) to support exponential growth (Extended Data Fig. 6e–h), while the synthetic auxotrophs failed to circumvent their dependencies. These results highlight the importance of establishing auxotrophies for compounds that are not environmentally available, and of ensuring the metabolic essentiality of enzymes intended to confer dependence.

Resistance to horizontal gene transfer

HGT is an important mechanism of evolution in any genetically rich environment[21]. We developed a conjugation escape assay to assess how DNA transfer within an ecosystem enables a GMO to escape biocontainment. Whereas any recombination event that replaces an inactivated gene could overcome metabolic auxotrophies[22], we hypothesized that conjugal escape would be disfavoured in GROs because donor DNA replacing bipA-dependent genes would also overwrite crucial genetic elements involved in genetic code reassignment (Fig. 5a). For example, reintroducing UAG stop codons into essential genes without restoring release-factor-1-mediated translational termination could be deleterious[9] or lethal[23]. Furthermore, reintroducing release factor 1 would result in competition between bipA incorporation and translational termination, undermining the recoded functions of the GRO.

To simulate a worst-case scenario in ecosystems containing a rich source of conjugal donors, we used Tn5 transposition to integrate an origin of transfer (*oriT*) into a population of *E. coli* MG1655 conjugal donor strains. We isolated a population of ~450 independent clones (one *oriT* for every ~10 kilobase portion of the 4.6 megabase (Mb) genome) and sequenced the flanking genomic regions of 96 donor colonies to confirm that *oriT* integration was well distributed throughout the population. We then conjugated this donor population into our auxotrophic strains at a ratio of 1 donor to 100 recipients to increase the probability that conjugal transfer will initiate from one *oriT* position per recipient. Conjugation was performed for durations of 50 min and 12 h (average conjugation times predicted to transfer 0.5 and 7.2 genomes) to simulate a single conjugal interaction and an ecological worst-case scenario, respectively. Conjugal escapees were selected on non-permissive media, and 23 alleles distributed throughout the genome (Fig. 5a) were screened using multiplex allele-specific colony PCR (mascPCR) to assess how much of the recoded genome is replaced by wild-type donor DNA.

Conjugal escape frequency decreases as the number of auxotrophic gene variants increases (Fig. 5b, top panel and Extended Data Fig. 8), consistent with larger portions of the genome that must be overwritten for conjugal escape of the multi-enzyme auxotrophs (Fig. 5b, bottom panel). The 12-h conjugations effect higher escape frequencies than do the 50-min conjugations, and the 12-h conjugations produce a larger diversity of conjugal escape genotypes, consistent with an increased opportunity to initiate new conjugal transfers during the mating period. Encouragingly, all 50-min conjugal escapees from multi-enzyme auxotrophs exhibit the wild-type donor sequence at all 23 assayed alleles (Fig. 5b, bottom panel and Supplementary Table 12), resulting in the reintroduction of release factor 1 and its UAG-mediated translational termination function. This collateral replacement of recombinant genomic DNA could be extended to other recombinant payloads such as toxins, antibiotic resistance genes, catabolic and genome editing enzymes, and orthogonal aminoacyl-tRNA synthetase/tRNA pairs used for NSAA incorporation.

Discussion

Effective biocontainment mechanisms for GMOs should place high barriers between modified organisms and the natural environment. Our NSAA design strategy produces organisms with an altered chemical language that isolates them from natural ecosystems. By conferring dependence on synthetic metabolites at the level of protein translation,

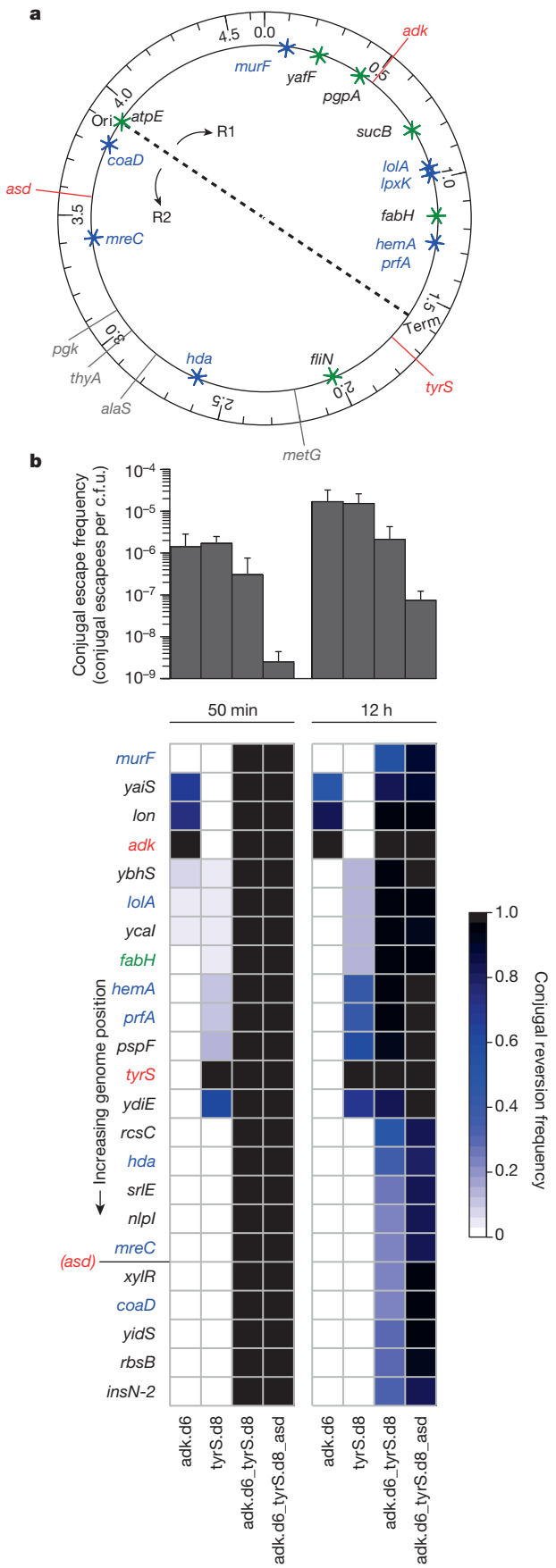


Figure 5 | Synthetic auxotrophy and genomic recoding reduce HGT-mediated escape. **a**, The positions of key alleles are plotted to scale on the genome schematic. Red lines indicate auxotrophies used in the multi-enzyme auxotrophs and grey lines indicate other auxotrophies that were not included in this assay. Asterisks indicate important alleles associated with the reassignment of UAG translation function (blue are essential genes and green are potentially important genes[9]). Conjugation-mediated reversion of the UAA codons back to the wild-type UAG is expected to be deleterious unless the natural UAG translational termination function is reverted. R1 and R2 denote replicores 1 and 2, respectively. **b**, Combining multiple synthetic auxotrophies in a single genome requires a large portion of the genome to be overwritten by wild-type donor DNA, reducing the frequency of conjugal escape (top panel) and increasing the likelihood of overwriting the portions of the genome (bottom panel) that provide expanded biological function (for example, *prfA* encodes release factor 1, which mediates translational termination at UAG codons). Positive error bars indicate standard deviation.

folding and function, synthetic auxotrophy addresses the need for GMOs that are refractory to mutational escape, metabolic supplementation and HGT. Because our NSAAs are incorporated into essential enzymes with second-site mutations, escapees are rare and are unfit to out-compete prototrophic microbial communities. In part, robustness emerges from simplicity: our most escape-resistant synthetic auxotrophs contain only 32 (adk.d6_tyrS.d8_int) and 49 (adk.d6_tyrS.d8_bipARS.d7) base pair substitutions across the 4.6 Mb parental genome and *bipARS*, with no essential genes deleted or toxic products added. Furthermore, NSAA-based biocontainment with bipA only modestly increases the cost per litre of proliferating culture (Extended Data Table 2).

This work highlights the delicate balance required to engineer essential proteins that are conditionally stabilized by a single NSAA. The design must confer sufficient instability in non-permissive conditions to deactivate the protein, while providing functional stability in the presence of the correct NSAA. Future design strategies could include polar or charged NSAAs to engineer hydrogen bonds requiring exquisite geometric specificity[24] for folding, enzyme–substrate interactions, or macromolecular associations. This approach may reduce susceptibility to suppressors, although fewer protein microenvironments may accommodate the burial of charged or polar residues. Reassigning additional codons would permit the incorporation of multiple NSAAs that confer dependence either in different structural motifs or in participation of a joint chemistry. Eventually, organisms with orthogonal genomic chemistries including expanded genetic alphabets[25] and their associated replication machinery could provide additional layers of isolation[26].

Our results demonstrate that mutational escape frequency under laboratory growth conditions is a necessary but insufficient metric to evaluate biocontainment strategies. Many genes considered to be essential have functions that can be complemented by environmental compounds, as demonstrated here for auxotrophies of natural (*asd*, *thy*) and designed (*pgk.d4*) enzymes. Furthermore, localizing biocontainment mechanisms to a small portion of the genome increases susceptibility to escape by uptake of foreign DNA. Distributing multiple NSAA-dependent enzymes throughout a recoded genome acts as a genomic safeguard against escape by HGT, and demands that conjugal escape replaces large portions of the recipient genome. This collateral replacement of GMO genomic DNA could be exploited to delete recombinant payloads upon exposure to conjugal donors in the environment. Additionally, by recoding restricted payloads with essential UAG codons, they can be prevented from functioning in natural organisms. Therefore, the expanded genetic code of GROs can be exploited both to prevent their undesired survival in natural ecosystems and to block incoming and outgoing HGT with natural organisms.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 April; accepted 26 November 2014.
Published online 21 January 2015.

1. Moe-Behrens, G. H., Davis, R. & Haynes, K. A. Preparing synthetic biology for the world. *Front. Microbiol.* **4,** 5 (2013).
2. Molin, S. *et al.* Conditional suicide system for containment of bacteria and plasmids. *Nature Biotechnol.* **5,** 1315–1318 (1987).

3. Li, Q. & Wu, Y.-J. A fluorescent, genetically engineered microorganism that degrades organophosphates and commits suicide when required. *Appl. Microbiol. Biotechnol.* **82,** 749–756 (2009).
4. Curtiss, R., III. Biological containment and cloning vector transmissibility. *J. Infect. Dis.* **137,** 668–675 (1978).
5. Ronchel, M. C. & Ramos, J. L. Dual system to reinforce biological containment of recombinant bacteria designed for rhizoremediation. *Appl. Environ. Microbiol.* **67,** 2649–2656 (2001).
6. Wright, O., Delmans, M., Stan, G. B. & Ellis, T. GeneGuard: a modular plasmid system designed for biosafety. *ACS Synth. Biol.* http://dx.doi.org/doi:10.1021/sb500234s (13 May 2014).
7. Knudsen, S. *et al.* Development and testing of improved suicide functions for biological containment of bacteria. *Appl. Environ. Microbiol.* **61,** 985–991 (1995).
8. Pasotti, L., Zucca, S., Lupotto, M., Cusella De Angelis, M. G. & Magni, P. Characterization of a synthetic bacterial self-destruction device for programmed cell death and for recombinant proteins release. *J. Biol. Eng.* **5,** 8 (2011).
9. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342,** 357–360 (2013).
10. Xie, J., Liu, W. & Schultz, P. G. A genetically encoded bidentate, metal-binding amino acid. *Angew. Chem.* **46,** 9239–9242 (2007).
11. Renfrew, P. D., Choi, E. J., Bonneau, R. & Kuhlman, B. Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. *PLoS ONE* **7,** e32637 (2012).
12. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).
13. Wu, H. C. & Wu, T. C. Isolation and characterization of a glucosamine-requiring mutant of *Escherichia coli* K-12 defective in glucosamine-6-phosphate synthetase. *J. Bacteriol.* **105,** 455–466 (1971).
14. Carr, P. A. *et al.* Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res.* (2012).
15. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).
16. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).
17. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27,** 379–423 (1948).
18. saiSree, L., Reddy, M. & Gowrishankar, J. IS*186* insertion at a hot spot in the *lon* promoter as a basis for *lon* protease deficiency of *Escherichia coli* B: identification of a consensus target sequence for IS*186* transposition. *J. Bacteriol.* **183,** 6943–6946 (2001).
19. Tomoyasu, T., Mogk, A., Langen, H., Goloubinoff, P. & Bukau, B. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the *Escherichia coli* cytosol. *Mol. Microbiol.* **40,** 397–413 (2001).
20. Steidler, L. *et al.* Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin 10. *Nature Biotechnol.* **21,** 785–789 (2003).
21. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480,** 241–244 (2011).
22. Wollman, E. L., Jacob, F. & Hayes, W. Conjugation and genetic recombination in *Escherichia coli* K-12. *Cold Spring Harb. Symp. Quant. Biol.* **21,** 141–162 (1956).
23. Mukai, T. *et al.* Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res.* **38,** 8188–8195 (2010).
24. Kortemme, T., Morozov, A. V. & Baker, D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. *J. Mol. Biol.* **326,** 1239–1259 (2003).
25. Malyshev, D. A. *et al.* A semi-synthetic organism with an expanded genetic alphabet. *Nature* **509,** 385–388 (2014).
26. Schmidt, M. & de Lorenzo, V. Synthetic constructs in/for the environment: managing the interplay between natural and engineered Biology. *FEBS Lett.* **586,** 2199–2206 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Renfrew for help with NSAA modelling in Rosetta, D. Goodman and R. Chari for sequence analysis assistance, M. Napolitano for advice on Lon-mediated escape assays, J. Teramoto and B. Wanner for the pJTE2 jumpstart plasmid, and F. Isaacs for manuscript comments. D.J.M. is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. M.J.L. was supported by a US Department of Defense National Defense Science and Engineering Graduate Fellowship. M.T.M. was supported by a Doctoral Study Award from the Canadian Institutes of Health Research. The research was supported by Department of Energy Grant DE-FG02-02ER63445.

Author Contributions D.J.M., M.J.L., M.T.M. and G.M.C. conceived the project and designed the study, with D.J.M. as computational lead and M.J.L. as experimental lead. D.J.M. computationally designed synthetic auxotrophs, performed strain engineering, characterized escape frequencies and fitness of synthetic auxotrophs, performed western blot analyses and prepared samples for mass spectrometry and X-ray crystallography. M.J.L. performed strain engineering, performed site-saturation mutagenesis at UAG positions, performed whole-genome sequencing of escapees, validated escape mechanisms and assessed HGT by conjugation. M.T.M. measured escape frequencies and fitness of natural metabolic auxotrophs, performed competition assays and assessed HGT by conjugation. R.T. and B.L.S. crystallized tyrS.d7 and determined the X-ray structure. G.K. analysed whole-genome sequencing data of escapees. J.E.N. and C.J.G. developed the *tdk* selection protocol. D.J.M., M.J.L. and M.T.M. wrote the paper.

Author Information Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data Bank under accession number 4OUD. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.M.C. (gchurch@genetics.med.harvard.edu).

METHODS

Essential protein selection and buried residue determination. Candidate genes were selected by searching the Keio collection[12] of comprehensive single-gene *E. coli* K-12 knockouts for genes classified as essential. X-ray structures were identified by mapping essential gene GenBank[27] protein genInfo identifiers (PIDs) to PDB[15] entries through the UniProtKB[28]. In cases of multiple PDB entries the highest-resolution structure was selected. 112 high-resolution X-ray structures (resolution ≤2.8 Å) were analysed. Structures were pre-processed to remove alternative side-chain conformers (the first listed conformer was kept), to remove atoms without occupancy, to remove heteroatoms, to convert selenomethionines to methionines, and to remove chains other than the first listed chain of the essential protein. The solvent-accessible surface area (SASA) of each position in each candidate structure was calculated using the PyRosetta[29] interface to the Rosetta SasaCalculator class with a 1.0 Å probe radius (a radius smaller than 1.4 Å allowed finer sampling of spaces around candidate positions). Positions were considered buried if their SASA was not more than 20% of the residue-specific average SASA value from a 30-member random ensemble of Gly-X-Gly peptides, where X is the residue type, as determined by the GETAREA method[30,31]. The average SASA values are Ala, 64.9; Arg, 195.5; Asn, 114.3; Asp, 113.0; Cys, 102.3; Gln, 143.7; Glu, 141.2; His, 154.6; Ile, 147.3; Gly, 87.2; Leu, 146.2; Lys, 164.5; Met, 158.3; Phe, 180.1; Pro, 105.2; Ser, 77.4; Thr, 106.2; Trp, 224.6; Tyr, 193.1; Val, 122.3. By these criteria 13,564 residues in the data set were considered buried.

Design and refinement of NSAA-dependent proteins. The side chains of each structure were relaxed into local minima of the Rosetta forcefield by the Rosetta sidechain_min application (Rosetta command lines, below). Three separate design simulations were then carried out for each target buried position using RosettaDesign[32]. The first simulation sets the target position to L-4,4′-biphenylalanine (bipA), as implemented in Rosetta (residue type B30)[11], and sets the surrounding residues to either redesign (varies both amino acid identity and side-chain conformation) or repack (varies only the conformation of the wild-type amino acid). Residues with Cα atoms within 6 Å of the target position, or with Cα atoms within 8 Å of the target position and Cβ atoms closer than the Cα atom to the target position, were set to redesign. Residues with Cα atoms within 10 Å of the target position, or with Cα atoms within 12 Å of the target position and Cβ atoms closer than the Cα atom to the target position, were set to repack. All other side chains were fixed at their minimized coordinates, together with all backbone atoms. The resulting energy terms were appended with the target position SASA as calculated by the PyRosetta SasaCalculator with a 1.0 Å probe radius. We term the Rosetta scores of these designs 'compensated scores'. In the second simulation, the same calculation is performed, except all positions previously set to redesign are restricted only to repack. We term the resulting scores 'uncompensated scores'. The difference between the 'compensated score' and the 'uncompensated score' reports on the extent to which the target site must change to accommodate bipA. In the third simulation, the target position maintains its wild-type identity, all coordinates are fixed at the positions output by the sidechain_min application, and the structure is rescored using the same scoring parameters as the other two simulations (Rosetta command lines, below). We term the resulting scores 'wild-type scores'. The difference between the compensated score and the wild-type score reports on the predicted stability of the redesigned core relative to the wild-type structure.

The design goal was to obtain variants that are functionally stable with bipA at the designed position, while being maximally destabilized with a natural amino acid at the bipA position. Accordingly, designs were filtered for the following criteria:

(1) The minimized wild-type score must be less than 10 Rosetta energy units to ensure that the starting structure can be reasonably modelled with the Rosetta forcefield.

(2) The compensated score must be less than or equal to the wild-type score, to select for redesigned cores that do not destabilize the protein relative to the wild-type sequence.

(3) The uncompensated score must be greater than the wild-type score, to ensure that compensatory mutations are necessary to accommodate bipA.

(4) The compensated score must be less than the uncompensated score, to select for compensatory mutations that improve the stability of the core in the presence of bipA relative to the uncompensated mutant. This requirement also selects for sequences that reduce the fitness of suppressors at the compensatory positions.

(5) The SASA score must be <0.75, to select for cores that tightly pack around bipA, both to select for stability in the presence of bipA and to reduce the fitness of standard amino acids at the bipA or compensatory positions.

The designs for 16 engineered UAG sites in 12 enzymes meeting these criteria were then ranked by the difference between compensated score and uncompensated score, as a key metric for bipA dependence, and were further filtered by the following criteria based on known structural and functional data from the literature:

(1) The redesigned residues must not participate in ligand binding, catalysis or be required for allosteric signal transduction via conformational rearrangements.

(2) The product of the reaction must not be environmentally available.

(3) The product of the reaction must not be complementable by another environmentally available molecule.

Using these criteria 10 designs were subject to refinement. Positions to design, repack, or revert to wild-type were selected by visual inspection. A second round of fixed backbone design was then applied to generate 100 designs from each unrefined structure (Rosetta command lines, below). Designs from six enzymes were carried forward for experimental characterization. Frequently occurring mutations in the refined designs assessed by visual inspection were included in MAGE oligonucleotides (Supplementary Table 1). For *tyrS*, eight additional all-atom designs were encoded by PCR primers for gene assembly (Supplementary Table 3).

Rosetta command lines. All Rosetta calculations were performed with Rosetta version 48561.

Example command line for preparative side-chain minimization of scaffold structures:

sidechain_min.linuxgccrelease -database __ROSETTA_DATABASE__ -loops::input_pdb __PREFIX__.pdb -output_tag __PREFIX__ -ex1 -ex2 -overwrite

Example command line for wild-type score:

*score.linuxgccrelease -database __ROSETTA_DATABASE__ -l *.pdb -score:hbond_His_Phil_fix -in:file:fullatom -no_optH -no_his_his_pairE -score:weights mm_std*

Example command line for initial design for compensated score and uncompensated score:

fixbb.linuxgccrelease -database __ROSETTA_DATABASE__ -ex1 -ex2 -s __PREFIX__.pdb -resfile __PREFIX__.resfile -minimize_sidechains -score:weights mm_std -score::hbond_His_Phil_fix -no_his_his_pairE -nstruct 1 -out:pdb_gz -overwrite

Example command line for refinement of dependent designs:

fixbb.linuxgccrelease -database __ROSETTA_DATABASE__ -s 2YXNA_min.pdb -minimize_sidechains -score::hbond_His_Phil_fix -no_his_his_pairE -ex1 -ex2 -nstruct 100 -resfile __PREFIX__.resfile -overwrite

Culture and selection conditions. Growth media consisted of LB^L (10 g l^{-1} bacto tryptone, 5 g l^{-1} sodium chloride, 5 g l^{-1} yeast extract). Permissive growth media for bipA-dependent auxotrophs was LB^L supplemented with sodium dodecyl sulfate (SDS), chloramphenicol, bipA and arabinose. Non-permissive media lacked bipA and arabinose. The following selective agents, nutrients and inducers were used when indicated: chloramphenicol (20 μg ml^{-1}), kanamycin (30 μg ml^{-1}), spectinomycin (95 μg ml^{-1}), tetracycline (12 μg ml^{-1}), zeocin (10 μg ml^{-1}), gentamycin (5 μg ml^{-1}), SDS (0.005% w/v), vancomycin (64 μg ml^{-1}), colicin E1 (ColE1; ~10 μg ml^{-1}), DAP (75 μg ml^{-1}), thymidine (100 μg ml^{-1}), bipA (10 μM), glucose (0.2% w/v), pyruvate (0.2% w/v), succinate (0.2% w/v), arabinose (0.2% w/v), anhydrotetracycline (30 ng μl^{-1}). For strains adk.d6_tyrS.d8_bipARS.d7 and adk.d6_tyrS.d8_int permissive media contained 100 μM bipA. Permissive media for metabolic auxotrophs is LB^L supplemented with 75 μg ml^{-1} DAP and 100 μg ml^{-1} thymidine. TolC selections (SDS) and counter selections (colicin E1) were performed as previously described[33]. Tdk selections used LB^L supplemented with 20 μg ml^{-1} 2′-deoxy-5-fluorouridine and 100 μg ml^{-1} deoxythymidine, and counter selections used LB^L supplemented with 5 μM azidothymidine.

Strain engineering. Two strategies were undertaken to engineer redesigned essential proteins. Strains adk.d4, adk.d6, alaS.d5, holB.d1, metG.d3 and pgk.d4 were generated by performing CoS–MAGE[14] with designed single stranded oligonucleotide pools (Supplementary Table 1) and *tolC* co-selection[14]. Recombined populations were plated on permissive media, and then replica plated on non-permissive media to screen for bipA-dependent clones. Top candidates were identified by kinetic growth monitoring (Biotek H1 or H4 plate reader) of 10–40 bipA-dependent clones in permissive and non-permissive liquid growth media. Strains showing robust growth in permissive media and little to no growth in non-permissive media were carried forward. The tyrS.d6, tyrS.d7 and tyrS.d8 gene variants were constructed by PCR amplification of the *E. coli* MG1655 *tyrS* gene with mutagenic primers, followed by full-length Gibson assembly[34] (Supplementary Tables 1 and 3) and recombination onto the genome using λ Red recombineering[35,36]. Strains tyrS.d6, tyrS.d7, tyrS.d8, adk.d6_tyrS.d6, adk.d6_tyrS.d7 and adk.d6_tyrS.d8 were produced by (1) deleting the endogenous *tdk* gene from C321.ΔA, (2) replacing the endogenous *adk* and *tyrS* genes with their codon-shuffled variants (*adk*(recode)-*tdk* and *tyrS*(recode)-*tdk*, Supplementary Table 3) transcriptionally fused to *tdk*, and (3) replacing the fusion cassettes with the adk.d6, tyrS.d6, tyrS.d7, or tyrS.d8 variants. Variants of adk.d6, tyrS.d7 and tyrS.d8 containing leucine and tryptophan at the bipA position were constructed by MAGE with oligonucleotides containing the appropriate mutations and clonal populations were isolated on LB^L plates lacking bipA and arabinose. Triple-enzyme auxotrophs were created by replacing *asd* with a *Δasd::spec*^{R} cassette. We reactivated mismatch repair using mutS_null_revert-2* in the *pgk*, *adk* and *tyrS* single-enzyme auxotrophs and all of the multi-enzyme auxotrophs. For construction of the quadruplet $tRNA_{bipA}$ (Supplementary

Discussion) QuikChange was used to replace the CUA anticodon with UCUA. Quadruplet versions of adk.d6 and tyrS.d8 with UAGA at the bipA positions were constructed by PCR and Gibson assembly followed by λ Red-mediated recombination into the genome as described above. All genotypes (Supplementary Table 3) were confirmed using mascPCR[37] and Sanger sequencing using primers from Supplementary Table 1.

Strain doubling time analysis. Strain doubling times were calculated as previously described[9]. Briefly, cultures were grown in flat-bottom 96-well plates (150 µl LB^L, 34 °C, 300 r.p.m.). Kinetic growth (OD_{600}) was monitored on a Biotek H1 plate reader at 5-min intervals. Doubling times were calculated by $t_{double} = \Delta t \times \ln(2)/m$, where $\Delta t = 5$ min per time point and m is the maximum slope of $\ln(OD_{600})$ calculated from the linear regression through a sliding window of 5 contiguous time points (20 min intervals). For escapee strains exhibiting growth rates that were too slow for this analysis, doubling times were calculated by $t_{double} = \Delta t \times \ln(2)/\ln(P_2/P_1)$, where Δt represents sliding windows of 15 min and P_2/P_1 represents initial/final OD_{600} values for the window. Strains that exhibited doubling times greater than 900 min and/or maximum OD_{600} values less than 0.2 after the specified culture duration were considered to exhibit no growth ('none observed') for the given conditions. Improved aeration doubling times for strains adk.d6_tyrS.d8_int and adk.d6_tyrS.d8_bipARS.d7 were obtained by growing strains in 3 ml LB^L in 28 ml culture tubes (three tubes dedicated for each time point), measuring OD_{600} of three technical replicates in 1 cm cuvettes in a spectrophotometer (Beckman DU640) at 20 min intervals for 3.66 h, and determining the slope of the log transformed data over each 40 min window. Least doubling time (20 min × ln(2)/slope) and corresponding R^2 values are reported (Supplementary Table 4).

Expression and purification of tyrS.d7. C321.ΔA cells were grown to mid-log phase in LB^L and co-electrotransformed with 5 ng each of plasmid pEVOL-bipA[10] and an additional plasmid containing full-length tyrS.d7 as an amino-terminal GST fusion under an anhydrotetracycline (aTc)-inducible promoter. After 90 min of recovery cells were plated on LB^L agar supplemented with chloramphenicol and kanamycin. Single colonies were used to inoculate 2 ml starter cultures of LB^L supplemented with chloramphenicol and kanamycin that were grown overnight at 34 °C. Saturated overnight growths were diluted 1:100 into six 1 l cultures containing LB^L supplemented with chloramphenicol and kanamycin, which were grown at 34 °C with shaking at 250 r.p.m. to an OD_{600} of 0.6. The temperature was then reduced to 18 °C, and bipA was added to a final concentration of 500 µM. After an additional 60–90 min aTc and arabinose were added to final concentrations of 30 ng ml^{-1} and 0.2%, respectively. After 24 h of expression, cells were harvested by centrifugation at 10,000*g* and snap frozen in a dry ice and ethanol bath. Approximately 10 g of thawed cell pellet was suspended in 100 ml of Buffer A (20 mM Tris-HCl (pH 7.2), 500 mM NaCl, and 5% (v/v) glycerol) supplemented with 1 mg ml^{-1} lysozyme. After sonication (6 cycles of 30 s each), the cell lysate was centrifuged at 20,000*g* for 20 min at 4 °C, and the supernatant was mixed with 5 ml of polyethyleneimine (pH 7.9) on ice. After centrifugation again at 20,000*g* for 10 min at 4 °C, the supernatant was filtered through a 0.45 µm PVDF membrane, and suspended with 3 ml of glutathione sepharose 4B beads (GE Healthcare Life Sciences). The beads were extensively washed with Buffer A supplemented with 1 mM dithiothreitol (DTT) and incubated with 120 units of PreScission protease (GE Healthcare Life Sciences) at 5 °C for 16 h. The untagged protein was eluted and dialysed against Buffer B (20 mM Tris–HCl (pH 7.5), 50 mM NaCl, 10 mM 2-mercaptoethanol and 5% (v/v) glycerol) at 4 °C. The redesigned enzyme was concentrated to approximately 5.5 mg ml^{-1}.

Determination of the tyrS.d7 crystallographic structure. One-microlitre drops of protein were mixed with an equal volume of a reservoir solution containing 0.1 M sodium malonate (pH 5.5) and 18% (w/v) polyethylene glycol 3350. Crystals were grown at room temperature via hanging drop vapour phase diffusion, and then were transferred into 0.1 M sodium malonate (pH 5.5), 25% (w/v) polyethylene glycol 3350 and 15% ethylene glycol. Crystals were frozen in liquid nitrogen, and diffraction data were collected at the Advanced Light Source (ALS) beamline 5.0.1. The data were processed using the HKL2000 package[38]. The crystal structure of a truncated *E. coli* TyrS variant protein (PDB code 2YXN) was used as a search model in molecular replacement. The crystallographic model was built using COOT[39], refined using REFMAC5 and Crystallography and NMR system (CNS)[40], and deposited in the RCSB Protein Data Bank (PDB code 4OUD). Statistics of the data collection and refinement are provided in Extended Data Table 1.

Mass spectrometry of NSAA-dependent enzymes. Strains adk.d6, tyrS.d7 and tyrS.d8 were grown to mid-log phase in 10 ml of permissive media. Cell pellets were obtained and soluble lysate fractions were collected as above. Samples were normalized to 250 µg (adk.d6) or 50 µg (tyrS.d7 and tyrS.d8) total protein content and resolved by SDS–PAGE. Gel slices from each strain containing the enzyme (resolved by size comparison to a known standard) were digested with trypsin. Peptide sequence analysis of each digestion mixture was performed by microcapillary reversed-phase high-performance liquid chromatography coupled with nanoelectrospray tandem mass spectrometry (µLC-MS/MS) on a LTQ-Orbitrap Elite mass spectrometer (ThermoFisher Scientific, San Jose, CA). The Orbitrap repetitively surveyed an m/z range from 395 to 1,600, while data-dependent MS/MS spectra on the 20 most abundant ions in each survey scan were acquired in the linear ion trap. MS/MS spectra were acquired with relative collision energy of 30%, 2.5-Da isolation width, and recurring ions dynamically excluded for 60 s. Preliminary sequencing of peptides was facilitated with the SEQUEST algorithm with a 30 p.p.m. mass tolerance against the Uniprot Knowledgebase *E. coli* K-12 reference proteome supplemented with a database of common laboratory contaminants, concatenated to a reverse decoy database. Using a custom version of Proteomics Browser Suite (PBS v.2.7, ThermoFisher Scientific), peptide-spectrum matches were accepted with mass error <2.5 p.p.m. and score thresholds to attain an estimated false discovery rate of ~1%.

Western blot analysis of tyrS.d7 variant GST fusions. Cell pellets for all variants were obtained as described above, with an expression culture volume of 10 ml. Cells were lysed using B-PER Bacterial Protein Extraction Reagent, lysozyme (100 mg ml^{-1}), DNaseI (5,000 U ml^{-1}), and Halt Protease Inhibitor Cocktail (all Thermo Scientific) according to the manufacturer's specifications. Lysates were centrifuged at 15,000*g* for 5 min and the soluble fractions were collected. Protein concentration was determined fluorometrically using the Qubit Protein Assay Kit (Life Technologies). Lysates were normalized to 5-µg samples, resolved by SDS–PAGE, and electro-blotted onto PVDF membranes (Life Technologies, number IB24002). Western blotting was performed using an anti-GST mouse monoclonal primary antibody (Genscript, number A00865-40) and anti-GAPDH mouse monoclonal loading control antibody (Thermo Scientific, number MA5-15738) followed by secondary binding to a HRP-conjugated anti-mouse antibody (Thermo Scientific, number 35080). Samples were imaged by luminol chemiluminescence on a ChemiDoc system (BioRad) and protein content was quantified by densitometry and normalized to GAPDH.

Solid media escape assays for natural metabolic and synthetic auxotrophs. All strains were grown in permissive conditions and harvested in late exponential phase. Cells were washed twice in LB^L and resuspended in LB^L. Viable c.f.u. were calculated from the mean and standard error of the mean (s.e.m.) of three technical replicates of tenfold serial dilutions on permissive media. Three technical replicates were plated on non-permissive media and monitored for 7 days. The order of magnitude of cells plated ranged from 10^2 to 10^9 depending on the escape frequency of the strain. Synthetic auxotrophs were plated on two different non-permissive media conditions: SC, LB^L with SDS and chloramphenicol (Supplementary Table 4); and SCA, LB^L with SDS, chloramphenicol and 0.2% arabinose (Supplementary Table 5). Metabolic auxotrophs were plated on LB^L for non-permissive conditions (Supplementary Table 11). If synthetic auxotrophs exhibited escape frequencies above the detection limit (lawns) on SC at day 1, 2 or 7 (alaS.d5, metG.d3, tyrS.d7), escape frequencies for those days were calculated from additional platings at lower density. Additional platings at higher density were also used to obtain day 1 and day 2 escape frequencies for pgk.d4 on SC. The s.e.m. $S_{\bar{X}}$ across technical replicates of the cumulative escape frequency ν scored for a given day was calculated as: $S_{\bar{X}} = \nu\sqrt{\left(\frac{S_{\bar{X}}\tau}{\tau}\right)^2 + \left(\frac{S_{\bar{X}}n}{n}\right)^2}$, where τ is the mean number of c.f.u. plated, $S_{\bar{X}}\tau$ is the s.e.m. of c.f.u. plated, n is the mean cumulative colony count up to the given day, and $S_{\bar{X}}n$ is the s.e.m. of the cumulative colony count up to the given day. If synthetic auxotroph escapees emerged on SC, three clones were isolated, their growth rates were calculated as described above, and the doubling time of the fastest escapee was recorded (Supplementary Table 4).

Site saturation mutagenesis at designed UAG positions. To site-specifically replace UAG with all other codons, we used MAGE oligonucleotide pools that exactly matched the sequence of the bipA-dependent gene except that the UAG was replaced by all 64 NNN codons (Supplementary Table 1). This allowed us to assess which canonical amino acid substitutions resulted in the best survival of synthetic auxotroph escapees. Although some of these amino acid substitutions may be unlikely to be evolutionarily sampled (evolution will favour amino acids with many tRNA gene copies and whose cognate codons are a single mutation from UAG[41]), this unbiased strategy avoided missing mechanisms of tRNA suppression that are not yet characterized. Immediately after introducing NNN codon diversity via MAGE[16], we recovered the cell populations in 1 ml of LB^L without supplementing antibiotics, arabinose, or bipA. At this point, functional proteins using bipA for proper expression, folding and function are still present in the cell, but protein turnover eventually replaces the bipA-dependent proteins with bipA-independent variants in which the UAG codon is replaced by one of the 64 codons. This in turn provides a strong selection for canonical amino acids that can replace bipA and maintain protein function. Samples of the population were taken at five time points after electrotransformation to track the population dynamics—after 1 h, 100 µl of culture was centrifuged at 16,000*g*, resuspended in 20 µl distilled water (dH_2O), and frozen

at –20 °C (time point 1); 2 ml of LB^L was added to the culture and then growth was allowed to proceed for 3 more hours before 100 µl of culture was centrifuged at 16,000*g*, resuspended in 20 µl dH_2O, and frozen at –20 °C (time point 2); the remaining culture was grown overnight to confluence after which 500 µl of culture was centrifuged at 16,000*g*, resuspended in 500 µl dH_2O, and frozen at –20 °C (time point 3); 30 µl of confluent culture was diluted into 3 ml of fresh LB^L and re-grown to confluence after which 500 µl of culture was centrifuged at 16,000*g*, resuspended in 500 µl dH_2O, and frozen at –20 °C (time point 4); finally, 30 µl of confluent culture was diluted into 3 ml of fresh LB^L and re-grown to confluence after which 500 µl of culture was centrifuged at 16,000*g*, resuspended in 500 µl dH_2O, and frozen at –20 °C (time point 5). After sampling was complete, we had obtained five time points from eight strains amounting to 40 total samples. Population dynamics were analysed by next-generation sequencing.

Next-generation sequencing of populations with degeneracy introduced at UAG positions. We designed custom primers to amplify ~127–146 base pairs (bp) surrounding the UAG codon of each variant and to add Illumina adapters and barcodes for sequencing. In order to reduce primer dimers, we redesigned the P5 primer binding sequence (Sol-P5_alt-PCR, Supplementary Table 1). We used PCR to introduce Illumina sequencing primer binding sites separated from the target amplicon by a 4–6 bp 'heterogeneity spacer' that allows low diversity Illumina libraries to be sequenced out of phase[42] (Supplementary Table 1). We estimate that ~10^6 cells (1 µl of a confluent culture containing ~10^9 cells per ml) were assayed at each time point. This PCR was performed in 20 µl reactions containing 10 µl of KAPA HiFi HotStart ReadyMix, 9 µl of dH_2O, 0.5 µl of each 20 µM primer, and 1 µl of template cells. Thermocycling (BioRad C1000 thermocycler) involved heat activation at 95 °C for 3 min, followed by 30 cycles of denaturation at 98 °C for 20 s, annealing at 62 °C for 15 s, and elongation at 72 °C for 30 s with a final elongation for 1 min (PCR1). PCR1 products (20 µl) were purified with magNA beads (40 µl)[43] and eluted in 20 µl of dH_2O. A second PCR (PCR2) amplification introduced Illumina adapters tagged with a unique 6 bp barcode (on the P7 adaptor) for each sample and time point. The PCR2 thermocycling and purification protocols were identical to those of PCR1 except that the products from PCR1 were used as template and different primers were used. The final DNA libraries were checked on a 1.5% w/v agarose gel and quantitated using a NanoDrop ND-1000 spectrophotometer. Equimolar samples of all 40 libraries were combined in a single tube and sequenced using a SE50 kit on an Illumina MiSeq (Dana Farber Cancer Institute Molecular Biology Core Facility). The P7 and Index1 reads were performed with standard sequencing primers, whereas the P5 read was sequenced with a custom primer (Sol-P5_alt-PCR, Supplementary Table 1).

Sequencing analysis of populations with NNN degeneracy at UAG positions. A simple Python script was written to tally each of the 64 UAG→NNN codon mutations and 21 amino acid/translational stop substitutions. We discarded all reads that were too short to discern the NNN codon. For all other reads, a constant seed sequence was indexed within the read, and the NNN codon was located based on proximity to this known seed sequence. The NNN codon identity and translated amino acid identities were stored in dictionaries entitled 'aas' and 'codons', respectively. The dictionaries and code are available together at GitHub (https://github.com/churchlab/NNN_sequencing_scripts).

Shannon entropy calculations. Shannon entropy was calculated using the standard relation $H(X) = -\sum_i P(x_i) \log P(x_i)$.

Whole-genome sequencing analysis of mutagenic escapees. We performed whole-genome sequencing on 20 escapees and their bipA-dependent parental strains. Sequencing libraries were prepared according to ref. 43 and sequenced with 150-bp paired-end reads on an Illumina MiSeq. We used Millstone (http://churchlab.github.io/millstone/) to automatically call single-nucleotide variants from raw fastq data with respect to our starting GRO C321.ΔA (NCBI GenBank Accession CP006698.1). Thus all variant positions are reported relative to the frame of this genome. All variant calls are available on Github (https://github.com/churchlab/dependence/tree/master/supplementary_materials). We then filtered these with custom scripts (https://github.com/churchlab/dependence) to identify alleles involved in hypothetical escape mechanisms: mutations in tRNAs that could lead to UAG suppression, mutations in translation machinery that could increase misincorporation of canonical amino acids, mutations in functionally related genes that could functionally complement the essential gene, and mutations in chaperones or proteases that could stabilize poorly folded Adk and TyrS proteins. Additionally, for strains with adequate coverage we performed *de novo* assembly of unmapped reads to uncover structural variants not reported by Millstone. We used Velvet[44] with a hash length of 21 and the following parameters for the graphing step: *-cov_cutoff 20 -ins_length 200 -ins_length_sd 90*. We then systematically queried NCBI BLAST to identify each *de novo* sequence and biased the BLAST results to prefer hits against the canonical MG1655 genome so that we could later group contigs by position (https://github.com/churchlab/dependence/blob/master/supplementary_materials/velvet_contigs_and_BLAST_data_non_permissive_8_strains.csv). Putative Lon insertions were visually confirmed using Millstone's JBrowse portal.

Integration of *bipARS* and $tRNA_{bipA}$. Genomic integration of *bipARS* and $tRNA_{bipA}$ was achieved in two steps by first replacing the endogenous *tdk* gene with the P_{araBAD}-inducible *bipARS* gene from pEVOL. Subsequently, the pEVOL tRNA and chloramphenicol resistance gene were inserted immediately downstream of P_{araBAD}-*bipARS*. Kapa HiFi Ready Mix was used to amplify each PCR product (see Supplementary Table 1 for primer sequences), and λ Red-mediated recombination was used to introduce the PCR products into the genome. Proper insertion of the desired cassettes was confirmed by PCR using tdk.seq-f and tdk.seq-r. We observed that 10 µM bipA was not adequate to support growth of adk.d6 or tyrS.d8 when *bipARS* was integrated into the genome; however, 100 µM bipA accommodated robust growth of adk.d6_int, tyrS.d8_int and adk.d6_tyrS.d8_int.

Design of the bipARS.d7 bipA-dependent synthetase. We applied our computational second-site suppressor strategy to position V290 of the bipyridylalanyl-tRNA synthetase X-ray structure (PDB code 2PXH, chain A). This position corresponds to bipA303 in our X-ray structure of tyrS.d7 when the structures are superimposed (alignable core backbone root mean square deviation of 3.6 Å). We hypothesized that this position may be amenable to redesign in homologous structures. Six designs covering sequence variability observed in the computational models (Supplementary Table 2) were produced by PCR amplification of *bipARS* with mutagenic primers (Supplementary Table 1) and isothermal assembly into the pEVOL vector, maintaining only the arabinose-inducible copy of *bipARS*. We also included the D286R mutation previously shown to increase synthetase activity[45] in all constructs. Since the *bipARS* designs should require bipA to translate, fold and function, all derived strains were initially co-transformed with a nonreplicating plasmid (pJTE2, R6γ origin of replication) containing a wild-type copy of *bipARS* to jumpstart production of $tRNA_{bipA}$. Designs were co-transformed with the jumpstart plasmid into C321.ΔA, and transformants were then transformed with a previously described GFP reporter plasmid containing a single UAG codon[9] to measure synthetase activity by GFP fluorescence. One design (bipARS.d7; T/A/G/G/A/bipA) produced >5-fold bipA-dependent induction of fluorescence in permissive media but failed to induce any bipA-dependent fluorescence after passaging overnight 1:150 in non-permissive media followed by an identical passage in permissive media. Since any functional synthetase remaining after non-permissive passaging should facilitate exponential production of additional synthetase, this behaviour suggests strong dependence of bipARS.d7 on bipA for translation and folding resulting in total clearance of bipARS.d7 and $tRNA_{bipA}$ after overnight growth in non-permissive conditions. The bipARS.d7/$tRNA_{bipA}$ and jumpstart vectors were co-transformed into C321.ΔA and then adk.d6 and tyrS.d8 were introduced as described above.

Growth competition assays. The assayed single- and double-enzyme synthetic auxotroph escapee strains (pgk.d4 esc. 1, 2 and 3; adk.d6_tyrS.d8 esc. 1, 2 and 3) were transformed with a pZE21 vector[46] bearing mCFP under aTc-inducible control in the multiple cloning site. The parental prototrophic C321.ΔA strain was similarly transformed with an identical vector except that the fluorophore is YFP. Strains were grown to late-exponential phase in LB^L supplemented with antibiotics (SDS, chloramphenicol, kanamycin), inducers (0.1% L-arabinose, 100 ng ml^{-1} aTc) and bipA. Cells were washed twice in M9 salts and adjusted to a cell concentration of roughly 1×10^9 cells per ml. Biological replicates of synthetic auxotroph escapees were mixed with the C321.ΔA strain at a ratio of 100:1 and diluted to a seeding concentration of roughly 2.5×10^7 cells per ml in non-permissive media (LB^L supplemented with SDS, chloramphenicol, kanamycin and aTc). Growth kinetics of the competition mixture were assayed in 200 µl sample volumes on microtitre plates incubated in a Biotek Synergy microplate reader at 34 °C. Cell mixtures were fixed in PBS with 1% paraformaldehyde at time 0 and at 8 h. Fixed cells were run on a BD LSRFortessa cell analyser and populations were binned based on YFP expression level. CFP was not used for species discrimination but rather to maintain consistent fitness costs associated with episomal DNA maintenance and fluorophore expression.

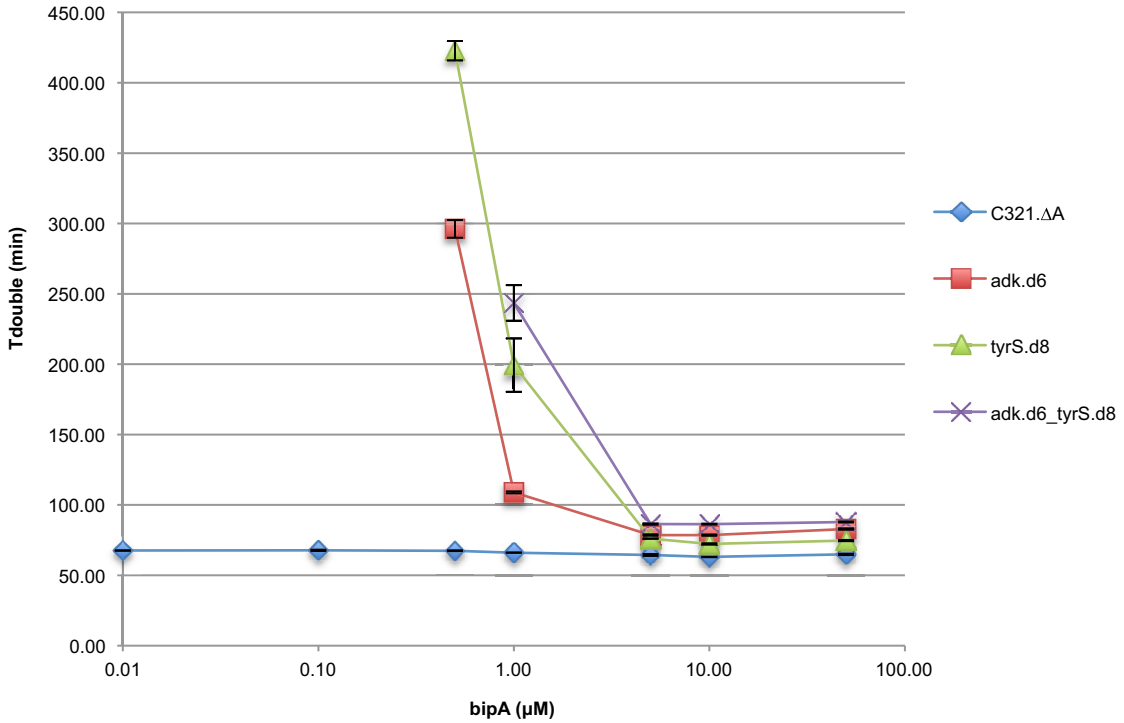
Bacterial lysate growth assays. All strains were grown up in permissive conditions and harvested in late-exponential phase. Cells were washed twice in M9 salts (6 g l^{-1} Na_2HPO_4, 3 g l^{-1} KH_2PO_4, 1 g l^{-1} NH_4Cl, 0.5 g l^{-1} NaCl) by centrifugation at 17,900*g* and then diluted 100-fold into LB^L supplemented with 166.66 ml l^{-1} trypsin-digested *E. coli* extract (Teknova catalogue number 3T3900). Growth kinetics were assayed in 200 µl sample volumes on microtitre plates as described above. Three biological replicates were performed by splitting a single well-mixed initial seeding population.

Conjugal escape assays. The conjugal donor population was produced using the Epicentre EZ-Tn5 Custom Transposome kit to insert a mosaic-end-flanked kan^R-oriT cassette into random positions of the *E. coli* MG1655 genome. The population of integrants was plated on LB^L agar plates supplemented with kanamycin. Approximately 450 clones were lifted from the plate and pooled, which corresponds

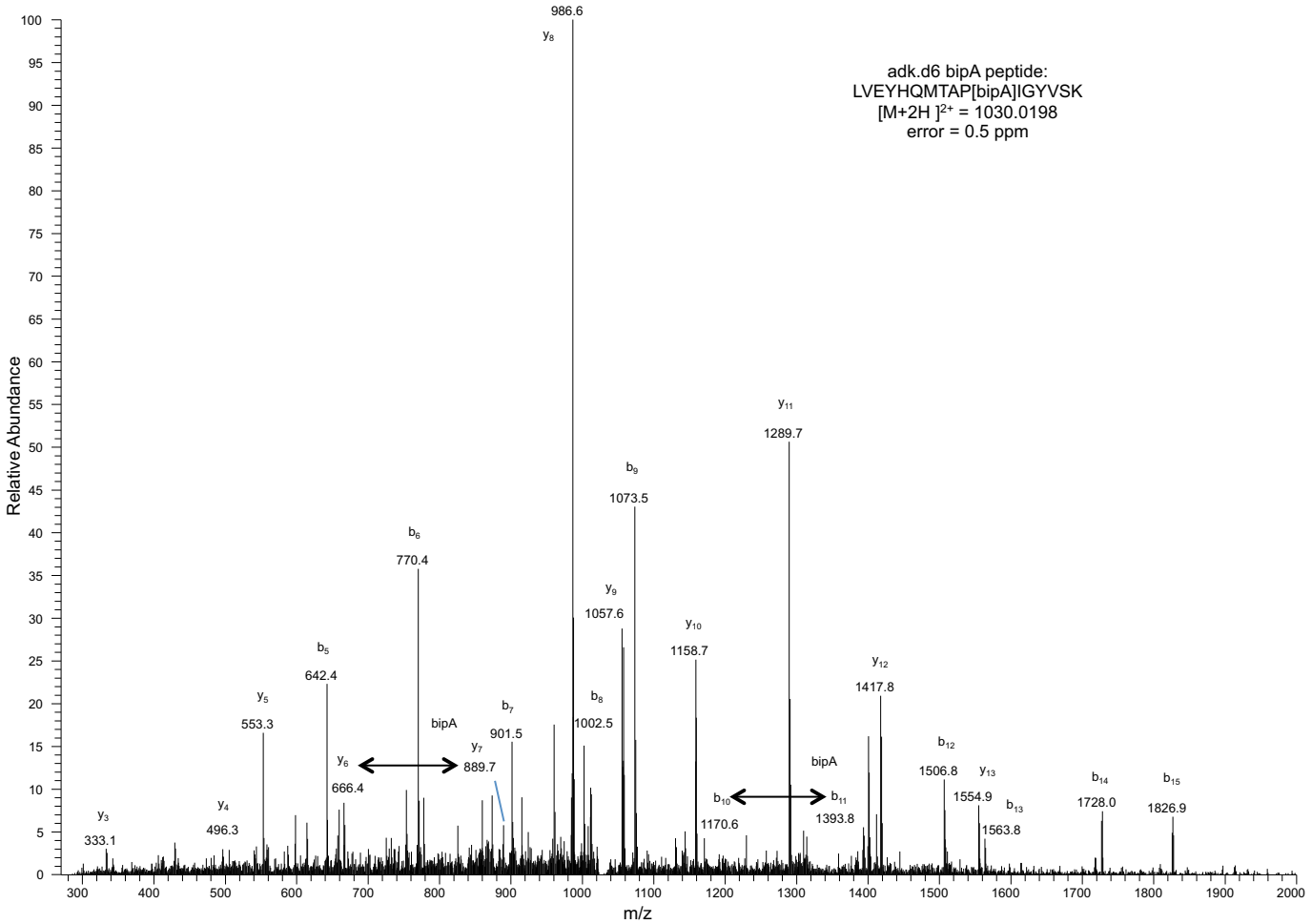
to one kan^R-oriT per ~10-kilobase pair region of the genome, assuming an equal distribution of transposition across the 4.6-megabase *E. coli* MG1655 genome. The pRK24 conjugal plasmid was conjugated[37] from *E. coli* strain 1100-2 (ref. 47) into the kan^R-oriT donor population. The kan^R-oriT insertion sites were confirmed to be well distributed. In brief, the donor population was sheared on a Covaris E210, end repaired, and ligated to Illumina adapters as described by ref. 43. Genomic sequences flanking the insertion site were amplified using the Sol-P5-PCR primer and a series of nested primers (Supplementary Table 1) that hybridize within the kan^R gene. PCR products corresponding to ~1 kilobase pair were gel purified from the smear and TOPO cloned (Invitrogen pCR-Blunt II-TOPO). Flanking genomic sequences were then identified by Sanger sequencing 96 TOPO clones. Conjugal escape assays were performed as described previously[37] with 50-min and 12-h conjugal duration and a donor:auxotroph ratio of 1:100. Three technical replicates of two biological replicates were performed for all conjugation assays with the exception of the double-enzyme synthetic auxotroph experiments, which were performed with three biological replicates (3 technical replicates each) to produce enough escapees for mascPCR screening. To determine the proportion of the genome overwritten by donor DNA the following numbers of colonies were scored for the 50-min/12-h time points: adk.d6 $n = 51/6$; tyrS.d8 $n = 44/7$; adk.d6_tyrS.d8 $n = 8/59$; adk.d6_tyrS.d8_asd:specR $n = 5/38$. This set omits a small collection of clones that could not be scored due to polyclonality.

Statistics. No statistical methods were used to predetermine sample size.

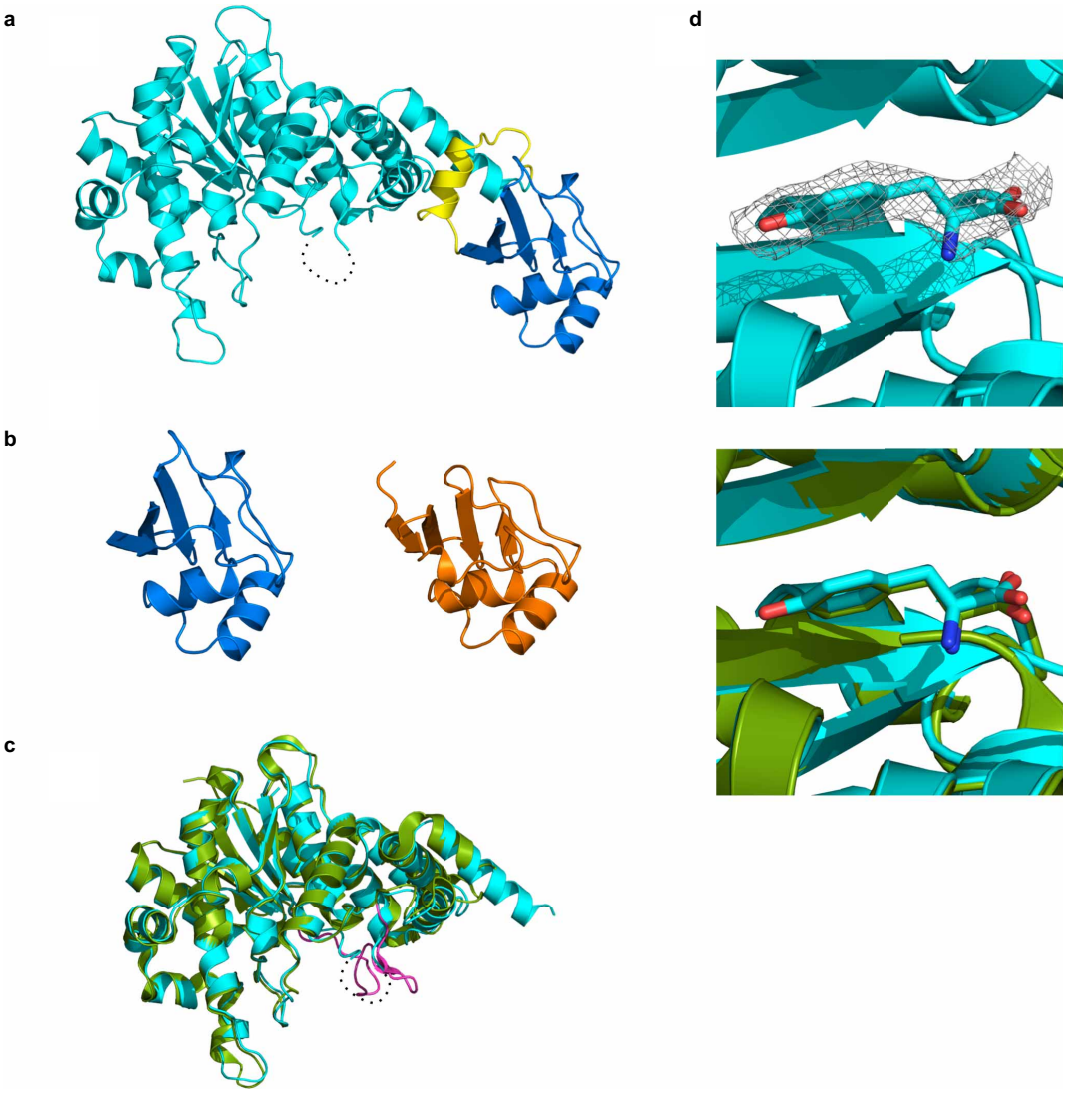
27. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res.* **33,** D34–D38 (2005).
28. UniProt Consortium. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* **41,** D43–D47 (2013).
29. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26,** 689–691 (2010).
30. Fraczkiewicz, R. & Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19,** 319–333 (1998).
31. Zhu, H., Fraczkiewicz, R. & Braun, W. *Solvent Accessible Surface Areas, Atomic Solvation Energies, and Their Gradients for Macromolecules* http://curie.utmb.edu/area_man.html (2012).
32. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl Acad. Sci. USA* **97,** 10383–10388 (2000).
33. Gregg, C. J. *et al.* Rational optimization of *tolC* as a powerful dual selectable marker for genome engineering. *Nucleic Acids Res.* **42,** 4779–4790 (2014).
34. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–345 (2009).
35. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97,** 6640–6645 (2000).
36. Yu, D. *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **97,** 5978–5983 (2000).
37. Isaacs, F. J. *et al.* Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333,** 348–353 (2011).
38. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 276 (ed Carter, C. W. Jr) 307–326 (Academic, 1997).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
40. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
41. Eggertsson, G. & Soll, D. Transfer ribonucleic acid-mediated suppression of termination codons in Escherichia coli. *Microbiol. Rev.* **52,** 354–374 (1988).
42. Fadrosh, D. W. *et al.* An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2,** 6 (2014).
43. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22,** 939–946 (2012).
44. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–829 (2008).
45. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* **395,** 361–374 (2010).
46. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25,** 1203–1210 (1997).
47. Tolonen, A. C., Chilaka, A. C. & Church, G. M. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol. Microbiol.* **74,** 1300–1313 (2009).



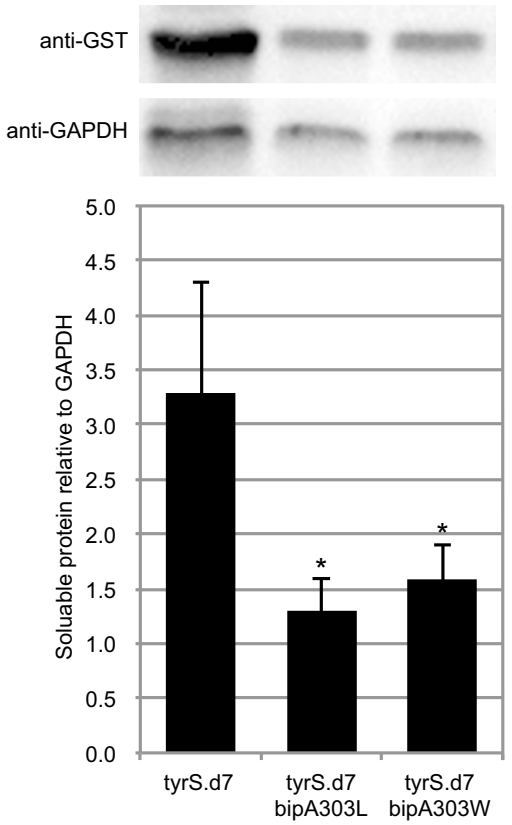
Extended Data Figure 1 | bipA dependence in synthetic auxotrophs. Prototrophic and synthetic auxotrophic strains were grown in titrations of bipA and monitored in a microplate reader (Methods). Media for all bipA concentrations contained SDS, chloramphenicol and arabinose. Doubling times for three technical replicates are shown. Positive and negative error bars are s.e.m. Growth was undetectable for synthetic auxotrophs at 0.00 μM, 0.01 μM and 0.10 μM bipA, as well as 0.50 μM bipA for adk.d6_tyrS.d8.



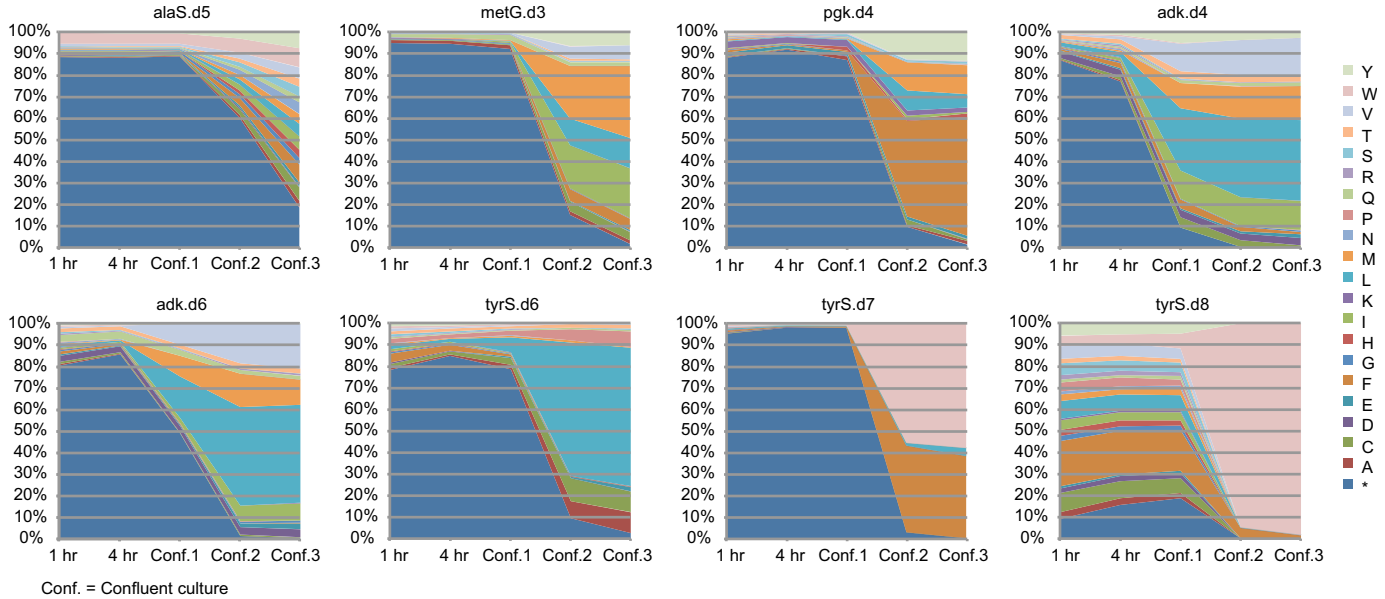
Extended Data Figure 2 | Mass spectrometry of NSAA-dependent enzymes. Mass spectrometry was performed and peptide spectrum matches (PSMs) were obtained as described in the Methods. Data sets were culled of minor contaminant PSMs and re-searched with SEQUEST against adk.d6, tyrS.d7 and tyrS.d8 sequences without taking into account enzyme specificity. To interrogate the sequences for bipA, tryptophan and leucine, the amino acid at the bipA position was given the mass of leucine and searches were performed with differential modifications of +110.01565 and +72.99525 to account for the masses of bipA and tryptophan, respectively. In all samples, only bipA, and not leucine or tryptophan, was detected at these positions. The PSM for adk.d6 is shown. Peptides observed to contain bipA are LVEYHQMTAP[bipA]IGYVSK (adk.d6), AQYV[bipA]AEQVTR (tyrS.d7) and AQYV[bipA]AEQATR (tyrS.d8).



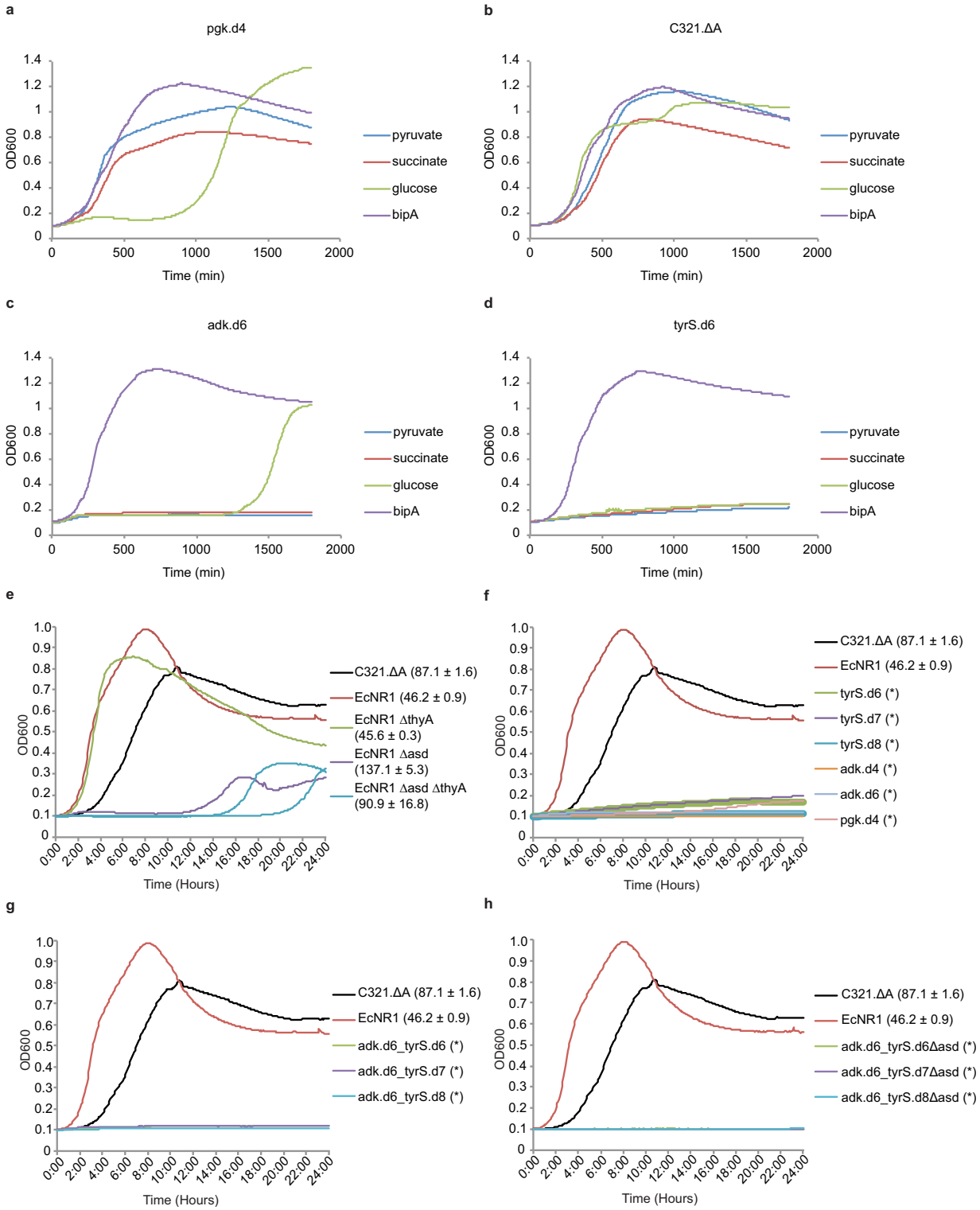
Extended Data Figure 3 | Crystal structure of tyrS.d7. **a**, Overall structure of the redesigned enzyme. The N-terminal domain (residues 4–330) that catalyses tyrosine activation, the carboxy-terminal tRNA-binding domain (residues 350–424) and their connecting region are coloured cyan, blue and yellow, respectively. The residues 232–241 are disordered (dash line). **b**, Comparison between the C-terminal tRNA recognition domains of tyrS.d7 (blue) and of *Thermus thermophilus* TyrS (orange; PDB code 1H3E). The residues 352–442 of the hyperthermophilic TyrS are shown. **c**, The N-terminal domain of the engineered protein is superposed on the crystal structure of its parental enzyme (green; PDB code 1X8X). The KMSKS loop of the parental enzyme is highlighted in magenta. **d**, Tyrosine molecule bound to tyrS.d7. An electron density map of L-tyrosine is shown as a grey mesh ($2F_o - F_c$ contoured at 1.2σ; top panel). A tyrosine and the surrounding protein fold of tyrS.d7 (cyan) are very similar to those of the wild-type TyrS structure (green; bottom panel).



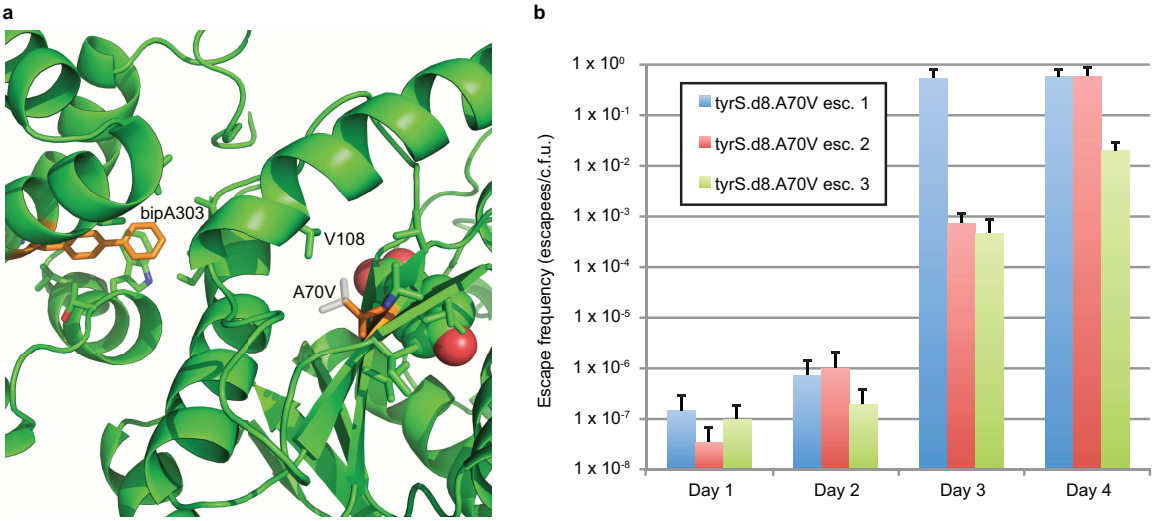
Extended Data Figure 4 | Western blot analysis of tyrS.d7 variants. Variants of tyrS.d7 with leucine or tryptophan at the bipA position were expressed as GST fusions under identical conditions and analysed by western blot (Methods). Soluble protein content was quantified by densitometry and normalized to GAPDH. Mutating bipA to leucine or tryptophan reduced soluble TyrS levels by 2.5- or 2.1-fold, respectively (*$P < 0.05$ by two-tailed unpaired Student's *t*-test with unequal variances). Three technical replicates were performed; a representative image is shown. Positive error bars are s.e.m.



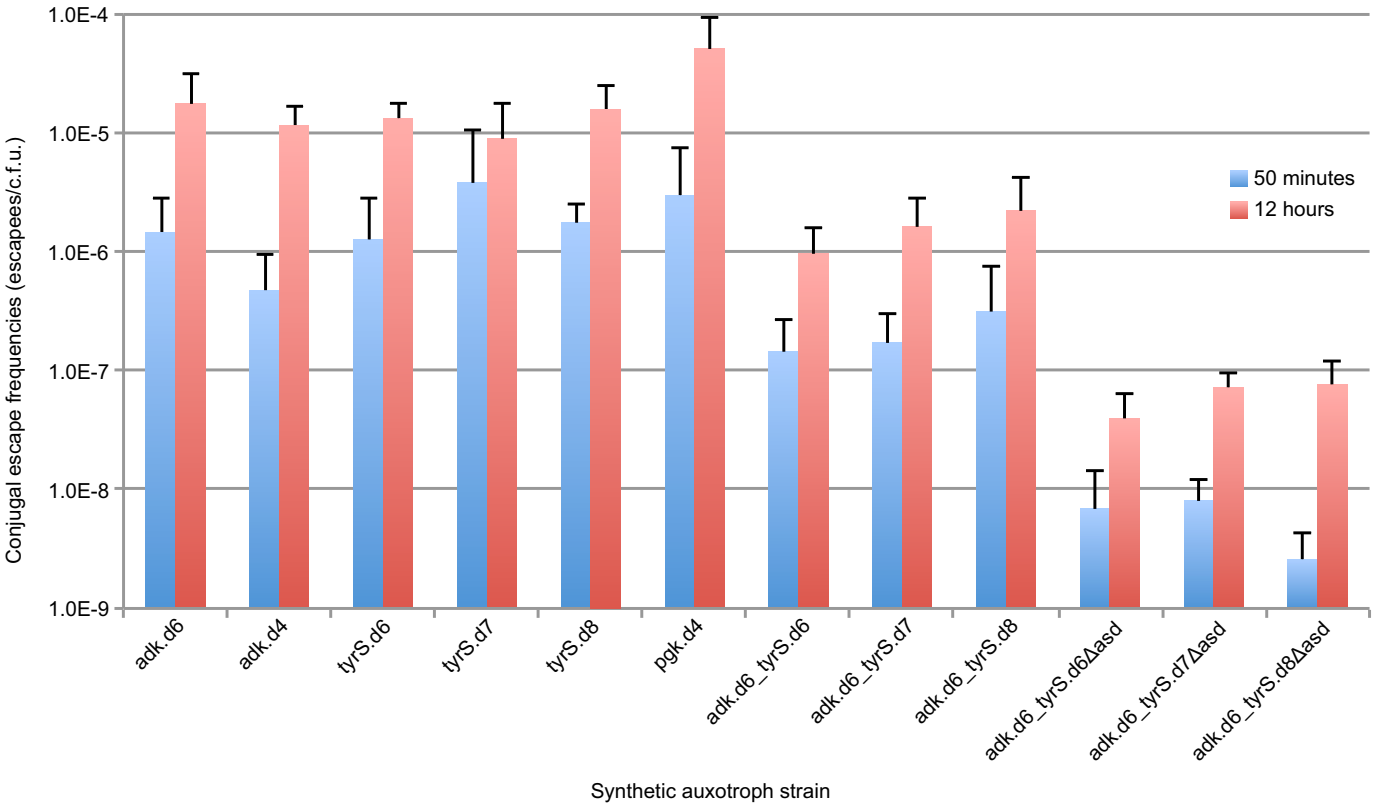
Extended Data Figure 5 | Population selection dynamics for canonical amino acid substitutions at designed UAG positions. For each plot, degenerate MAGE oligonucleotides were used to create a population of cells in which the UAG codon was mutated to all 64 codons. Codon substitutions leading to survival in the absence of bipA were selected by growth in LB^L media without bipA and arabinose supplementation. Aliquots of the culture population were taken at 1 h, 4 h, confluence 1 (once the culture reached confluence), confluence 2 (after regrowth of a 100-fold dilution of confluence 1) and confluence 3 (after regrowth of a 100-fold dilution of confluence 2). The amino acid identity at the bipA position was probed by targeted Illumina sequencing. Residual bipA-containing proteins were expected to remain active until intracellular protein turnover cleared them from the cell, making the 1-h time point a reasonable representation of initial diversity present in the population. These data show the relative fitness of amino acid substitutions in a given protein variant; relative fitness across multiple protein variants cannot be accurately assessed from these data.



Extended Data Figure 6 | Natural metabolites can circumvent auxotrophies. a–d, Synthetic auxotrophs of *pgk* can be complemented by pyruvate or succinate. Strains were cultured in LB^L in the presence of pyruvate, succinate, glucose or bipA (10 μM) and monitored by kinetic growth. The single-enzyme synthetic auxotroph pgk.d4 (**a**) grows similarly to prototrophic C321.ΔA (**b**) in the presence of pyruvate and succinate, but not glucose. Synthetic auxotrophs of *adk* (**c**) and *tyrS* (**d**) grow robustly in bipA but cannot be complemented by pyruvate or succinate. Growth of pgk.d4 and adk.d6 in glucose after 1,000 min is due to mutational escape (loss of bipA dependence). **e**, The synthetic auxotroph parental strain (C321.ΔA), a second prototrophic MG1655-derived strain (EcNR1), and three natural auxotroph derivatives of EcNR1 were grown in LB^L supplemented with 166.66 ml^{-1} bacterial lysate (Teknova). Growth curves are shown with doubling times ± one standard deviation of three technical replicates next to the labels. The conditions fully complement the metabolic auxotrophy of EcNR1.*ΔthyA*, which doubles as robustly as prototrophic EcNR1. Strains lacking the *asd* gene (EcNR1.*Δasd* and the EcNR1.*ΔasdΔthyA* double knockout) show more impairment but enter exponential growth with doubling times of 91 to 137 min, respectively. **f**, **g**, Single- (**f**) and double-enzyme (**g**) synthetic auxotrophies are not complemented by natural products in rich media or bacterial lysate. **h**, When the *Δasd* auxotrophy is combined with double-enzyme synthetic auxotrophies the natural products are no longer sufficient to support growth. No growth is indicated by an asterisk in **f**–**h**.



Extended Data Figure 7 | Analysis of the A70V mutation as an escape mechanism for tyrS.d8. **a**, The X-ray structure of tyrS.d7 is shown; tyrS.d8 varies by the single mutation V307A. BipA303, A70 and their neighbouring side chains are shown in stick representation, with bipA303 and A70 coloured orange. The bound tyrosine substrate is shown in spacefill. The A70V mutation (white sticks) may stabilize the catalytic domain when bipA is replaced by natural amino acids by tightly packing with neighbouring side chains including V108. **b**, Escape frequencies on non-permissive media for three separately constructed tyrS.d8 A70V strains are shown for days 1 through 4. Although escapees are growth-impaired in the absence of bipA (Supplementary Table 10), all cells form colonies after 5 days, suggesting that A70V confers 100% survival on non-permissive media. Positive error bars indicate s.e.m.



Extended Data Figure 8 | Conjugal escape frequencies of synthetic auxotrophs. Single-, double- and triple-enzyme auxotrophs were assayed to determine the frequency of escape by HGT and recombination from a prototrophic donor as described in the Methods. These results highlight the benefit of having multiple auxotrophies distributed throughout the genome. Notably, scaling from a single synthetic auxotrophy to three distributed auxotrophies results in a reduction of conjugal escape by at least two orders of magnitude. Positive error bars indicate standard deviation.

Extended Data Table 1 | Data collection and refinement statistics

	tyrS.d7
Data collection	
Space group	*P* 12_11
Cell dimensions	
a, *b*, *c* (Å)	81.3, 67.2, 90.7
a, *b*, *g* (°)	90.0, 102.6, 90.0
Resolution (Å)	50.0 - 2.65 (2.74-2.65) *
R_{sym} or R_{merge}	0.074 (0.497)
I/s*I*	29.2 (4.65)
Completeness (%)	99.0 (98.4)
Redundancy	7.6 (7.7)
Refinement	
Resolution (Å)	45 – 2.65
No. reflections	26407
R_{work}/ R_{free}	0.222/0.306
No. atoms	
Protein	6038
Ligand/ion	13
Water	57
B-factors	
Protein	58.66
Ligand/ion	52.10
Water	48.24
R.m.s deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.530

The data were collected using a single crystal.
* Highest-resolution shell is shown in parentheses.

Extended Data Table 2 | Cost per litre of culture for commonly used NSAAs

NSAA	Vendor	Name at vendor	CAS#	MW	Cat# for 1g	Price of 1g	Optimal conc. (mM)	Cost per liter of culture
pAcF	peptech	L-4-Acetylphenylalanine	122555-04-8	207.23	AL624-1	$500.00	1.0	$103.62
pAzF	Bachem	H-4-Azido-Phe-OH	33173-53-4	206.2	F-3075.0001	$285.00	5.0	$293.84
pCNF	peptech	L-4-Cyanophenylalanine	167479-78-9	190.2	AL240-1	$150.00	1.0	$28.53
bpa	peptech	L-4-Benzoylphenylalanine	104504-45-2	269.3	AL660-1	$100.00	1.0	$26.93
napA	peptech	L-2-Naphthylalanine	58438-03-2	215.25	AL121-1	$80.00	1.0	$17.22
bipA	peptech	L-4,4'-Biphenylalanine	155760-02-4	241.29	AL506-1	$150.00	0.1	$3.62
pIF	peptech	L-4-Iodophenylalanine	24250-85-9	291.09	AL261-1	$40.00	1.0	$11.64
bipyA	Asis Chem	Bipyridylalanine	custom synthesis	245.282	(25 g price)	$10,000/25g	1.0	$98.11