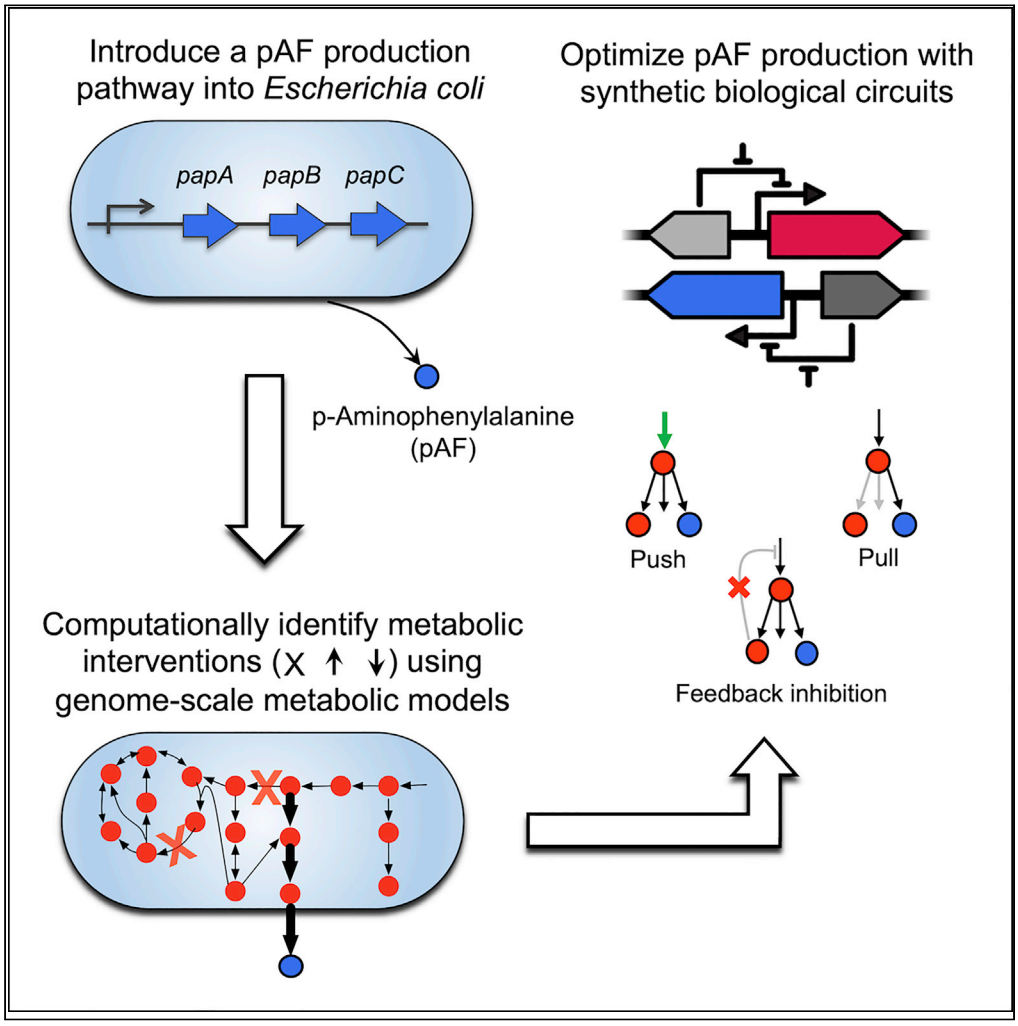
iScience

CellPress
OPEN ACCESS

Article

Computational design and engineering of an *Escherichia coli* strain producing the nonstandard amino acid *para*-aminophenylalanine



Ali R. Zomorrodi, Colin Hemez, Pol Arranz-Gibert, Terrence Wu, Farren J. Isaacs, Daniel Segrè

farren.isaacs@yale.edu (F.J.I.)
dsegre@bu.edu (D.S.)

Highlights

Sought to optimize *para*-aminophenylalanine (pAF) production and growth in *E. coli*

Identified interventions in the host native metabolism using genome-scale models

Constructed multiple mutant strains involving gene knockouts and/or overexpressions

Flux modification in chorismate biosynthesis pathway significantly raised pAF titer

Zomorrodi et al., iScience *25*, 104562
July 15, 2022
https://doi.org/10.1016/j.isci.2022.104562

iScience

CellPress
OPEN ACCESS

Article

Computational design and engineering of an *Escherichia coli* strain producing the nonstandard amino acid *para*-aminophenylalanine

Ali R. Zomorrodi,[1,2,3,11] Colin Hemez,[4,5,6,11] Pol Arranz-Gibert,[4,5] Terrence Wu,[7] Farren J. Isaacs,[4,5,6,*] and Daniel Segrè[3,8,9,10,12,*]

SUMMARY

Introducing heterologous pathways into host cells constitutes a promising strategy for synthesizing nonstandard amino acids (nsAAs) to enable the production of proteins with expanded chemistries. However, this strategy has proven challenging, as the expression of heterologous pathways can disrupt cellular homeostasis of the host cell. Here, we sought to optimize the heterologous production of the nsAA *para*-aminophenylalanine (pAF) in *Escherichia coli*. First, we incorporated a heterologous pAF biosynthesis pathway into a genome-scale model of *E. coli* metabolism and computationally identified metabolic interventions in the host's native metabolism to improve pAF production. Next, we explored different approaches of imposing these flux interventions experimentally and found that the upregulation of flux in the chorismate biosynthesis pathway through the elimination of feedback inhibition mechanisms could significantly raise pAF titers (~20-fold) while maintaining a reasonable pAF production-growth rate trade-off. Overall, this study provides a promising strategy for the biosynthesis of nsAAs in engineered cells.

INTRODUCTION

Engineering microbes with diverse natural and nonnaturally occurring functions is a major endeavor in synthetic biology, important for multiple goals. One goal is the microbial production of industrial and therapeutic small molecules through metabolic engineering. Another goal involves the ribosomal production of proteins containing nonstandard amino acids (nsAAs), which expand the chemistry of the canonical set of 20 amino acids that organisms in all kingdoms of life use for protein biosynthesis (Lajoie et al., 2013; Soye et al., 2015). Proteins that incorporate nsAAs enable diverse biochemistries not typically found in nature, such as hydrocarbon-based secondary structure stabilization (Walensky and Bird, 2014), site-specific antibody-drug conjugation (Zimmerman et al., 2014), and covalent linkage of proteins to form functional biopolymers (Albayrak and Swartz, 2014; Jin et al., 2019). Although short nsAA-bearing proteins, such as stapled peptides, have traditionally been generated via asymmetric synthesis methods (Williams, 1999) and cell-free translation techniques exist for the production of larger proteins (Hong et al., 2014), the ribosomal incorporation of nsAAs into the proteins of living cells greatly expands the scope and utility of nsAA-containing proteins. For instance, ribosomally synthesized nsAA-containing proteins can be incorporated into biocontainment strategies that limit the survival and propagation of engineered microbes to specified environments (Mandell et al., 2015; Rovner et al., 2015), fluorescent proteins that serve as *in vivo* probes for enzymatic activities (Xuan et al., 2017), and sequence-defined synthetic biomaterials containing multiple instances of nsAAs (Amiram et al., 2015; Arranz-Gibert et al., 2018; Shapiro et al., 2022; Vanderschuren et al., 2022).

In order to synthesize proteins that contain ribosomally synthesized nsAAs, engineered organisms require a pool of nsAAs from which to draw during translation. To date, the problem of provisioning organisms with nsAAs has been predominantly tackled by exogenously supplementing the desired amino acids (Amiram et al., 2015). In addition, there are only a few reports of *in vivo* nsAA production for genetic code expansion by manipulating endogenous amino acid biosynthetic pathways to favor the intracellular accumulation of intermediates that are used as nsAAs (Gaston et al., 2011; Mehl et al., 2003). Both strategies are subject to limitations. In the case of exogenous supplementation, some nsAAs may have low cell membrane

[1]Mucosal Immunology and Biology Research Center, Pediatrics Department, Massachusetts General Hospital, Boston, MA, USA

[2]Harvard Medical School, Boston, MA, USA

[3]Bioinformatics Graduate Program, Boston University, Boston, MA, USA

[4]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT, USA

[5]Systems Biology Institute, Yale University, West Haven, CT, USA

[6]Department of Biomedical Engineering, Yale University, New Haven, CT, USA

[7]Yale West Campus Analytical Core, 600 West Campus Drive, West Haven, USA

[8]Department of Biology, Boston University, Boston, MA, USA

[9]Department of Biomedical Engineering, Boston University, Boston, MA, USA

[10]Biological Design Center, Boston University, Boston, MA, USA

[11]These authors contributed equally

[12]Lead contact

*Correspondence: farren.isaacs@yale.edu (F.J.I.), dsegre@bu.edu (D.S.)

https://doi.org/10.1016/j.isci.2022.104562

permeability, compromising their transport into the cell for ribosomal incorporation. In the case of native pathway production, the set of nsAAs that can be synthesized is limited to the biosynthetic capabilities of the host organism.

Incorporating and optimizing heterologous biosynthetic pathways into a host cell's metabolism constitutes a third possible strategy for provisioning organisms with nsAAs; this circumvents the need for exogenous nsAA supplementation and enables the biosynthesis of a vast range of nsAAs beyond the host organism's native metabolic capabilities. Heterologous pathways for synthesizing nsAAs can be obtained from bacteria, fungi, and plants that produce a wide variety of amino acids with nonstandard functional groups as intermediates when synthesizing antibiotics, toxins, and other bioactive small molecules (Sulzbach and Kunjapur, 2020). For example, the Gram-positive bacterium *Streptomyces cattleya* is capable of synthesizing a terminal-alkyne amino acid (Marchand et al., 2019) and 4-fluorothreonine (Murphy et al., 2003). Importing naturally occurring biosynthetic pathways for nsAAs into model organisms with a high capacity for ribosomal nsAA incorporation, such as genomically recoded *Escherichia coli* (Lajoie et al., 2013), could be an effective way to synthesize proteins using nsAAs. Compared with exogenous nsAA supplementation, endogenous production would increase the efficiency and decrease the cost of nsAA-containing protein biosynthesis. Another as-yet unrealized application that would be enabled by endogenous nsAA production is the construction of synthetic microbial consortia wherein each strain is biocontained on an nsAA produced by another strain in the community; this requires the optimization of nsAA production by microbial community members in addition to the optimization of ribosomal nsAA incorporation.

Despite this promise, the efficient biosynthesis of nsAAs and other small molecules via expression of heterologous pathways in laboratory host strains has proven challenging because it requires laborious efforts to mine, characterize, and optimize heterologous biosynthetic pathways in hosts, given that such pathways can disrupt cellular homeostasis by co-opting native metabolic resources for small molecules production (Breit et al., 2019; Pitera et al., 2007). Although a previous study aimed to address this challenge by engineering the native metabolism of an *E. coli* strain for producing an nsAA (Masuo et al., 2016), systems-level studies of nsAA overproduction, which take into account the entire scope of cellular metabolism, are still lacking.

GEnome-scale Models (GEMs) of metabolism can help address this gap (Long et al., 2015; Zomorrodi et al., 2012). These models consist of the full inventory of metabolic reactions encoded by the genome of an organism and can be computationally simulated to systematically explore metabolic tradeoffs in engineered organisms. A wide spectrum of computational approaches has been developed to design overproducing microbial strains by using GEMs as a basis (see Long et al., 2015 and Zomorrodi et al., 2012 for a review of these approaches). These tools computationally identify candidate flux changes in the network, such as knockouts, upregulations, or downregulations that lead to the enhanced production of a biochemical of interest. The identified flux changes are then mapped to genetic manipulations that can be implemented experimentally. Several studies have reported on the successful utilization of these computational pipelines to guide the design of engineered microbial strains overproducing commodity chemicals, biofuel precursors, and a variety of other chemicals (Fong et al., 2005; Ranganathan et al., 2012; Suástegui et al., 2017; Xu et al., 2011; Yim et al., 2011).

We hypothesized that the efficient synthesis of nsAAs using heterologous pathways also requires systematic engineering of the host's native metabolism to proportionately allocate metabolic resources to cellular growth and the bioengineering objective, i.e., nsAA production. In this study, we tackle this by combining computational systems biology approaches based on GEMs of metabolism and synthetic biology techniques to explore the feasibility of engineering highly efficient nsAA producer microbes. We focus our efforts on optimizing the biosynthesis of *para*-aminophenylalanine (pAF) in *E. coli*, generated from intracellular chorismate using a heterologously expressed gene cluster from *Pseudomonas fluorescens* (Masuo et al., 2016). Motivated by the observation that a trade-off exists between pAF production and growth rate in this engineered *E. coli* cultured in carbon-limited environments, we sought to explore opportunities for modulating this trade-off by using computational modeling. To this end, we used a GEM of *E. coli* metabolism and a computational strain design pipeline to better understand how the introduction of the heterologous pAF-producing pathway co-opts native metabolic resources and to computationally identify rational ways of rewiring the host metabolism to improve pAF production. We then used the predicted

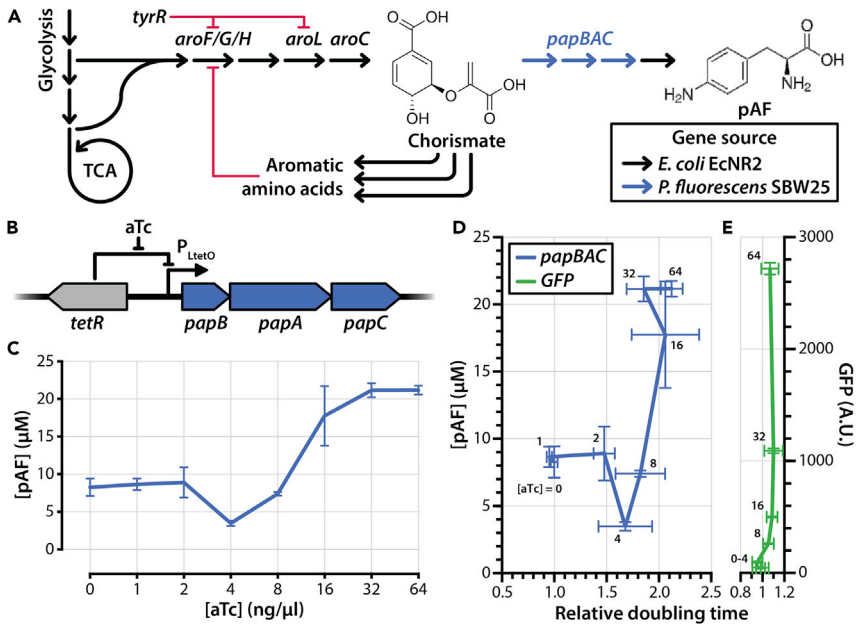


Figure 1. Trade-off between pAF production and growth in nonengineered strains

(A) Overview of pAF production via heterologous expression of *papBAC*. Chorismate, the end-product of the shikimate pathway, is converted to *p*-aminophenylpyruvate by the PapBAC enzymes, which is then converted to pAF by host cell deaminases. Red lines indicate inhibitory interactions on the specified genes.

(B) aTc-inducible papBAC overexpression circuit used in this study. *tetR*: tetracycline repressor.

(C) pAF production as a function of aTc concentration in *E. coli* strain EcNR2 bearing the aTc-inducible papBAC overexpression cassette and grown in M9 minimal medium.

(D) pAF production and doubling time (relative to growth without aTc) for EcNR2 bearing the aTc-inducible *papBAC* overexpression cassette.

(E) GFP fluorescence and doubling time (relative to growth without aTc) for EcNR2 bearing GFP in the place of *papBAC* under an aTc-inducible expression cassette. All data in this figure are represented as mean +/− SD of three replicates (technical replicates for pAF titer; biological replicates for all other measurements).

metabolic flux interventions as a starting point to apply multiplex genome engineering technologies (Gallagher et al., 2014; Wang et al., 2009; Wannier et al., 2021) to experimentally construct and test engineered strains for pAF production. We found that upregulation of metabolic flux in the chorismate biosynthesis pathway through the elimination of feedback inhibition mechanisms is the most promising strategy to increase pAF production. However, the optimized strains continued to exhibit a trade-off between growth rate and pAF production. Our study provides a basis for the systematic exploration of host cell metabolism to optimize the biosynthesis of natural products via heterologous expression. The strategy presented here may be applied to diverse biosynthetic pathways in a wide range of host organisms.

RESULTS

A trade-off between nsAA production and growth in carbon-limited environments

To enable pAF production by *E. coli*, we engineered the *E. coli* strain EcNR2 (Isaacs et al., 2011) to synthesize pAF using the *papBAC* gene cluster derived from *P. fluorescens* strain SBW25. The enzymes PapA, PapB, and PapC convert chorismate into *para*-aminophenylpyruvate, which is then converted to pAF by native cellular aminotransferases (Masuo et al., 2016) (Figure 1A). We constructed a pAF production circuit by placing the *papBAC* genes downstream of the P_{LtetO} anhydrotetracycline (aTc)-inducible promoter (Figure 1B) and cloned the gene cluster into a plasmid with a p15A origin of replication (Lutz and Bujard, 1997). The aTc-inducible promoter enables titratable expression of the genes under its control, a characteristic we confirmed by examining an analogous circuit with green fluorescent protein (GFP) in the place of *papBAC* (Isaacs et al., 2004) (Figure 1E).

We induced *papBAC* expression at a range of aTc concentrations in M9 minimal medium supplemented with 0.4% glucose and measured extracellular pAF concentrations after 24 h of incubation using tandem

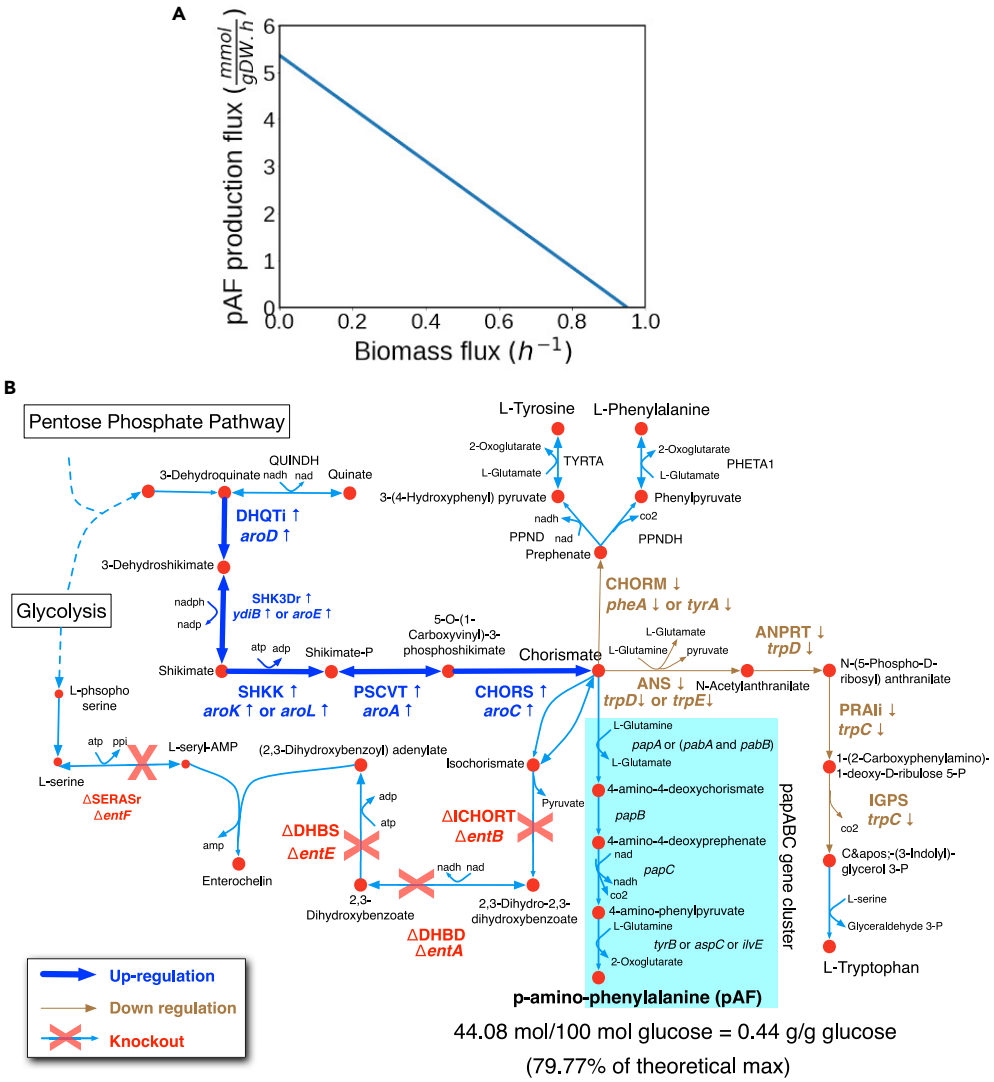

A
pAF production flux ($\frac{mmol}{gDW.h}$)
Biomass flux (h^{-1})
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
B
Pentose Phosphate Pathway
Glycolysis
L-Tyrosine
L-Phenylalanine
2-Oxoglutarate
L-Glutamate
TYRTA
PHETA1
3-(4-Hydroxyphenyl) pyruvate
Phenylpyruvate
nadh
co2
PPND
nad
PPNDH
Prephenate
QUINDH
3-Dehydroquinate
nadh
nad
Quinate
DHQTi ↑
aroD ↑
3-Dehydroshikimate
nadph
nadp
SHK3Dr ↑
ydiB ↑ or aroE ↑
Shikimate
atp
adp
Shikimate-P
5-O-(1-Carboxyvinyl)-3-phosphoshikimate
Chorismate
SHKK ↑
aroK ↑ or aroL ↑
PSCVT ↑
aroA ↑
CHORS ↑
aroC ↑
CHORM ↓
pheA ↓ or tyrA ↓
L-Glutamate
pyruvate
L-Glutamine
ANPRT ↓
trpD ↓
N-(5-Phospho-D-ribosyl) anthranilate
ANS ↓
trpD↓ or trpE↓
N-Acetylanthranilate
PRAIi ↓
trpC ↓
1-(2-Carboxyphenylamino)-1-deoxy-D-ribulose 5-P
IGPS
trpC ↓
co2
C'-(3-Indolyl)-glycerol 3-P
L-Tryptophan
L-serine
Glyceraldehyde 3-P
L-phospho serine
atp
ppi
L-seryl-AMP
ΔSERASr
ΔentF
(2,3-Dihydroxybenzoyl) adenylate
Isochorismate
Pyruvate
ΔICHORT
ΔentB
ΔDHBS
ΔentE
amp
Enterochelin
adp
atp
nadh
nad
2,3-Dihydroxybenzoate
2,3-Dihydro-2,3-dihydroxybenzoate
ΔDHBD
ΔentA
L-Glutamine
papA or (pabA and pabB)
L-Glutamate
4-amino-4-deoxychorismate
papB
4-amino-4-deoxyprephenate
nad
papC
nadh
co2
4-amino-phenylpyruvate
L-Glutamine
tyrB or aspC or ilvE
2-Oxoglutarate
p-amino-phenylalanine (pAF)
papABC gene cluster
Up-regulation
Down regulation
Knockout
44.08 mol/100 mol glucose = 0.44 g/g glucose
(79.77% of theoretical max)
C
Reaction-level manipulations
Gene-level manipulations
Min pAF production level
34.97%
19.33
0.19
79.77%
44.08
0.44
% of theoretical max
mol/100 mol glucose
g/g glucose
DHQTi 8x OR SHK3Dr 8x
OR
SHKK 8x OR CHORS 8x
OR
PSCVT 8x
AND
ANS 60% OR ANPRT 60%
OR
PRAIi 60% OR IGPS 60%
AND
CHORM 43%
AND
ΔSERASr
OR
ΔDHBS
OR
ΔDHBD
OR
ΔICHORT
aroD OR ydiB
OR
aroE OR aroC
OR
aroA OR aroK
OR
aroL
AND
trpD OR trpE
OR
trpC
AND
pheA AND tyrA
AND
ΔentF
OR
ΔentE
OR
ΔentA
OR
ΔentB

Figure 2. Computational design of a pAF-producing *E. coli* strain using a heterologous pathway

(A) The predicted pAF production levels as a function of the biomass production flux (i.e., growth).

(B) Identified metabolic interventions to enhance pAF production.

(C) The impact of predicted metabolic flux interventions (left panel) and the corresponding gene-level interventions inferred from gene-protein-reaction associations in the model (right panel) on the minimum pAF production in the metabolic network. A minimum of three reaction (or four gene) interventions are needed to achieve a non-zero minimum pAF production. Implementing interventions shown on the right in the bar graphs (i.e., any of the four reaction or gene knockouts) in addition to those shown on the left further improves the pAF production. The minimum pAF production level was obtained by minimizing the pAF production in the model while imposing the identified interventions.

mass spectrometry. We observed low levels of pAF production even in the absence of aTc, a consequence of both leaky expression under P_{LtetO} and of peak overlap in MS spectra. Concentrations of aTc above 8 ng/mL yielded pAF titers above baseline, with pAF yield plateauing at ~20 μM (0.001 g pAF per g of glucose) beyond aTc concentrations of 32 ng/mL (Figure 1C). However, strains grown in aTc concentrations of 2 ng/mL and higher exhibited substantial increases in doubling time relative to the uninduced control (Figure 1D). Notably, the growth impairment of the strain increases concomitantly with higher *papBAC* expression, which suggests a trade-off between pAF production and growth rate. We did not observe a similar trade-off between protein expression and doubling time when we expressed nonenzymatic GFP under an analogous circuit (Figure 1E), indicating that the trade-off observed is not due to increased translational load. We hypothesized that the trade-off is a consequence of rerouting chorismate flux toward pAF production and away from native downstream pathways, which presumably attenuate the production of metabolites essential for growth in minimal medium (Figure 1A). Of note, the growth defect we observed at higher aTc concentrations is unlikely to be attributed to the toxicity or antimetabolite effect of pAF, as past studies that assessed the effects of pAF supplementation on the growth of *E. coli* in minimal media observed only minor growth impairments at very high (>10 mM) pAF concentrations (Mohammadi Nargesi et al., 2018). Nevertheless, we cannot completely rule out the possibility that one of the intermediate metabolites between chorismate and pAF (e.g., 4-amino-4-deoxychorismate and 4-amino-4-deoxyprephenate) is toxic or exhibits an antimetabolite effect, as intermediate metabolite toxicity from the heterologous expression of other biosynthetic pathways has been observed before (Carothers et al., 2009).

Construction of an in silico pAF-producing *E. coli*

Our observation that pAF production trades off with growth in *E. coli* illustrates the disruption of cellular homeostasis caused by the expression of heterologous enzymatic pathways, a phenomenon observed in other studies (Keasling et al., 2021; Pitera et al., 2007; Santos et al., 2011; Tsoi et al., 2018); this motivates the need to engineer the host cell's native metabolism in order to optimize nsAA production in dynamic environments. To address this, we leveraged computational systems biology methods to systematically study nsAA production in *E. coli*. To construct the corresponding pAF-producing *E.* coli in silico, metabolic reactions encoded by the *papBAC* gene cluster were extracted and consolidated from several online databases including MetaCyc (Caspi et al., 2014), KEGG (Kanehisa et al., 2014), and Model SEED (Henry et al., 2010) and were incorporated into the *i*JO1366 GEM of metabolism for *E. coli* (Orth et al., 2011) (Figure 2). We next used this engineered metabolic model to examine the production capacity of pAF by the engineered *E. coli* strain in silico. This analysis revealed that this strain is able to produce a maximum of 0.54 mmol of pAF per mmole of glucose (0.54 g pAF per g glucose) under the aerobic minimal M9 medium (equivalent of 2.16 g/L or 120 mM of pAF in an M9 medium with 4% glucose). However, this maximum is achieved at the expense of zero growth, implying that pAF production is in direct competition with growth, which was further confirmed by plotting the predicted maximum pAF production level by the model as a function of maximum growth (biomass production flux; Figure 2A); this is due to the fact that glutamine and chorismate, which are the main precursors for the pAF biosynthesis in the network, also serve as a precursor for biomass production (i.e., growth) and for the biosynthesis of a number of other essential amino acids such as phenylalanine, tyrosine, and tryptophan (Figure 2B). These results suggest that further engineering of native metabolism of the host *E. coli* strain is required to enhance pAF production.

Computational design of metabolic interventions for the native metabolism of the host *E. coli* strain

We used the computational strain design pipeline OptForce (Ranganathan et al., 2010) (STAR Methods) to identify metabolic interventions leading to the improved production of pAF in the *E. coli* strain harboring

the *papBAC* gene cluster under the aerobic minimal M9 medium using glucose as the carbon source. The first set of interventions that we identified consists of the upregulation of reactions encoded by the *papBAC* gene cluster that results in a minimum pAF production of ~80% of theoretical maximum when the growth is set to be at least 20% of maximum. These reaction flux increases draw the metabolic flow toward pAF and can be also readily inferred intuitively. However, pull strategies alone may not always be sufficient to provide a desired production level for a target biochemical due to the presence of bottlenecks in other parts of the metabolic network, which cannot be directly captured by GEMs of metabolism. Therefore, in the next step, we sought to identify metabolic interventions in other parts of the host's native metabolism that push metabolic flow toward pAF; this was done by re-running OptForce while preventing the upregulation of reactions encoded by the *papBAC* gene cluster. This analysis identified a set of reaction manipulations that collectively lead to an increased pool of chorismate (Figure 2B). Here, we found that at least three simultaneous reaction manipulations are required to enable a minimum pAF production based on the model. These reaction manipulations result in a minimum pAF production level of 34.97% of theoretical maximum (Figure 2C, left panel) and include (1) the increasing flux in any of the reactions in the shikimate pathway that are directly involved in the biosynthesis of chorismate (encoded by *aro* family genes or *ydiB*) (Figure 2B); (2) decreasing flux in any of the reactions that convert chorismite to aromatic amino acids namely L-tryptophan (encoded by *trpC*, *D* and *E*) and to phenylalanine and tyrosine (chorismate mutase "CHORM", encoded by *pheA* or *tyrA*) (Figure 2B). In addition, we found that a fourth flux intervention, i.e., the deletion of any of the reactions encoded by *entA*, *entB*, *entE*, and *entF* to prevent the conversion of chorismate to enterochelin, will further increase the minimum pAF production more than 2-fold (from 34.97% to 79.77% of theoretical maximum) (Figure 2C). Although the deletion of aromatic amino acid genes to improve pAF production has been explored before (Masuo et al., 2016), such modifications would lead to auxotrophies, and other interventions identified in this study have not been reported, thus highlighting the importance of taking into account the entire scope of metabolism (afforded by GEMs) to infer potentially promising metabolic interventions while minimizing impact on growth and fitness.

As shown in Figure 2C, a minimum of three reaction level and four gene level interventions (inferred from gene-protein-reaction associations in the GEM) are needed to obtain a non-zero minimum pAF production by the model. We further confirmed this by analyzing the effect of all single, pair, triple, and quadruple combinations of these genetic interventions on the pAF production. This analysis did confirm that no single, pair, or triple combinations of these genetic interventions is enough to enable a non-zero minimum pAF production in the network (Figure S1). This analysis, in principle, simulates epistatic interactions among our identified genetic manipulations with respect to pAF production as the phenotype of interest (an epistatic interaction between two genes means that the phenotype of a double-gene mutant strain cannot be easily inferred from phenotypes of single-gene mutant strains [Snitkin and Segrè, 2011]). Epistasis analysis shows that with simultaneous implementation of four genetic interventions, a high metabolic flux is pushed toward chorismate while all metabolic sinks of chorismate are blocked, thereby funneling a high carbon flux toward pAF (Figures 2B and S1). This nonlinearity highlights the importance of computational analyses to identify interventions that may need to be applied concurrently in order to have a significant effect on the phenotype of interest.

Construction and characterization of computationally designed overproducing strains

Computational modeling identified numerous opportunities for metabolic flux manipulations to enhance pAF production in *E. coli*, including the downregulation of metabolic flux downstream of chorismate biosynthesis and the upregulation of flux upstream of chorismate biosynthesis. These predictions serve as a starting point to experimentally design an engineered *E. coli* strain with an enhanced pAF production capacity. Although we can infer genetic manipulations corresponding to predicted flux interventions using gene-protein-reaction associations in GEMs (Figure 2C), there can be different ways of imposing these flux changes experimentally, not all of which could be directly captured by GEMs. For example, increasing metabolic flux in the shikimate pathway (toward chorismite; Figure 2B) can be achieved experimentally by either upregulating the *aro* family genes or by removing feedback inhibitions in this pathway (even though the latter is not directly captured by stoichiometric models used in this study). In particular, the redundant 3-deoxy-7-phosphoheptlonate synthases AroF, AroG, and AroH are allosterically inhibited by tyrosine, phenylalanine, and tryptophan, respectively (Kikuchi et al., 1997; Ray et al., 1988; Smith et al., 1962), and the transcriptional repressor TyrR reduces the expression of numerous *aro* family genes under conditions of tyrosine abundance (Pittard et al., 2005) (Figure 1A). Thus, we hypothesized that these feedback mechanisms could play a strong role in enhancing or attenuating metabolic flux toward chorismate

synthesis. Therefore, we sought to eliminate these inhibition loops in our engineered strains as a way of increasing metabolic flux toward chorismate predicted by computational studies.

Likewise, because gene knockouts are more amenable to experimental implementation than gene downregulations, we decided to impose the model-guided metabolic flux downregulations by knocking out the genes downstream of chorismate biosynthesis (*pheA*, *trpDE*, *tyrA*, and *entA*) and supplementing the growth medium with the respective aromatic amino acids; this also permitted fine-tuning of these amino acid concentrations in the growth medium to tease out their impact on pAF production. Furthermore, to implement the predicted flux upregulations, we cloned an additional copy of the penultimate gene upstream of chorismate biosynthesis (*aroC*) into episomal expression vectors. We used multiplex genome engineering and recombineering (Sharan et al., 2009; Wang et al., 2009) to construct *E. coli* strains with combinatorial knockouts of *pheA*, *trpDE*, *tyrA*, and *entA*, genes downstream of chorismate biosynthesis. The knockout of aromatic amino acid biosynthesis genes generates auxotrophies for the respective amino acids. To compensate for these auxotrophies, we grew engineered strains in minimal medium supplemented with phenylalanine, tryptophan, and tyrosine. We also cloned *aroC* and placed it under the control of a vanillic-acid-inducible promoter (P_{vanR}) (Meyer et al., 2019) to enable titratable overexpression. To verify whether genes under the control of the two synthetic promoters P_{LtetO} and P_{vanR} could be induced independently of one another, we also constructed a circuit analogous to the one used for *papBAC* and *aroC* expression with GFP (green fluorescent protein) under the control of P_{LtetO} and RFP (red fluorescent protein) under the control of P_{vanR} (Figure S2). We observed that GFP expression was unaffected by vanillic acid concentrations of up to 4 μM and that RFP expression was unaffected by aTc concentrations of up to 32 ng/mL (Figure S3). Finally, we used MAGE (multiplex automated genome engineering) (Wang et al., 2009) to introduce nonsynonymous point mutations that render AroF and AroG allosterically insensitive to tyrosine and phenylalanine, respectively (Kikuchi et al., 1997; Smith et al., 1962). Because AroF and AroG together account for >99% of 3-deoxy-7-phosphoheptlonate synthase activity within metabolically active *E. coli* (Tribe et al., 1976), we did not introduce analogous feedback-inactivating mutations to AroH.

We investigated the effects of these interventions, alone and in combination, on pAF production after 24 h of induction in minimal media supplemented with tyrosine, phenylalanine, and tryptophan (Figure 3A). Titration of vanillic acid to overexpress *aroC* led to modest increases in pAF production, whereas no combination of downstream gene knockouts increased pAF production contrary to our expectation. Feedback-inhibitory mutations did not lead to increases in pAF production on their own, but we observed a > 4-fold increase in pAF production in strains with both AroFG feedback inhibition mutations (*aroFG*-FIM) and a *tyrR* knockout (Δ*tyrR*); this suggests the existence of redundant feedback control mechanisms within the shikimate pathway that regulate flux toward chorismate, leading to the observed epistatic pattern on pAF production.

We hypothesized that other epistatic patterns may exist among the genomic interventions targeted, and we reassessed the effects of gene overexpression and knockout in the context of a feedback-insensitive strain. Tuning the expression of the *papBAC* genes through aTc titration had a strong effect on pAF production in EcNR2.*aroFG*-FIM.Δ*tyrR*, up to a maximal pAF yield of 85 μM after 40 h of induction (0.004 g pAF per g of glucose) (Figure 3B, top panel). *aroC* overexpression via titration of vanillic acid led to no increases in pAF yield when paired with any level of *papBAC* induction. The expression of *aroC* was in fact detrimental to pAF production at high aTc concentrations, possibly because of the high translational load placed on cells expressing both AroC and PapB/A/C in abundance. These same patterns—of pAF production driven predominantly by *papBAC* gene overexpression and of detrimental effects of *aroC* overexpression on pAF production—were also observed in feedback-insensitive strains with additional knockouts in *tyrA*, *pheA*, and *trpDE* (Figure 3B, bottom panel).

Relationship between pAF production and growth in engineered strains

Given our initial observation of a trade-off between pAF production and growth rate (Figure 1D), we sought to determine how the model-guided genomic interventions (Figures 2B and 3A) influence the balance between cellular growth and pAF production in *E. coli*. Our efforts to culture engineered strains with *tyrA*, *pheA*, and *trpDE* knockouts in 96-well plate readers under conditions of *papBAC* induction were unsuccessful, presumably because suboptimal agitation and aeration conditions in such instruments inhibit the growth of strains with fitness-reducing gene knockouts. We thus measured the growth rates of all

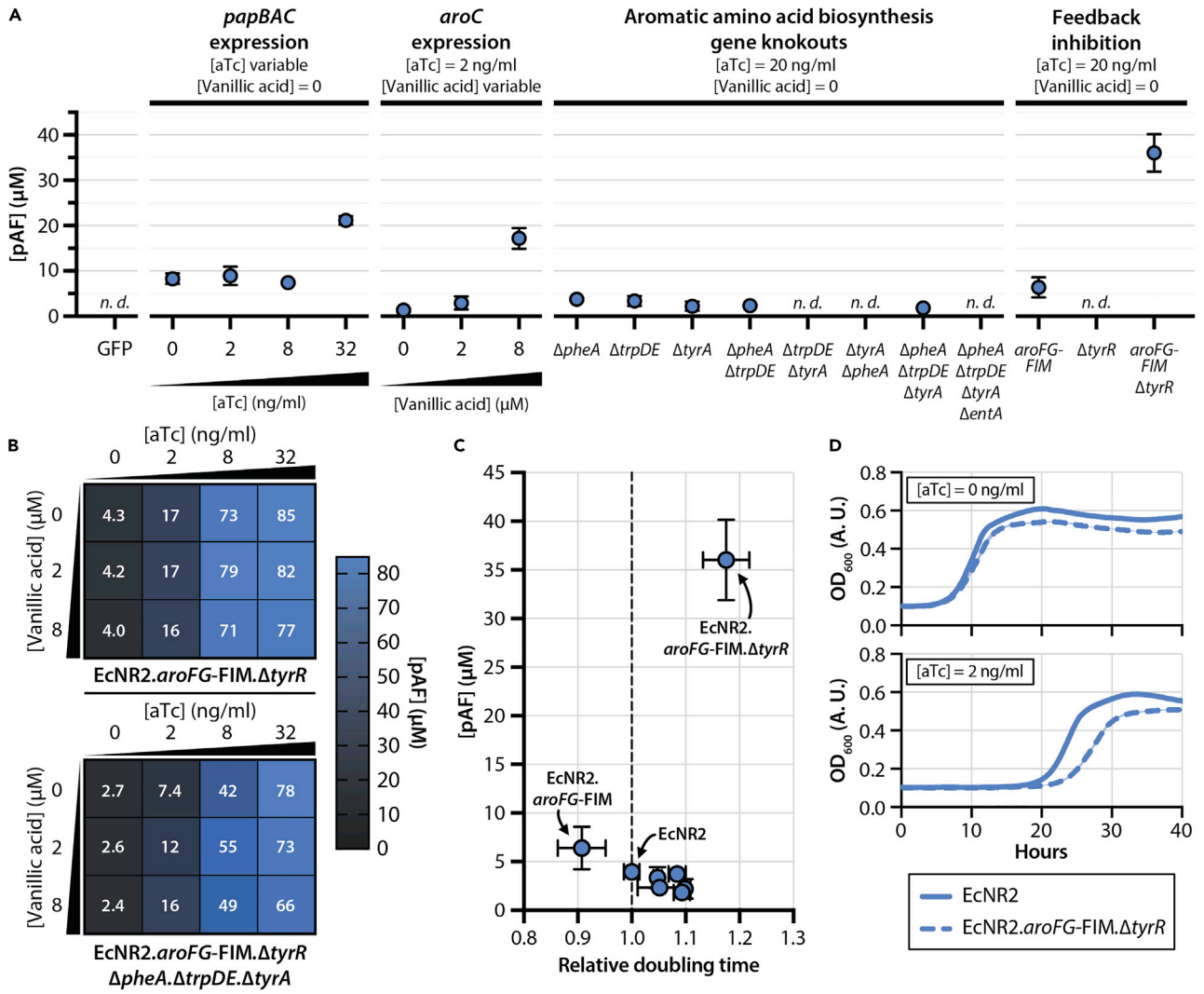


Figure 3. pAF production and growth rate trade-offs in engineered strains

(A) pAF production in strains with modulations in aTc-induced *papBAC* overexpression, vanillic acid-induced *aroC* overexpression, aromatic amino acid biosynthesis gene knockouts, and shikimate pathway feedback inhibitions. "GFP" condition denotes a strain bearing an aTc-inducible GFP overexpression cassette. All strains were grown in M9 minimal medium supplemented with phenylalanine, tryptophan, and tyrosine. Data are represented as mean +/− SD of two technical replicates.

(B) pAF production in strains with shikimate pathway feedback inhibition mutations (*aroFG*-FIM and Δ*tyrR*) and modulated *papBAC* and *aroC* overexpression. The strain in the top panel bears feedback inhibition mutations only, and the strain in the bottom panel bears knockouts in *pheA*, *trpDE*, and *tyrA* in addition to feedback inhibition mutations. Data are represented as the mean of three technical replicates.

(C) pAF production and doubling time (relative to the growth of EcNR2) for strains with different combinations of genomic interventions. Strains in which pAF production was not detected are not included. We observed an epistatic interaction between *aroFG*-FIM and Δ*tyrR*—both interventions in combination lead to substantial increases in pAF titer, whereas they lead to only minor increases in titer on their own. pAF data are represented as mean +/− SD of two technical replicates; doubling time data are represented as mean +/− SD of three biological replicates.

(D) Representative growth curves of EcNR2 and EcNR2.*aroFG*-FIM.Δ*tyrR* bearing aTc-inducible *papBAC* overexpression cassettes at 0 aTc (top panel) and 2 ng/μL aTc (bottom panel). Curves are the mean of three biological replicates.

strains in the absence of *papBAC* expression (aTc = 0) and compared these rates against pAF productions obtained from the same strains in conditions optimal for pAF production (see STAR Methods) (Figures 3C and Table S1). Most engineered strains exhibited longer doubling times than EcNR2 without improving pAF titers. EcNR2.*aroFG*-FIM showed a modestly lower doubling time (0.91 ± 0.04 relative to EcNR2) and modestly higher pAF yields (6.4 ± 2.2 μM), whereas EcNR2.Δ*tyrR* showed a longer doubling time (1.14 ± 0.03 relative to EcNR2) and no detectable pAF production.

EcNR2.*aroFG*-FIM.Δ*tyrR* was nonetheless capable of growing in a plate reader under *papBAC* induction conditions. Compared with EcNR2, the strain showed an 18% slower growth rate (161 ± 5 min for EcNR2.*aroFG*-FIM.Δ*tyrR* versus 137 ± 1 min for EcNR2) in the absence of *papBAC* expression (Figure 3D, top panel). However, both EcNR2 and EcNR2.*aroFG*-FIM.Δ*tyrR* exhibited profound lag phases at low levels of *papBAC* expression ([aTc] = 2 ng/mL; Figure 3D, bottom panel). This suggests that feedback inhibition mutations can increase pAF yields but do not resolve the trade-off between pAF production and growth rate in carbon-limited media.

DISCUSSION

The optimization of heterologous pathways for nsAA production is a promising strategy for engineering microbes with diverse biochemical capabilities. Heterologous expression often places significant metabolic burdens on engineered strains, which may hinder their utility in diverse industrial, biomedical, and environmental contexts. In this study, we leveraged GEMs of metabolism to guide the design of *E. coli* strains capable of producing the nsAA pAF from a heterologous pathway while mitigating disruptions to cellular homeostasis. Previous studies have successfully engineered native *E. coli* metabolism to maximize pAF production in fermentation settings (Masuo et al., 2016), wherein the active growth of cells is of minimal concern and did not take into account the entire metabolism of the host. Here, we built on this prior work by simultaneously accounting for growth and pAF production in our strain engineering efforts and by considering the entire scope of the host's native metabolism using GEMs; this enables us to systematically pinpoint and test the effects of metabolic interventions that span the entire *E. coli* metabolome.

Our initial efforts for pAF production in nonengineered *E. coli* hosts were hindered by a conspicuous trade-off between pAF production titers and growth rate (Figure 1D), presumably because expression of the pAF-producing *papBAC* genes disrupts cellular homeostasis by shunting essential metabolites away from central metabolism. Computational modeling using GEMs of metabolism identified numerous metabolic flux interventions that were predicted to improve the production of pAF for a given minimal required growth rate. These interventions included decreasing metabolic flow downstream of chorismate biosynthesis pathway through the downregulation of genes *tyrA*, *pheA*, *trpDE*, and *entA* and increasing metabolic flux upstream of chorismate biosynthesis through the upregulation of genes in the *aro* family (such as *aroC)*. GEMs of metabolism, which are based on a stoichiometric representation, do not account for the nonlinear effects of regulatory interactions. Therefore, we also decided to consider the elimination of feedback inhibition mechanisms within the shikimate pathway as an alternative to implement flux upregulations to increase metabolic flow toward chorismite as predicted by using GEMs of metabolism. Likewise, we decided to knock out genes *tyrA*, *pheA*, *and trpDE* (instead of implementing the predicted downregulations by computational simulations using GEMs) while supplementing the growth medium with the respective aromatic amino acids. Although the removal of genes involved in the aromatic amino acid biosynthesis (more specifically pheA) has been implemented before to increase pAF production (Masuo et al., 2016), their combination with other interventions that we identified in this study has not previously been explored. In order to optimize the expression levels of genes for pAF production, we constructed a genetic circuit based on a pair of orthogonal titratable gene expression systems (Figure S2). The use of orthogonal titratable promoters for metabolic pathway optimization has been demonstrated successfully in the past (Meyer et al., 2019). This approach obviates the need for complex combinatorial cloning of constitutive gene expression elements (including promoters, ribosome binding sites, and plasmid copy numbers) in order to screen for an optimized pathway (Jeschek et al., 2017).

Interestingly, we found that increased metabolic flow in the shikimate pathway through *aroC* upregulation, either alone or in combination with other predicted metabolic interventions, did not lead to increased pAF production compared with the nonengineered host (Figure 3A); this could be due to the presence of other rate limiting enzymatic reactions upstream of AroC in the shikimate pathway, implying that additional gene upregulations in that pathway are needed to impose the desired effect. However, we observed that imposing an increased metabolic flow toward chorismate through the elimination of feedback inhibition loops (via the elimination of allosteric feedback on AroF and AroG and deletion of the tyrosine-sensitive transcriptional repressor TyrR) led to a measurable increase in pAF production. Contrary to our expectation, combining these feedback elimination interventions with other computationally predicted interventions did not improve pAF production any further. These observations suggest that specific genetic interventions inferred directly from gene-protein-reaction associations in the GEMs of metabolism may not always behave as expected due to regulatory/allosteric effects that are not captured by these models;

this alludes to the need to consider additional regulatory interactions in these pathways in future engineering efforts, e.g., by using large-scale kinetic models of metabolism (Khodayari and Maranas, 2016; Khodayari et al., 2014) (instead of stoichiometric models used in this study), which can take into account such regulatory and allosteric interactions.

The interventions that did increase pAF production in our engineered strain, AroFG-FIM and Δ*tyrR*, showed an epistatic effect on pAF production. Neither AroFG feedback elimination nor *tyrR* knockout raised pAF titers when implemented in isolation but implementing both interventions within the same strain led to substantial increases in pAF titer (Figure 3A). This epistatic pattern could be due to redundant regulatory mechanisms within the shikimate pathway that control chorismate production. Both AroFG and TyrR respond to high tyrosine conditions by reducing flux through the shikimate pathway, and the elimination of feedback loops on both genes is thus needed to restore chorismate production in the tyrosine-supplemented minimal media used in this study. We observed substantial (~20-fold) increases in pAF production in strains with engineered chorismate pathway regulation compared with unmodified strains (Figure 3B). In future studies, we propose that it is possible to achieve further increases in pAF production through more extensive metabolic feedback loop modulation but that these interventions do not necessarily alter the trade-off between pAF production and growth (Figure 3D).

Sourcing nonstandard amino acids through biosynthesis *in vivo* holds promise for more efficient and cost-effective production of synthetic proteins and biomaterials. This strategy could also serve as a mediator of cellular signaling or of the establishment of synthetic cross-feeding channels between members of engineered microbial communities. More specifically, we envision the exchange of nsAAs between organisms capable of producing them and organisms dependent upon them for survival as a promising approach for constructing stable microbial consortia that are minimally affected by surrounding inter-species metabolite exchanges. Engineered cross-feeding channels may also be resilient to invasion by natural strains through the exchange of metabolites that are not a component of natural cellular milieu. Crucially, organisms participating in such cross-feeding channels would need to both produce nsAAs and grow robustly in dynamic environments. Engineering microbial consortia thus requires the simultaneous optimization of two phenotypes—growth and metabolite production.

Conclusions

In this study, we integrated computational systems biology modeling using GEMs of metabolism and synthetic biology techniques to construct an *E. coli* strain with enhanced pAF production capability. We found that the elimination of feedback inhibition loops in the chorismate biosynthesis pathway has the highest impact on enhancing pAF production while minimizing the disruption of cellular homeostasis and growth. Imposing other interventions in pathways downstream of chorismate, which was predicted by our computational models and seemed promising, did not lead to a further increase in pAF production contrary to our expectation. These observations warrant future studies that consider tunable regulatory circuits in this part of the metabolism to design a more efficient pAF-producing strain. Overall, our study sets a basis for the engineering of overproducing strains, which can be used in synthetic microbial consortia that meet the dual objectives of high-level product formation and sustained growth. We anticipate that this approach could be applied to heterologous pathways beyond pAF production such as other nsAAs or biochemicals.

Limitations of the study

Our study identified three classes of genomic interventions that could influence pAF production in *E. coli*: (1) the knockout of genes downstream of chorismate biosynthesis, (2) the upregulation of genes upstream of chorismate biosynthesis, and (3) the elimination of regulatory and allosteric feedback mechanisms that reduce chorismate biosynthesis in conditions of high aromatic amino acid abundance. Although in this study we aimed to explore all these interventions both individually and in combination, we only explored one intervention of the second class—the overexpression of *aroC*, which encodes the final enzyme in chorismate biosynthesis. We did not observe higher pAF titers with *aroC* overexpression, which highlights the possibly that there may exist other rate-limiting reactions upstream of *aroC* that limit flux through the shikimate pathway not explored in our studies. The systematic overexpression of genes upstream of chorismate biosynthesis could reveal these rate-limiting reactions and lead to increased pAF production.

As noted earlier, stoichiometric GEMs of metabolism do not capture interventions of the third class—regulatory and allosteric feedback mechanisms; this limits their capability for predicting promising metabolic intervention

strategies. We did observe increases in pAF production with the inactivation of these feedback mechanisms, namely by eliminating allosteric regulation of AroF/G/H and knockout of the transcriptional repressor *tyrR* (Figure 3A). However, feedback mechanisms are difficult to model computationally and to tune experimentally—for instance, tuning the allosteric response of AroF/G/H would require protein sequence alterations, whose effects are more difficult to predict than genetic up- or downregulations. Although our study highlights feedback modulation as a promising strategy for the biosynthesis of nsAAs in heterologous hosts, the generalizability of this approach is limited in the absence of synthetic biology tools to control these feedback mechanisms.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Computational methods
  - Experimental methods

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104562.

ACKNOWLEDGMENTS

We gratefully acknowledge funding from the Defense Advanced Research Projects Agency (Purchase Request No. HR0011515303, Contract No. HR0011-15-C-0091), the Dutch Research Council (NWO), and National Science Foundation (EF-1935120). DS also acknowledges funding from the Kilachand Multicellular Design Program, the US Department of Energy, Office of Science, Office of Biological & Environmental Research through the Microbial Community Analysis and Functional Evaluation in Soils Science Focus Area Program (m-CAFEs) under contract number DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory, and the NSF Center for Chemical Currencies of a Microbial Planet (C-CoMP, publication #005). We also thank members of the Segrè lab and Paul Muir and Koen Vanderschuren in the Isaacs lab for their invaluable feedback on this work.

AUTHOR CONTRIBUTIONS

ARZ, DS, and FJI conceived the study. ARZ performed all computational simulations. CH performed all experiments except mass spectrometry analyses; PA and TW performed the mass spectrometry analyses. ARZ and CH analyzed the computational and experimental results, respectively. ARZ and CH wrote the first draft of the manuscript. All authors contributed to writing and editing the manuscript.

DECLARATION OF INTERESTS

FJI is a cofounder of Pearl Bio, Inc. Other authors declare no competing interests.

Received: March 20, 2022
Revised: May 16, 2022
Accepted: June 2, 2022
Published: July 15, 2022

REFERENCES

Albayrak, C., and Swartz, J.R. (2014). Direct polymerization of proteins. ACS Synth. Biol. *3*, 353–362. https://doi.org/10.1021/sb400116x.

Amiram, M., Haimovich, A.D., Fan, C., Wang, Y.S., Aerni, H.R., Ntai, I., Moonan, D.W., Ma, N.J., Rovner, A.J., Hong, S.H., et al. (2015). Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. Nat. Biotechnol. *33*, 1272–1279. https://doi.org/10.1038/nbt.3372.

Arranz-Gibert, P., Vanderschuren, K., and Isaacs, F.J. (2018). Next-generation genetic code expansion. Curr. Opin. Chem. Biol. *46*, 203–211. https://doi.org/10.1016/j.cbpa.2018.07.020.

Breit, K.R., Zamudio, B., and Thomas, J.D. (2019). The effects of alcohol and cannabinoid exposure during the brain growth spurt on behavioral

development in rats. Birth Defects. Res. *111*, 760–774. https://doi.org/10.1002/bdr2.1487.

Bynum, M.L., Hackebeil, G.A., Hart, W.E., Laird, C.D., Nicholson, B.L., Siirola, J.D., Watson, J.-P., and Woodruff, D.L. (2021). Pyomo–optimization Modeling in python, Third Edition (Springer Science \& Business Media).

Carothers, J.M., Goler, J.A., and Keasling, J.D. (2009). Chemical synthesis using synthetic biology. Curr. Opin. Biotechnol. *20*, 498–503. https://doi.org/10.1016/j.copbio.2009.08.001.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., Holland, T.A., Keseler, I.M., Kothari, A., Kubo, A., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. *42*, D459–D471. https://doi.org/10.1093/nar/gkt1103.

DeVito, J.A. (2008). Recombineering with tolC as a selectable/counter-selectable marker: remodeling the rRNA operons of Escherichia coli. Nucleic Acids Res. *36*, e4. https://doi.org/10.1093/nar/gkm1084.

Fong, S.S., Burgard, A.P., Herring, C.D., Knight, E.M., Blattner, F.R., Maranas, C.D., and Palsson, B.O. (2005). In silico design and adaptive evolution of Escherichia coli for production of lactic acid. Biotechnol. Bioeng. *91*, 643–648. https://doi.org/10.1002/bit.20542.

Gallagher, R.R., Li, Z., Lewis, A.O., and Isaacs, F.J. (2014). Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. Nat. Protoc. *9*, 2301–2316. https://doi.org/10.1038/nprot.2014.082.

Gaston, M.A., Zhang, L., Green-Church, K.B., and Krzycki, J.A. (2011). The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. Nature *471*, 647–650. https://doi.org/10.1038/nature09918.

Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Linsay, B., and Stevens, R.L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. *28*, 977–982. https://doi.org/10.1038/nbt.1672.

Hong, S.H., Kwon, Y.C., and Jewett, M.C. (2014). Non-standard amino acid incorporation into proteins using Escherichia coli cell-free protein synthesis. Front. Chem. *2*, 34. https://doi.org/10.3389/fchem.2014.00034.

Isaacs, F.J., Carr, P.A., Wang, H.H., Lajoie, M.J., Sterling, B., Kraal, L., Tolonen, A.C., Gianoulis, T.A., Goodman, D.B., Reppas, N.B., et al. (2011). Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science *333*, 348–353. https://doi.org/10.1126/science.1205822.

Isaacs, F.J., Dwyer, D.J., Ding, C., Pervouchine, D.D., Cantor, C.R., and Collins, J.J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. Nat. Biotechnol. *22*, 841–847. https://doi.org/10.1038/nbt986.

Jeschek, M., Gerngross, D., and Panke, S. (2017). Combinatorial pathway optimization for streamlined metabolic engineering. Curr. Opin. Biotechnol. *47*, 142–151. https://doi.org/10.1016/j.copbio.2017.06.014.

Jin, X., Park, O.J., and Hong, S.H. (2019). Incorporation of non-standard amino acids into proteins: challenges, recent achievements, and emerging applications. Appl. Microbiol. Biotechnol. *103*, 2947–2958. https://doi.org/10.1007/s00253-019-09690-6.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. *42*, D199–D205. https://doi.org/10.1093/nar/gkt1076.

Keasling, J., Garcia Martin, H., Lee, T.S., Mukhopadhyay, A., Singer, S.W., and Sundstrom, E. (2021). Microbial production of advanced biofuels. Nat. Rev. Microbiol. *19*, 701–715. https://doi.org/10.1038/s41579-021-00577-w.

Khodayari, A., and Maranas, C.D. (2016). A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. Nat. Commun. *7*, 13806. https://doi.org/10.1038/ncomms13806.

Khodayari, A., Zomorrodi, A.R., Liao, J.C., and Maranas, C.D. (2014). A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data. Metab. Eng. *25*, 50–62. https://doi.org/10.1016/j.ymben.2014.05.014.

Kikuchi, Y., Tsujimoto, K., and Kurahashi, O. (1997). Mutational analysis of the feedback sites of phenylalanine-sensitive 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase of Escherichia coli. Appl. Environ. Microbiol. *63*, 761–762. https://doi.org/10.1128/aem.63.2.761-762.1997.

Lajoie, M.J., Rovner, A.J., Goodman, D.B., Aerni, H.R., Haimovich, A.D., Kuznetsov, G., Mercer, J.A., Wang, H.H., Carr, P.A., Mosberg, J.A., et al. (2013). Genomically recoded organisms expand biological functions. Science *342*, 357–360. https://doi.org/10.1126/science.1241459.

Long, M.R., Ong, W.K., and Reed, J.L. (2015). Computational methods in metabolic engineering for strain design. Curr. Opin. Biotechnol. *34*, 135–141. https://doi.org/10.1016/j.copbio.2014.12.019.

Lutz, R., and Bujard, H. (1997). 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. https://doi.org/10.1093/nar/25.6.1203.

Mandell, D.J., Lajoie, M.J., Mee, M.T., Takeuchi, R., Kuznetsov, G., Norville, J.E., Gregg, C.J., Stoddard, B.L., and Church, G.M. (2015). Biocontainment of genetically modified organisms by synthetic protein design. Nature *518*, 55–60. https://doi.org/10.1038/nature14121.

Marchand, J.A., Neugebauer, M.E., Ing, M.C., Lin, C-I, Pelton, J.G., and Chang, M.C.Y. (2019). Discovery of a pathway for terminal-alkyne amino acid biosynthesis. Nature *567*, 420–424. https://doi.org/10.1038/s41586-019-1020-y.

Masuo, S., Zhou, S., Kaneko, T., and Takaya, N. (2016). Bacterial fermentation platform for producing artificial aromatic amines. Sci. Rep. *6*, 25764. https://doi.org/10.1038/srep25764.

Mehl, R.A., Anderson, J.C., Santoro, S.W., Wang, L., Martin, A.B., King, D.S., Horn, D.M., and Schultz, P.G. (2003). Generation of a bacterium with a 21 amino acid genetic code. J. Am. Chem. Soc. *125*, 935–939. https://doi.org/10.1021/ja0284153.

Meyer, A.J., Segall-Shapiro, T.H., Glassey, E., Zhang, J., and Voigt, C.A. (2019). Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. Nat. Chem. Biol. *15*, 196–204. https://doi.org/10.1038/s41589-018-0168-3.

Mohammadi Nargesi, B., Trachtmann, N., Sprenger, G.A., and Youn, J.W. (2018). Production of p-amino-L-phenylalanine (L-PAPA) from glycerol by metabolic grafting of Escherichia coli. Microb Cell Fact *17*, 149. https://doi.org/10.1186/s12934-018-0996-6.

Murphy, C.D., Schaffrath, C., and O'Hagan, D. (2003). Fluorinated natural products: the biosynthesis of fluoroacetate and 4-fluorothreonine in Streptomyces cattleya. Chemosphere *52*, 455–461. https://doi.org/10.1016/S0045-6535(03)00191-7.

Neidhardt, F.C., Bloch, P.L., and Smith, D.F. (1974). Culture medium for enterobacteria. J. Bacteriol. *119*, 736–747. https://doi.org/10.1128/jb.119.3.736-747.1974.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., and Palsson, B. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. Mol. Syst. Biol. *7*, 535. https://doi.org/10.1038/msb.2011.65.

Pitera, D.J., Paddon, C.J., Newman, J.D., and Keasling, J.D. (2007). Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. Metab. Eng. *9*, 193–207. https://doi.org/10.1016/j.ymben.2006.11.002.

Pittard, J., Camakaris, H., and Yang, J. (2005). The TyrR regulon. Mol. Microbiol. *55*, 16–26. https://doi.org/10.1111/j.1365-2958.2004.04385.x.

Ranganathan, S., Suthers, P.F., and Maranas, C.D. (2010). OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput. Biol. *6*, e1000744. https://doi.org/10.1371/journal.pcbi.1000744.

Ranganathan, S., Tee, T.W., Chowdhury, A., Zomorrodi, A.R., Yoon, J.M., Fu, Y., Shanks, J.V., and Maranas, C.D. (2012). An integrated computational and experimental study for overproducing fatty acids in Escherichia coli. Metab. Eng. *14*, 687–704. https://doi.org/10.1016/j.ymben.2012.08.008.

Ray, J.M., Yanofsky, C., and Bauerle, R. (1988). Mutational analysis of the catalytic and feedback sites of the tryptophan-sensitive 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase of Escherichia coli. J. Bacteriol. *170*, 5500–5506. https://doi.org/10.1128/jb.170.12.5500-5506.1988.

Rovner, A.J., Haimovich, A.D., Katz, S.R., Li, Z., Grome, M.W., Gassaway, B.M., Amiram, M., Patel, J.R., Gallagher, R.R., Rinehart, J., and

Isaacs, F.J. (2015). Recoded organisms engineered to depend on synthetic amino acids. Nature *518*, 89–93. https://doi.org/10.1038/nature14095.

Santos, C.N.S., Koffas, M., and Stephanopoulos, G. (2011). Optimization of a heterologous pathway for the production of flavonoids from glucose. Metab. Eng. *13*, 392–400. https://doi.org/10.1016/j.ymben.2011.02.002.

Shapiro, D., Mandava, G., Yalcin, S.E., Yalcin, S., Arranz-Gibert, P., Dahl, P.J., Dahl, P., Shipps, C., Gu, Y., Srikanth, V., et al. (2022). Protein nanowires with tunable functionality and programmable self-assembly using sequence-controlled synthesis. Nat. Commun. *13*, 829. https://doi.org/10.1038/s41467-022-28206-x.

Sharan, S.K., Thomason, L.C., Kuznetsov, S.G., and Court, D.L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. Nat. Protoc. *4*, 206–223. https://doi.org/10.1038/nprot.2008.227.

Smith, L.C., Ravel, J.M., Lax, S.R., and Shive, W. (1962). The control of 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthesis by phenylalanine and tyrosine. J. Biol. Chem. *237*, 3566–3570. https://doi.org/10.1016/s0021-9258(19)70858-8.

Snitkin, E.S., and Segrè, D. (2011). Epistatic interaction maps relative to multiple metabolic phenotypes. PLoS Genet. *7*, e1001294. https://doi.org/10.1371/journal.pgen.1001294.

Des Soye, B.J., Patel, J.R., Isaacs, F.J., and Jewett, M.C. (2015). Repurposing the translation apparatus for synthetic biology. Curr. Opin. Chem. Biol. *28*, 83–90. https://doi.org/10.1016/j.cbpa.2015.06.008.

Sulzbach, M., and Kunjapur, A.M. (2020). The pathway less Traveled: engineering biosynthesis of nonstandard functional groups. Trends Biotechnol. *38*, 532–545. https://doi.org/10.1016/j.tibtech.2019.12.014.

Suástegui, M., Yu Ng, C., Chowdhury, A., Sun, W., Cao, M., House, E., Maranas, C.D., and Shao, Z. (2017). Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in Saccharomyces cerevisiae for high production of polymer and drug precursors. Metab. Eng. *42*, 134–144. https://doi.org/10.1016/j.ymben.2017.06.008.

Tribe, D.E., Camakaris, H., and Pittard, J. (1976). Constitutive and repressivle enzymes of the common pathway of aromatic biosynthesis in Escherichia coli K-12: regulation of enzyme synthesis at different growth rates. J. Bacteriol. *127*, 1085–1097. https://doi.org/10.1128/jb.127.3.1085-1097.1976.

Tsoi, R., Wu, F., Zhang, C., Bewick, S., Karig, D., and You, L. (2018). Metabolic division of labor in microbial systems. Proc. Natl. Acad. Sci. USA *115*, 2526–2531. https://doi.org/10.1073/pnas.1716888115.

Vanderschuren, K., Arranz-Gibert, P., Khang, M., Hadar, D., Gaudin, A., Yang, F., Folta-Stogniew, E., Saltzman, W.M., Amiram, M., and Isaacs, F.J. (2022). Tuning protein half-life in mouse using sequence-defined biopolymers functionalized with lipids. Proc. Natl. Acad. Sci. U. S. A. *119*. e2103099119. https://doi.org/10.1073/pnas.2103099119.

Walensky, L.D., and Bird, G.H. (2014). Hydrocarbon-stapled peptides: principles, practice, and progress. J. Med. Chem. *57*, 6275–6288. https://doi.org/10.1021/jm4011675.

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature *460*, 894–898. https://doi.org/10.1038/nature08187.

Wannier, T.M., Ciaccia, P.N., Ellington, A.D., Filsinger, G.T., Isaacs, F.J., Javanmardi, K., Jones, M.A., Kunjapur, A.M., Nyerges, A., Pal, C., et al. (2021). Recombineering and MAGE. Nat. Rev. Methods Primers *1*, 7. https://doi.org/10.1038/s43586-020-00006-x.

Williams, R.M. (1999). Asymmetric syntheses of unnatural amino acids and hydroxyethylene Peptide isosteres. Methods Mol. Med. *23*, 339–356. https://doi.org/10.1385/0-89603-517-4:339.

Xu, P., Ranganathan, S., Fowler, Z.L., Maranas, C.D., and Koffas, M.A. (2011). Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab. Eng. *13*, 578–587. https://doi.org/10.1016/j.ymben.2011.06.008.

Xuan, W., Yao, A., and Schultz, P.G. (2017). Genetically encoded fluorescent probe for detecting Sirtuins in living cells. J. Am. Chem. Soc. *139*, 12350–12353. https://doi.org/10.1021/jacs.7b05725.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., et al. (2011). Metabolic engineering of Escherichia coli for direct production of 1, 4-butanediol. Nat. Chem. Biol. *7*, 445–452. https://doi.org/10.1038/nchembio.580.

Zimmerman, E.S., Heibeck, T.H., Gill, A., Li, X., Murray, C.J., Madlansacay, M.R., Tran, C., Uter, N.T., Yin, G., Rivers, P.J., et al. (2014). Production of site-specific antibody-drug conjugates using optimized non-natural amino acids in a cell-free expression system. Bioconjug Chem *25*, 351–361. https://doi.org/10.1021/bc400490z.

Zomorrodi, A.R., Suthers, P.F., Ranganathan, S., and Maranas, C.D. (2012). Mathematical optimization applications in metabolic networks. Metab. Eng. *14*, 672–686. https://doi.org/10.1016/j.ymben.2012.09.005.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Escherichia coli EcNR2	Addgene	Addgene #26931
Chemicals, peptides, and recombinant proteins		
LB growth medium	AmericanBio	N/A
Solid chemicals: amino acids (phenylalanine, tryptophan, tyrosine), antibiotics (carbenicillin, spectinomycin), inducers (anhydrotetracycline, vanillic acid), trace minerals (Table S2), vitamins (Table S2)	SigmaAldrich	N/A
Colicin E1 for *tolC* counterselections	This paper	N/A
Experimental models: Organisms/strains		
E. coli EcNR2.ΔpheA, ΔtrpDE, ΔtyrA, ΔpheA.ΔtrpDE, ΔtrpDE.ΔtyrA, ΔtyrA.ΔpheA, ΔpheA.ΔtrpDE.ΔtyrA, ΔpheA.ΔtrpDE.ΔtyrA.entA	This paper	CFH748-755
Oligonucleotides		
MAGE oligo for *aroF*-FIM: T*T*CAGTTGTTCCGGAGTCATTAAAACCTGTTCGTCG GTAATATGTACTTTATTCAGCGCGTCTTTTTGcatGATG GCGATCCTGTTTAT	This paper	oCFH007
MAGE oligo for *aroG*-FIM: T*G*CGCCCCAGCTCATCAGGTCAGCGAGATATTGT GGGGTGATCATGTTGAGAAACTCACCTGCCGCTGG CAGACCGCTGTCGTTAATATC	This paper	oCFH008
Primers for cloning pPAF-v5 and pGFP-RFP	This paper	rCFH404-415
MAGE oligos for knockout of *tolC* insertions at *pheA*, *trpDE*, *tyrA*, and *entA*	This paper	oCFH026-29
Primers for amplifying *tolC* cassettes to recombiner into *pheA*, *trpDE*, *tyrA*, *entA*, and *tyrR*	This paper	rCFH334-337, 344-347, 511-512
Recombinant DNA		
pPAF-v5	This paper	CFH489
pGFP-RFP	This paper	CFH496
papBAC synthetic genes	GenScript	N/A
Software and algorithms		
Python scripts for OptForce simulations	This paper	https://doi.org/10.5281/zenodo.6569900
Python v2.7.11 for Unix	Python Software Foundation	www.python.org
MATLAB r2019a for doubling time calculations	MathWorks Inc	https://www.mathworks.com
Gurobi Optimizer v9.0.0	Gurobi Optimization	https://www.gurobi.com
COBRApy Toolbox v0.9.1	cobrapy - constraint-based metabolic modeling in Python	https://opencobra.github.io/cobrapy/
Other		
SCC Shared Computing Cluster, SCC1, 2 fourteen-core 2.4 GHz Intel Xeon E5-2680v4, 256 Gb RAM, 2 nodes	Boston University Research Computing, Massachusetts Green High Performance Computing Center (MGHPCC), Holyoke, MA	https://www.bu.edu/tech/support/research/computing-resources/scc/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Yale West Campus Analytical Core	Yale University	https://wcac.yale.edu
Experimental data and MATLAB analysis code for Figures 1 and 3	This paper	https://doi.org/10.5281/zenodo.6569900

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Daniel Segrè (dsegre@bu.edu).

Materials availability

All strains and plasmids generated in this study are available upon request to Farren J. Issacs (farren.isaacs@yale.edu).

Data and code availability

- All experimental data related to the main Figures 1 and 3 has been deposited at Zenodo (DOI: 10.5281/zenodo.6569900). DOI is also listed in the key resources table. Any other data reported in this paper will be available via the lead contact upon request.
- All original code has been deposited at Zenodo (DOI: 10.5281/zenodo.6569900) and is publicly available as of the date of publication. DOI is also listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.
- Experimental data are represented as the mean of technical or biological replicates. Details for each data representation can be found in the figure legends.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All strains in this study are derived from *Escherichia coli* EcNR2 (Wang et al., 2009), which is available from Addgene (ID 26931). EcNR2 is F^-. LB min media supplemented with 50 μg/ml carbenicillin was used for routine strain growth and cloning. Strains were grown at 34°C.

METHOD DETAILS

Computational methods

OptForce algorithm: In brief, OptForce (Ranganathan et al., 2010) first characterizes the phenotypic space of a reference (e.g., wild-type) strain using any available metabolic flux data by minimizing and maximizing the flux of each reaction in the network subject to these experimental flux data. It next identifies the phenotypic space of an overproducing strain in a similar fashion by imposing a constraint on the minimum required biomass production (growth) in the network as well as a desired production level for the target product. By super-imposing the flux ranges in the reference and overproducing strains for each reaction, the list of reactions whose fluxes must increase, decrease, or shrink to zero in order to achieve the target production level of the desired product is identified. This list of reactions is then used to identify a minimal set of direct metabolic flux interventions (upregulations, downregulations, or knockouts) to achieve a desired production level for the biochemical of interest. In doing so, OptForce simulates a worst-case scenario where the network fights against imposed interventions by maximizing the minimum product formation. OptForce has been previously used successfully to design overproducing strains (Ranganathan et al., 2012; Xu et al., 2011).

Details of OptForce implementation: To characterize the phenotypic space of the wild-type strain, we used experimental flux data for 35 reactions in the central metabolism of *E. coli* from a previous study (Ranganathan et al., 2012). We performed the OptForce simulations by requiring a biomass production of at least 20% of theoretical maximum, for which the maximum achievable pAF production (per 100 mol of glucose) is predicted to be 44.27 (~80% of theoretical maximum). Therefore, we set 80% of theoretical maximum as

our desired pAF production level. In our analyses, reactions with no gene association in the metabolic model were prevented from any type of manipulation. Furthermore, reactions associated with in silico or *in vivo* essential genes in the minimal M9 medium were prevented from being removed. Re-running the OptForce with different thresholds for growth and pAF production did not affect the set of identified metabolic interventions. All simulations were performed in Python 2 using customized scripts (see key resources table). The OptForce optimization problem was solved using the Gurobi solver (https://www.gurobi.com) in Pyomo, an optimization modeling environment in Python (Bynum et al., 2021).

Experimental methods

Materials, strains, and media: All strains in this study are derived from *Escherichia coli* EcNR2 (Wang et al., 2009). LB min media from AmericanBio (Canton, MA) was used for routine strain growth and cloning. Growth and pAF production assays were performed in M9 minimal medium supplemented with 0.25 μg/L betaine and vitamin and trace mineral mixes from Neidhardt EZ rich defined medium (Neidhardt et al., 1974) (see Table S2 for media components). Where indicated, M9 medium was also supplemented with 0.2 mM tyrosine, 0.4 mM phenylalanine, and 0.1 mM tryptophan to support the growth of strains with knockouts in aromatic amino acid biosynthesis genes. 0.4% (w/v) glucose was used as a carbon source for all M9 media formulations. 50 μg/ml carbenicillin was used for the growth of EcNR2 and derivatives in all conditions. Strains harboring *papBAC*, *aroC*, and GFP expression plasmids were cultured in carbenicillin and 95 μg/ml spectinomycin.

Cloning and strain engineering: Synthetic DNA fragments encoding codon-optimized *papA*, *papB*, and *papC* genes from *Pseudomonas fluorescens* were synthesized by GenScript (Piscataway, NJ). Primers and single-stranded oligonucleotides used for cloning and multiplex genome engineering (MAGE) were synthesized by Integrated DNA Technologies (Coralville, IA). Synthetic inducible circuits were generated by cloning DNA fragments into linearized plasmids with a p15A origin of replication and a spectinomycin resistance cassette. Linear DNA fragments were amplified using high-fidelity PCR kits from Kapa Biosystems sourced from Millipore Sigma (Burlington, MA). Gibson assembly (New England Biolabs; Boston, MA) was used for plasmid assembly.

MAGE (Wang et al., 2009) and recombineering (Sharan et al., 2009) were used to implement all genomic perturbations described in the study. Small genomic interventions (i.e., *aroF* and *aroG* feedback inhibition mutations) were implemented by performing 3 rounds of MAGE using 90-mer single-stranded DNA oligonucleotides encoding the desired mutation. Cultures were then plated on LB agar to isolate individual clones, which were then screened for the desired mutation(s) using multiplex allele-specific colony PCR (Isaacs et al., 2004). For scarless gene-scale deletions, we first recombineered the counter-selectable marker *tolC* (DeVito, 2008) into the locus to be deleted and selected for recombinants by plating on LB + 0.1% sodium dodecyl sulfate. We then used MAGE to delete *tolC* and selected for scarless mutants by growing cultures for 8 h in LB containing colicin E1. Cultures passing the liquid selection were then plated on LB agar to isolate individual clones. All genomic interventions were then confirmed via Sanger sequencing.

Growth assays: Cells were plated for single colonies on LB agar containing the appropriate antibiotics and grown overnight. Replicate (n = 3) single colonies were picked and inoculated into 3 mL LB min broth with the appropriate antibiotics. Once cell cultures reached mid-log phase (OD_{600} 0.4-0.5), they were diluted 1:100 into 3 mL M9 media without inducers. Cells were again grown to mid-log, OD-normalized to 0.4, and washed 1x via centrifugation with M9 medium to remove residual LB medium. 3 μL of OD-normalized culture was then inoculated into 147 μL M9 with appropriate inducers in a 96-well bioassay plate. Cultures were grown with shaking at 34°C in a Biotek Synergy HT microplate reader and OD_{600} was measured every 10 min. Growth curves were analyzed using MATLAB scripts written to calculate doubling times.

pAF production assays: Cells were plated for single colonies on LB agar containing the appropriate antibiotics and grown overnight. Single colonies were picked and inoculated into 3 mL LB min broth with the appropriate antibiotics. Upon reaching mid-log phase, cultures were OD-normalized to 0.4 and washed 1x via centrifugation with M9 medium to remove residual LB medium. OD-normalized cultures were diluted 1:100 into 3 mL of fresh M9 medium with supplements (see "Materials, strains, and media"). Cells were grown at 34°C with shaking for 5 h, and *papBAC* and *aroC* were induced with the addition of aTc and vanillic

acid, respectively, to the cultures. Cells were incubated at 34°C with shaking for 24 or 40 h. After the induction period, cultures were centrifugated and supernatant was collected for analysis via LC/MS.

Sample Preparation and Liquid Chromatography/Mass Spectrometer analysis (LC/MS analysis): Supernatant from bacterial cell cultures was collected. To precipitate protein, the supernatant was treated with a 1:10 volume of 100% (w/v) ice-cold TCA, incubated on ice for 20 min, then centrifuged at 17,000 x g for 10 min. This supernatant contained the compounds of interest. Samples were injected onto an Agilent Eclipse Plus C18 RRHD column (2.1 × 50mm, 1.8um particle size) and run over a gradient of 8 min. The mobile phase consisted of a linear gradient of (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile using LCMS grade solvents (Optima, Fisher Chemical). The gradient parameters were as follows: 0-1 min, 10% B (isocratic wash); 1-8 min, 10-95% B; 8-10 min, 95-10% B. The flow rate throughout the procedure was 0.5 mL/min. Injection volume was 12ul. The source gas temperature was set at 280 C at 11L/min, sheath gas was set to 350 C at 11 L/min. Nebulizer was set at 40 psig. Positive ion mode capillary voltage was 5.5 kV, and the nozzle voltage set at 2.0 kV. Separation and analysis were performed using an Agilent 1290 Infinity UPLC system coupled to an Agilent 6550 QToF iFunnel mass spectrometer in ESI + mode. Data was collected in ms mode, scan range of 110 -1700, at 3 spectra/sec. Lock mass spray was employed for all analysis. pAF concentration was quantified using standards consisting of known concentrations of pAF dissolved in supernatant of EcNR2 grown in M9 minimal medium (Figure S4). Reference standards were generated in technical triplicate for each LC/MS experiment. A standard curve was generated from extracted ion chromatograms using the mean of the area under the curve of reference standards to interpolate pAF concentration in biological samples (Figure S5).