




A platform for distributed production of synthetic nitrated proteins in live bacteria

Received: 31 May 2022

Accepted: 13 April 2023

Published online: 15 May 2023

 Check for updates

Neil D. Butler¹, Sabyasachi Sen¹, Lucas B. Brown² , Minwei Lin¹ & Aditya M. Kunjapur¹  

The incorporation of the nonstandard amino acid *para*-nitro-L-phenylalanine (pN-Phe) within proteins has been used for diverse applications, including the termination of immune self-tolerance. However, the requirement for the provision of chemically synthesized pN-Phe to cells limits the contexts where this technology can be harnessed. Here we report the construction of a live bacterial producer of synthetic nitrated proteins by coupling metabolic engineering and genetic code expansion. We achieved the biosynthesis of pN-Phe in *Escherichia coli* by creating a pathway that features a previously uncharacterized nonheme diiron *N*-monooxygenase, which resulted in pN-Phe titers of $820 \pm 130 \mu\text{M}$ after optimization. After we identified an orthogonal translation system that exhibited selectivity toward pN-Phe rather than a precursor metabolite, we constructed a single strain that incorporated biosynthesized pN-Phe within a specific site of a reporter protein. Overall, our study has created a foundational technology platform for distributed and autonomous production of nitrated proteins.

The modification of proteins to contain nitroaromatic functionalities can facilitate unique interactions due to the electron delocalizing properties of nitro groups in aromatic systems. However, none of the 20 standard amino acids contain nitroaromatic functionality nor ring substituents that exhibit similar levels of electron-withdrawing potential. One approach to overcome this limitation is to incorporate nitroaromatic nonstandard amino acids, also referred to as noncanonical amino acids or unnatural amino acids, at specified protein residues in live cells using genetic code expansion technology^{1–3}. Using this approach, the nitroaromatic nonstandard amino acid *para*-nitro-L-phenylalanine (pN-Phe) has been of heightened interest given its diversity of applications (Fig. 1a), including as a peptide distance marker³, an enhancer of activity for nitroreductase enzymes⁴ and a means to break immune self-tolerance^{5–7}. However, these demonstrations relied on the supplementation of chemically synthesized pN-Phe to bacterial cultures for the production of nitrated proteins, which can limit the contexts where these proteins or cells are used.

One example of where proteins that contain pN-Phe have been used is for immunomodulation to aid in the treatment of autoimmune disorders or cancer. By substituting a single surface residue of an

autologous protein with pN-Phe and administering it to mouse models, Schultz and colleagues demonstrated the termination of tolerance toward the wild-type autologous protein, enabling robust IgG antibody production against the protein. This immunogenic enhancement is believed to occur due to the π electron delocalizing properties of the nitroaryl group within aromatic systems⁸ and the recognition of these neoepitopes by CD4⁺ T cells during binding to the peptide and major histocompatibility complex class II (MHC-II) complex⁷. The T cells subsequently activate B cells that produce antibodies capable of binding to the wild-type protein. A broad range of proteins has been targeted using this strategy, including tumor necrosis factor α (TNF α), complement component 5a (C5a), receptor activator of nuclear factor- κ B ligand (RANKL), human epidermal growth factor receptor 2 (HER2) and programmed death-ligand 1 (PD-L1) (refs. 5–7,9–11). Compared to other nonstandard amino acids, pN-Phe proved the most effective at eliciting an immune response and has been reported in the largest number of investigations⁷. However, one major drawback of the approach of administering nitrated proteins was the need for aggressive immunization schedules ranging from four injections over 28 days⁷ to eight injections over 17 days⁵.

¹Department of Chemical & Biomolecular Engineering, University of Delaware, Newark, DE, USA. ²Systems, Synthetic, and Physical Biology Program, Rice University, Houston, TX, USA. ✉e-mail: kunjapur@udel.edu

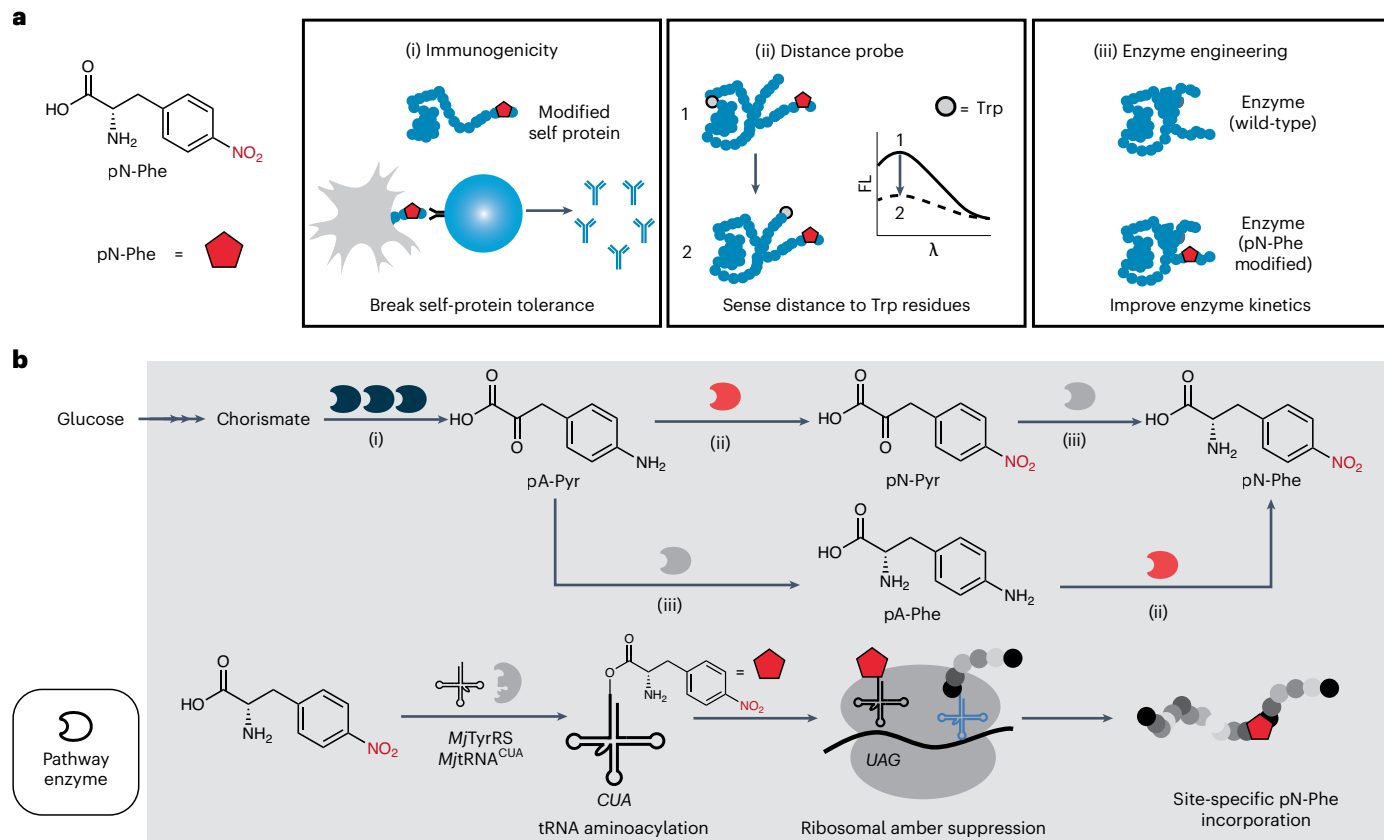


Fig. 1 | De novo biosynthesis and site-specific incorporation of pN-Phe.

a, Within the field of genetic code expansion, the site-specific incorporation of *para*-nitro-L-phenylalanine (pN-Phe) has been used in several applications. Those depicted are as follows: (i) incorporation into self-proteins for immunochemical termination of tolerance toward an antigen and polyclonal antibody production (antigen-presenting cell delivering modified antigen to a T-cell is shown); (ii) use of pN-Phe as a fluorescent quencher to measure relative distance toward tryptophan and (iii) substitution within the binding pocket of enzymes for improved enzyme kinetics. **b**, Schematic representation of our de novo synthesis

pathway for pN-Phe in *E. coli* that can subsequently be used for site-specific protein incorporation. The de novo biosynthesis pathway makes use of (i) three heterologous enzymes from the chloramphenicol biosynthesis pathway in *Streptomyces venezuelae* that generate pA-Phe from chorismate known as papABC (shown in blue to indicate prior use in engineered pathways); (ii) an *N*-oxygenase with activity on pA-Phe or pA-Pyr (shown in red to indicate a lack of prior use in engineered pathways) and (iii) an endogenous, promiscuous aminotransferase capable of converting aromatic keto acids to amino acids (shown in gray to indicate that it is an endogenous enzyme).

Microbial biosynthesis of nonstandard amino acids offers a potential solution to many limitations of biomanufacturing proteins that contain an expanded genetic code by forming the nonstandard amino acids intracellularly and on-demand before incorporation within proteins^{12,13}. The generation of a bacterial strain that can autonomously access an expanded genetic code could also offer a new strategy to solve the problem of aggressive immunization associated with the use of pN-Phe for immunomodulatory applications by producing nitrated proteins within a patient. Additionally, a live bacterial vector could offer benefits of tropism for regions of disease¹⁴, high antigen-loading potential (potentially 100,000 copies displayed/cell in *Escherichia coli*)¹⁵ and adjuvanticity arising from the bacterial cell¹⁶.

Toward the on-demand production of proteins with nonstandard functionalities more generally, a handful of nonstandard amino acids have been metabolically biosynthesized and incorporated in distinct engineering efforts, including *para*-amino-L-phenylalanine (pA-Phe)¹⁷, L-sulfotyrosine¹⁸, L-dihydroxyphenylalanine¹⁹, 4-nitro-L-tryptophan²⁰, 5-hydroxy-L-tryptophan²¹ and L-propargylglycine²². While substantial advancements, most of these nonstandard amino acids are found in nature and were produced in model microbes by transplanting naturally occurring pathways. However, to our knowledge, pN-Phe has not been found in nature. Indeed, natural production of nitro compounds is relatively rare, with only approximately 200 naturally produced nitro compounds previously identified²³. As such, pN-Phe biosynthesis would require a de novo pathway design.

To construct a metabolic pathway that achieves biosynthesis of pN-Phe, we anticipated that nitro group formation could proceed from aromatic amines via oxidation of the amine, which we saw as a promising route for pN-Phe synthesis for two reasons. First, the metabolic synthesis of the amine precursor, pA-Phe, has been demonstrated from multiple natural gene clusters (commonly referred to as papABC^{17,24,25}). Second, a promising amine-oxidizing enzyme, or nonheme diiron *N*-monooxygenase (ObiL, otherwise referred to in the literature as ObaC), was recently discovered in *Pseudomonas fluorescens* strain ATCC 39502 in the biosynthetic gene cluster for the antibiotic obaflourin. Prior work suggests that this *N*-oxygenase has native activity on *para*-aminophenylpyruvate (pA-Pyr), the immediate precursor to pA-Phe in metabolic synthesis^{26,27}. Multiple *N*-oxygenases have demonstrated activity when expressed in *E. coli*, including AurF²⁸ (with a native substrate of *para*-aminobenzoic acid) and CmlI²⁹ (with a native substrate of the amine precursor to chloramphenicol). However, there are no reported examples of engineered biosynthetic pathways that feature *N*-oxygenases to our knowledge.

Here we constructed and tested a de novo biosynthetic pathway for pN-Phe in *E. coli* that used four heterologous enzymes, including an *N*-oxygenase enzyme for nitro group formation and the papABC operon from *Streptomyces venezuelae*. Through chassis engineering and *N*-oxygenase bioprospecting, we were able to improve pN-Phe titers to near millimolar levels and decrease the accumulation of the nonstandard amino acid intermediate pA-Phe. We then identified an

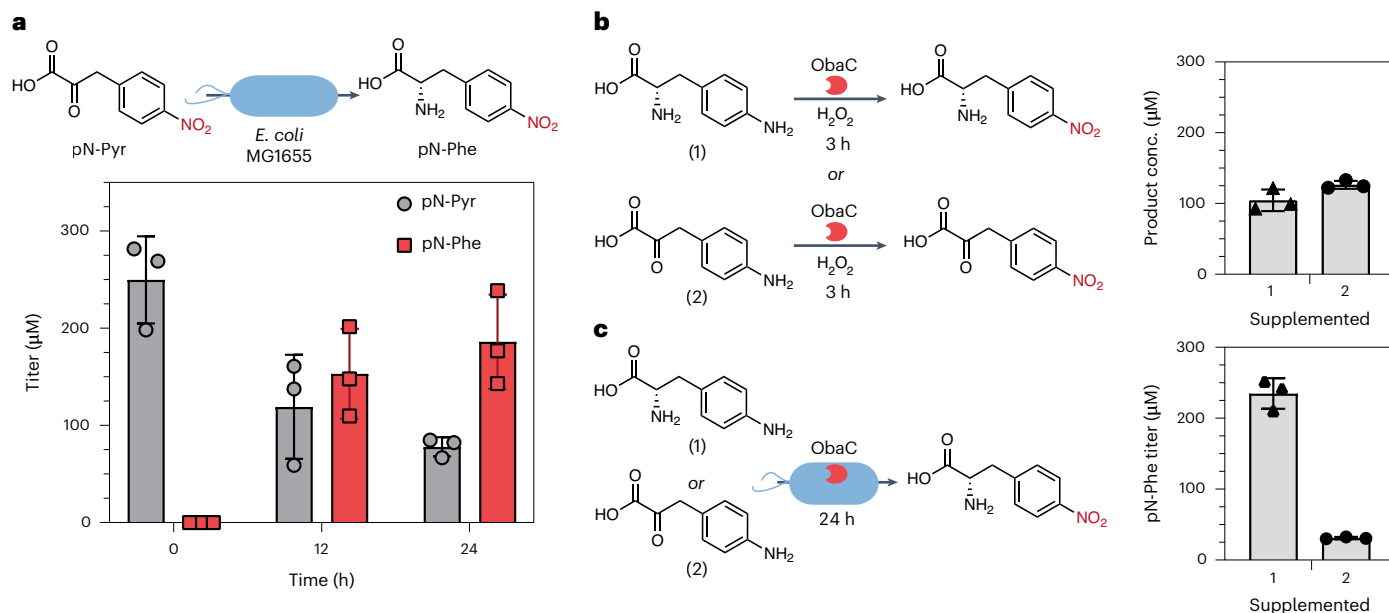


Fig. 2 | Initial characterization for de novo synthesis pathway. a, Evaluation of the native aminotransferase activity of *E. coli* MG1655 for the conversion of pN-Pyr to pN-Phe when fermentative cultures in LB media are supplemented with 250 μM pN-Pyr at midexponential phase. **b**, In vitro investigation of ObiL activity

on pA-Phe and pA-Pyr with 1 mM substrate provided. **c**, Conversion of pA-Phe or pA-Pyr to pN-Phe by *E. coli* strains transformed to express ObiL and cultivated in LB media. Samples sizes are $n = 3$ using biological replicates. Data shown are mean \pm s.d.

engineered variant of the tyrosyl-tRNA synthetase and associated tRNA from *Methanocaldococcus jannaschii* (*Mj*TyrRS and *Mj*tRNA^{Tyr}_{CUA}) that was capable of selectively charging tRNA with pN-Phe rather than Tyr or the metabolic pathway intermediate pA-Phe. By coupling this system with our synthesis pathway, in addition to further genomic edits and gene expression improvements within the strain, we established an engineered *E. coli* capable of de novo synthesis and incorporation of pN-Phe within proteins.

Results

Establishing a pathway for pN-Phe synthesis in *E. coli*

To devise possible routes for pN-Phe biosynthesis, we initially attempted to use established computational pathway identification tools. State-of-the-art programs that we tested—RetroBioCat³⁰ and ATLAS of Biochemistry³¹—were unable to predict biosynthetic routes to pN-Phe, which is consistent with the limited documentation of nitroaromatic biosynthesis. Thus, we proposed our own retrobiosynthesis with the following enzymatic elements (Fig. 1b): (1) three heterologous enzymes from the chloramphenicol biosynthesis pathway in *S. venezuelae* that generate pA-Phe from chorismate; (2) the putative *N*-oxygenase ObiL and (3) a promiscuous aminotransferase capable of converting aromatic keto acids to amino acids. Before investigating this pathway, we first assessed whether *E. coli* would be a compatible host for pN-Phe synthesis given prior confirmation of nitroreductase activity³² and known toxicity from nitroaromatic degradation. Thus, we performed toxicity and stability testing in *E. coli* MG1655 with primary pathway intermediates pA-Pyr, *para*-nitrophenylpyruvate (pN-Pyr), pA-Phe and pN-Phe (Supplementary Fig. 1). Here only pN-Pyr exhibited toxicity at relevant (1 mM) concentrations, and these compounds were fairly stable as well, again with the exception of pN-Pyr. When we added 0.25 mM pN-Pyr to metabolically active cultures, we observed approximately 75% conversion to pN-Phe over 24 h (Fig. 2a). This was consistent with the reported polyspecificity of native aromatic aminotransferases measured in vitro³³.

Next, we sought to identify an *N*-oxygenase that could fully oxidize pA-Phe or pA-Pyr. Recent work has identified the putative diiron monooxygenase ObiL in the synthesis of obafluorin²⁶. In that study,

the characterization of ObiL featured a colorimetric assay to measure the oxidation of pA-Pyr rather than direct chromatographic or spectrometric confirmation. To directly assay ObiL activity on pA-Pyr using high-performance liquid chromatography (HPLC) and to evaluate pA-Phe as a potential substrate, we purified ObiL with a C-terminal hexahistidine tag and an N-terminal β-galactosidase fusion (Supplementary Fig. 2). We then performed in vitro characterization on ObiL using H₂O₂ as a reductant to recycle the diiron core of the enzyme²⁸. Here we demonstrated that ObiL was active on both pA-Pyr and pA-Phe, although only 10% conversion of 1 mM substrate was achieved in vitro after 3 h (Fig. 2b, Supplementary Fig. 3). We next investigated the heterologous activity of ObiL in MG1655 using metabolite supplementation to fermentations. In this case as well, we observed the formation of nitroaromatic products when supplementing pA-Pyr and pA-Phe (Fig. 2c). This result indicated that native reductant systems in *E. coli* can reduce the diiron cluster ObiL from its fully oxidized state. The reduced form of diiron monooxygenases is required for reaction with oxygen to form the peroxo intermediate directly implicated in amine oxidation.

De novo synthesis of pN-Phe from glucose in *E. coli*

To establish de novo biosynthesis of pN-Phe in *E. coli*, we initially sought to confirm the synthesis of pA-Phe using the previously established *papABC* operon from *S. venezuelae*. We placed the native operon on a plasmid under the control of a single isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible T7-promoter system. Given the lack of modifications to the host metabolism, this system alone served as a negative control and generated less than 10 μM pA-Phe titer following 24 h growth in M9-glucose minimal media (Fig. 3a, strain NB04; strain details found in Supplementary Table 1). To increase flux to the shikimate pathway, we further expressed a feedback-resistant 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase³⁴ (*aroG*_{D146N} or *aroG*^{*}) on a separate plasmid from *papABC* genes (strain NB06), resulting in a pA-Phe titer of 624 \pm 79 μM achieved 24 h after induction.

We next integrated the *N*-oxygenase *obiL* into the pathway in an operon with *aroG*^{*} to investigate whether coexpression of these genes with the *papABC* operon would achieve pN-Phe biosynthesis.

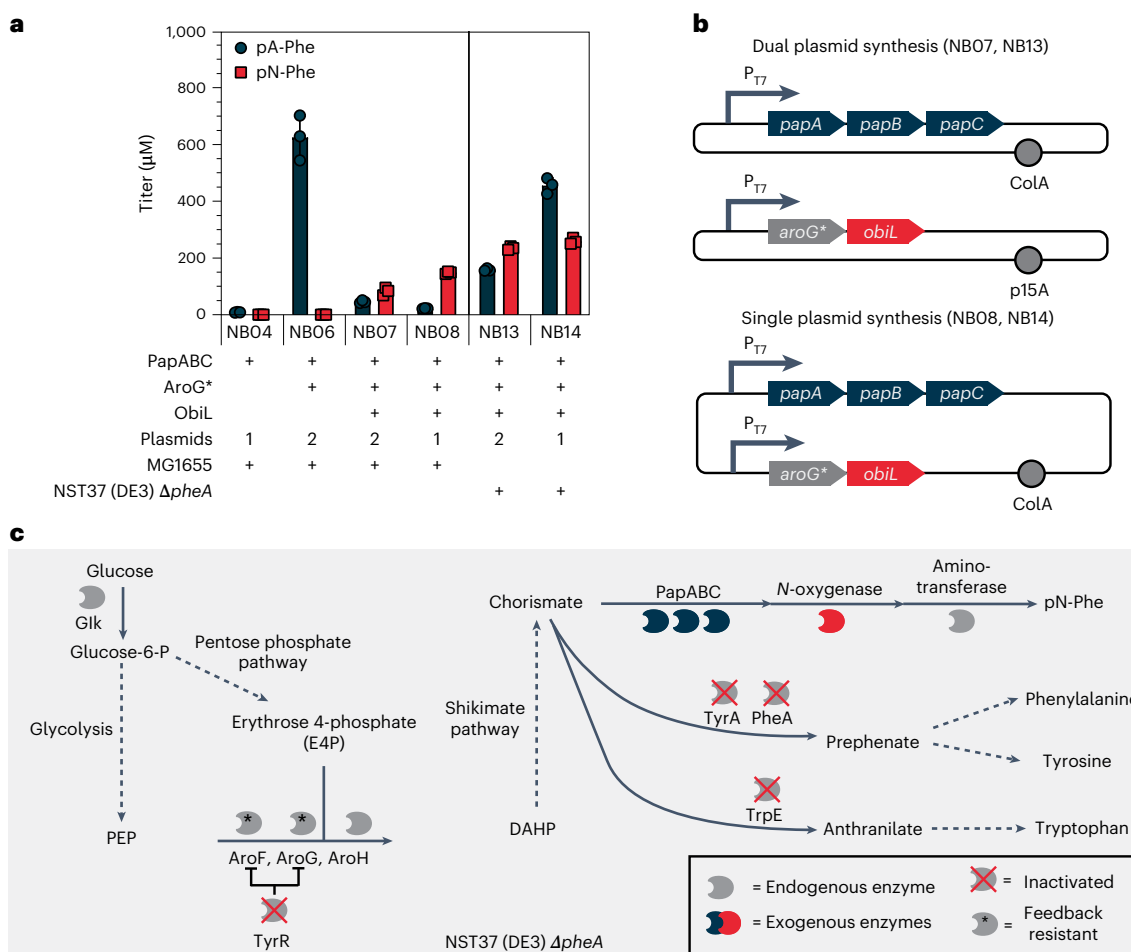


Fig. 3 | Metabolic engineering for the synthesis of pN-Phe in *E. coli*. **a**, De novo synthesis titer of pN-Phe using M9-minimal media. Samples size is $n = 3$ using biological replicates. Data shown are mean \pm s.d. **b**, Depiction of the different arrangement of pathway plasmids for the synthesis of pN-Phe. **c**, Representation of the knockouts and modifications from *E. coli* MG1655 present in the *E. coli* NST37(DE3) $\Delta pheA$ strain engineered for enhanced pN-Phe titer through chorismate. Inactivated enzymes are shown with a red cross and feedback-resistant versions of enzymes are shown with an asterisk. GIK, glucokinase; AroG, phenylalanine-sensitive 3-deoxy-d-arabino-heptulosonate-7-phosphate (DAHP) synthase; AroF, tyrosine-sensitive DAHP synthase; AroH, tryptophan-sensitive DAHP synthase; TyrA and PheA, TyrA and PheA subunits of the chorismate mutase; TrpE, anthranilate synthase component I; TyrR, transcriptional regulatory protein; PEP, phosphoenolpyruvate.

resistant versions of enzymes are shown with an asterisk. GIK, glucokinase; AroG, phenylalanine-sensitive 3-deoxy-d-arabino-heptulosonate-7-phosphate (DAHP) synthase; AroF, tyrosine-sensitive DAHP synthase; AroH, tryptophan-sensitive DAHP synthase; TyrA and PheA, TyrA and PheA subunits of the chorismate mutase; TrpE, anthranilate synthase component I; TyrR, transcriptional regulatory protein; PEP, phosphoenolpyruvate.

We were encouraged to observe pN-Phe, albeit with a fairly low 24 h titer of $83 \pm 13 \mu\text{M}$ (strain NB07, Supplementary Fig. 4). To reduce the metabolic burden and limit the number of plasmids for downstream site-specific incorporation steps, we consolidated both the *obiL-aroG** operon and the *papABC* operon onto one plasmid (Fig. 3b). This strain achieved a twofold increase in pN-Phe titer to $153 \pm 22 \mu\text{M}$ (strain NB08). Next, to further enhance titer, we purchased a commercially available phenylalanine overproducer strain and performed several modifications to improve its suitability for pN-Phe biosynthesis. The initial strain contained mutations to eliminate the functional synthesis of tyrosine and tryptophan ($\Delta tyrA$ and $\Delta trpE$), as well as mutations to increase shikimate pathway flux (inactivation of *tyrR* and mutations to create feedback-resistant versions of AroG and AroF proteins). We then inactivated the expression of PheA, the initial enzyme that directs chorismate flux to phenylalanine. This inactivation rendered the strain auxotrophic to all three standard aromatic amino acids and should permit the strain to accumulate a greater pool of chorismate, increasing potential pN-Phe titer. Additionally, we performed genomic integration of the T7 polymerase, *lacI* and *lacZ* genes for use with our T7-polymerase-based plasmid setup, creating the NST37(DE3) $\Delta pheA$ strain (Fig. 3c). Using this strain in M9-glucose media with aromatic amino acids supplemented (0.04 mg ml^{-1} L-phenylalanine, 0.04 mg ml^{-1}

L-tyrosine and 0.04 mg ml^{-1} L-tryptophan), we achieved our highest measured pN-Phe titer of $260 \pm 11 \mu\text{M}$ with the single plasmid system (strain NB14), although final cell density for the dual plasmid system was lower than that of the single plasmid system (Supplementary Table 2). Despite promising pN-Phe titer, a higher titer of pA-Phe ($454 \pm 28 \mu\text{M}$) was measured using the engineered strain with the single plasmid system. Given that pA-Phe is a structurally related nonstandard amino acid that engineered aminoacyl-tRNA synthetases (AARSs) have been shown to accept as a substrate, we reasoned that we would need to improve the selectivity of pN-Phe biosynthesis to limit eventual misacylation of orthogonal tRNA.

Screening an *N*-oxygenase library for enhanced activity

We hypothesized that we could increase pN-Phe titers by identifying a more active *N*-oxygenase to improve the conversion of pA-Phe to pN-Phe. Previously characterized nonheme diiron monooxygenases have demonstrated activity on diverse aromatic amines and compatibility with expression in *E. coli*. Thus, although ObiL was the only reported diiron monooxygenase to exhibit full oxidation of pA-Phe to pN-Phe, we hypothesized that additional *N*-oxygenases with activity on pA-Phe or pA-Pyr could exist. To assess the broader space of candidate enzymes, we generated a protein sequence similarity network (SSN).

We observed several clusters, with characterized *N*-oxygenases found in four of them, and we selected 21 distinct sequences from 18 different clusters for heterologous expression (Fig. 4a, Supplementary Fig. 5 and Table 3). Upon cloning these with a C-terminal hexahistidine tag, all the *N*-oxygenases except four (NO6, NO13, NO18 and NO20) expressed in the soluble fraction (Supplementary Fig. 6). We then assessed whether the soluble *N*-oxygenases were active on pA-Phe by supplementing it to aerobic cultures of *E. coli* MG1655 transformed to express each *N*-oxygenase. Under these conditions, only one additional *N*-oxygenase (NO16) resulted in pN-Phe production (Fig. 4b). NO16 was in the same cluster as ObiL and was the only protein to reside in a predicted bacterial gene cluster with 85% similarity to that of the β -lactam obafuorin (Supplementary Table 4). ObiL and NO16 are structurally similar, sharing 73% sequence identity. To obtain further structural insight, we generated protein structural predictions using AlphaFold2 via ColabFold^{35,36} (Supplementary Fig. 7). Alignment of AlphaFold2-generated structures of ObiL and NO16 to the crystal structure of the *N*-oxygenase AurF, when bound to its product *para*-nitrobenzoic acid (Protein Data Bank (PDB): 3cht (ref. 37)), showed binding pocket differences that could influence pA-Phe binding (Supplementary Fig. 8).

We next sought to compare the performance of ObiL and NO16 on substrates of interest *in vitro* and in the full pathway context *in vivo*. We purified NO16 and found that it catalyzed a twofold higher yield of pN-Phe from pA-Phe compared to the yield catalyzed by ObiL, converting 1 mM pA-Phe to $196 \pm 23 \mu\text{M}$ pN-Phe over 3 h (Fig. 4c). Conversion of pA-Pyr to pN-Pyr was comparable for ObiL and NO16. We then compared ObiL or NO16 in the full pathway context using the single plasmid expression system (Fig. 4d), initially using M9-minimal media supplemented with 1.5% glucose. We observed an improved pN-Phe titer at 24 h when coexpressing NO16 ($330 \pm 75 \mu\text{M}$), as compared to ObiL ($197 \pm 25 \mu\text{M}$) (Fig. 4e, Supplementary Table 5). The lower final titer of pA-Phe observed using NO16 supports the notion that increased *N*-oxygenase activity improved pathway selectivity. To explore the role of culture medium in influencing titer, we also tested a rich defined media (MOPS EZ Rich with 1.5% glucose). Using this media, we achieved an encouraging pN-Phe titer of $820 \pm 130 \mu\text{M}$ after 24 h using NO16 compared to a titer of $435 \pm 23 \mu\text{M}$ using ObiL (Fig. 4f). The use of lysogeny broth (LB) media with glucose resulted in no distinguishable difference in activity between ObiL and NO16 (Supplementary Fig. 9). To further investigate these results, we repeated *in vivo* supplementation experiments for ObiL and NO16 in LB media compared to M9-glucose (Supplementary Fig. 10). Although ObiL and NO16 appear to perform comparably in most conditions, in M9-glucose media, the use of NO16 was superior, achieving complete conversion of 0.5 mM pA-Phe to pN-Phe after 24 h. Given that the highest pN-Phe titer was generated using MOPS EZ Rich media, we used this media in the remaining studies. Given that we achieved a pN-Phe titer above 0.8 mM, which is close to the typical concentration of 1 mM supplemented to cells for genetic code expansion, we began to investigate tools for site-specific incorporation within proteins.

Identification of a selective AARS

To achieve site-specific incorporation of biosynthesized pN-Phe within a target protein, we required an AARS and tRNA pair that exhibited sufficient activity at submillimolar concentrations of pN-Phe and selectivity for pN-Phe rather than pA-Phe. The original system reported to enable pN-Phe incorporation at the nonsense amber codon UAG in *E. coli* used an evolved *MjTyrRS* (referred to as pNFRS) paired with *Mj*tRNA^{Tyr}_{CUA}³. However, the activity of this system on other nonstandard amino acids has not been reported, and many *MjTyrRS* derivatives exhibit substrate polyspecificity^{38,39}. Hence, we evaluated the activity and selectivity of a panel of 20 *MjTyrRS* derivatives that were reported to have high demonstrated activity on one or more nonstandard amino acids (Supplementary Table 6). To test site-specific incorporation of pN-Phe within a target protein, we used a Ubiquitin-superfolder green

fluorescent protein (sfGFP) fusion reporter system (Ub-GFP) that contains one in-frame UAG stop codon in the N-terminal region of the translated sfGFP domain (position 77 of the fully translated protein; position 1 of sfGFP). The anticodon on the orthogonal tRNA is designed to pair with the UAG codon on mRNA such that the nonstandard amino acid is added to the growing polypeptide at that position. In addition, to improve the signal-to-noise ratio, we performed screening in an aromatic amino acid dropout variation of MOPS EZ Rich medium (Fig. 5a). In this initial screen, we identified the following five AARSs that achieved over fivefold enhancement in OD₆₀₀-normalized sfGFP fluorescence after supplementation of pN-Phe: (1) TetRS-C11 (ref. 40), (2) pCNFRS³⁸, (3) NapARS⁴¹, (4) pAcFRS⁴² and (5) pNFRS (Fig. 5b, Supplementary Fig. 11). In addition, we tested variants of the orthogonal pyrrolysyl and chimeric phenylalanine amber suppression systems, but for both categories, incorporation efficiency was poor (Supplementary Fig. 12). We then investigated the dose-response to pN-Phe at concentrations from 200 to 1,000 μM (Fig. 5c) and the selectivity for pN-Phe compared to pA-Phe or Tyr (Fig. 5d) for these five AARSs. We observed that the originally reported pNFRS accepted pA-Phe and exhibited poor activity at lower pN-Phe concentrations as compared to the AARSs TetRS-C11 and NapARS. Of the two, TetRS-C11 exhibited the most efficient incorporation of pN-Phe within proteins at lower concentrations and had low fold-change enhancement at relevant pA-Phe concentrations. We thus chose to investigate whether we could engineer this AARS for improved selectivity and titer (Supplementary Figs. 13 and 14). We created a 44-member library consisting of single mutants at binding pocket residues to positively charged amino acids (R/H/K) with the intention of favoring the electronegative nitro group. However, as no members of the library exhibited improved selectivity or activity, we continued with TetRS-C11.

To test the coupling of nonstandard amino acid biosynthesis and incorporation in stages, we first investigated whether we could couple an *N*-oxygenase with the orthogonal translation system and reporter. We coexpressed the TetRS-C11 AARS/*Mj*tRNA^{Tyr}_{CUA} pair along with ObiL and a C-terminal heptahistidine-tagged version of the Ub-GFP protein previously described, and we supplemented these cultures with 1 mM pA-Phe (Fig. 5e). Following fermentation in LB media, we purified the GFP reporter and measured its mass via intact protein mass spectrometry (MS), where we saw two major masses, one at 37274.5 Da and one at 37305.5 Da (Fig. 5f). The latter mass was the expected value with pN-Phe incorporation, and the former corresponded to either pA-Phe or Tyr present from the media (the masses of pA-Phe and Tyr are within 1 Da). The experiment successfully demonstrated that pN-Phe produced by cells could be incorporated within our target protein. The level of incorrect protein product observed may have been due to the high concentration of pA-Phe (1 mM) added during this assay or the high Tyr levels present in LB media.

Coupled *de novo* biosynthesis and incorporation of pN-Phe

Our last steps to create an autonomous producer of nitrated proteins explored the coupling of *de novo* biosynthesis and incorporation of pN-Phe. To start, we expressed our pN-Phe biosynthesis plasmid with the AARS/tRNA plasmid and the Ub-GFP fusion reporter plasmid. We staggered inductions by inducing the pN-Phe biosynthesis at mid-exponential phase (OD ~0.5) and inducing the AARS and Ub-GFP 2 h later to allow accumulation of pN-Phe before induction of the AARS. After purification of our reporter and analysis by intact protein MS, we obtained a mass of ~63 Da less than predicted, corresponding to the incorporation of either glutamine or lysine at the desired site (Fig. 6a). Although the listed phenotype for the NST37 strain does not note any suppressor tRNA mutations, the *supE44* allele, which encodes a glutaminyl suppressor tRNA, *glnX* has been found in many strains of *E. coli* K12. In our case, sequencing of the *glnX* tRNA indeed uncovered a suppressor CUA anticodon. Thus, we modified the anticodon to the wild-type CUG (Fig. 6b).

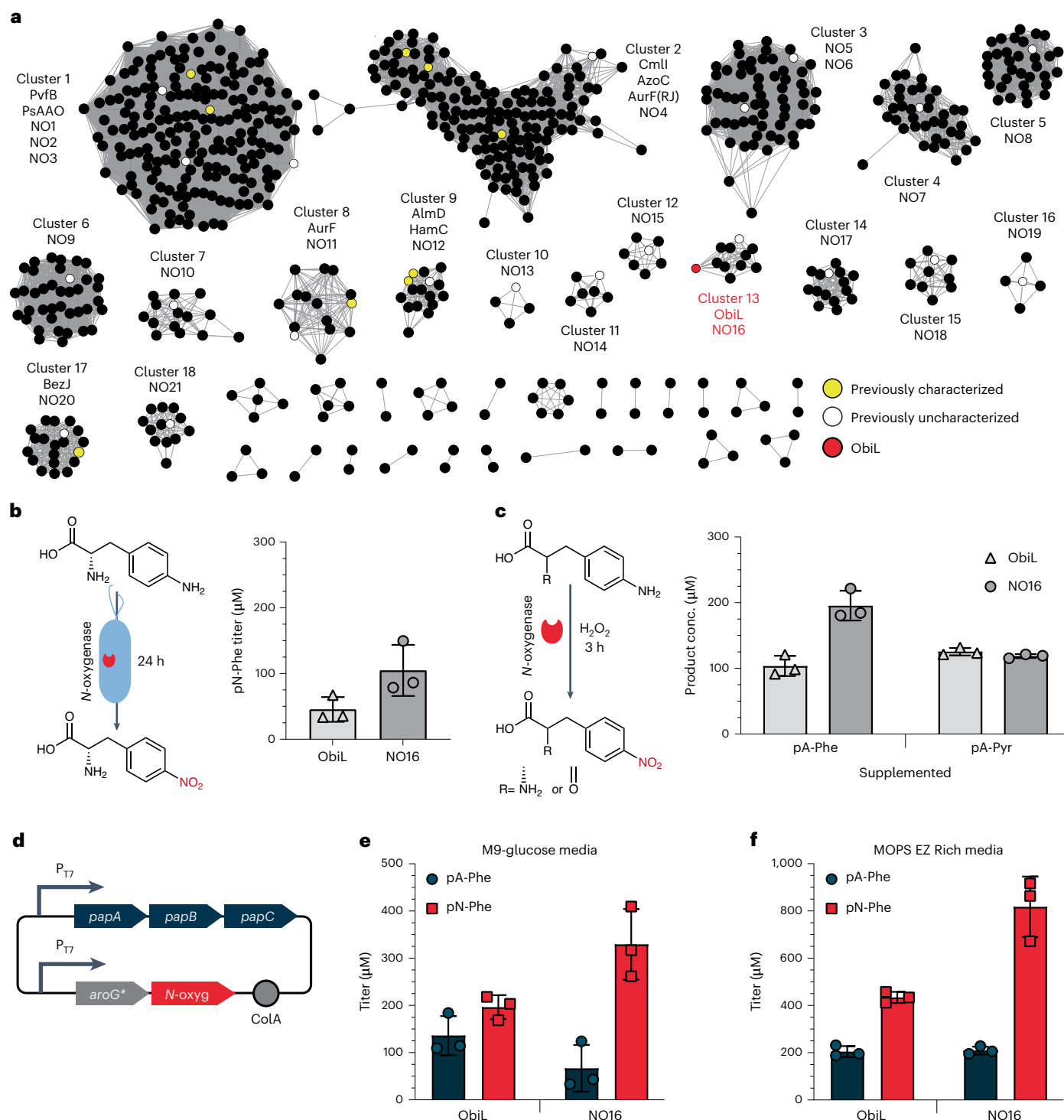


Fig. 4 | Bioprospecting of nonheme diiron monooxygenases for activity on pA-Pyr and pA-Phe. a, Sequence similarity network generated using 2,134 unique putative diiron monooxygenase sequences determined from NCBI BLAST of the *N*-oxygenases AurF, CmlI, AzoC and ObiL represented as 775 unique nodes. Edges are drawn between nodes with a minimum alignment score of 100. Unclustered nodes (56 total nodes representing 60 sequences) are not displayed. Previously characterized *N*-oxygenases are highlighted in yellow (PvfB, PsAAO, CmlI, AzoC, AurF(RJ), AurF, AlmD, HamC and BezJ). Previously uncharacterized *N*-oxygenases

cloned and tested in this study are highlighted in white. ObiL is shown in red. All other *N*-oxygenases are shown in black. **b**, In vivo screening of *N*-oxygenase activity on pA-Phe via supplementation testing of 1 mM pA-Phe in cultures expressing *N*-oxygenases via measurement of pN-Phe titer after 24 h fermentation in LB media at 37 °C. **c**, In vitro investigation of NO16 activity on pA-Phe and pA-Pyr. **d**, Depiction of the pN-Phe synthesis plasmid setup. **e**, De novo pN-Phe synthesis in M9-glucose media. **f**, Synthesis of pN-Phe in MOPS EZ Rich media. Samples sizes are $n = 3$ using biological replicates. Data shown are mean \pm s.d.

Next, we evaluated how pN-Phe titer would be affected by the burden of expressing biosynthetic genes, incorporation machinery and Ub-GFP reporter using separate plasmids. Here we measured titer at 3

and 24 h from fermentative growth using MOPS EZ Rich glycerol media to enable the induction of the P_{araBAD} promoter for AARS translation (Fig. 6c). We obtained low pN-Phe titers at both 3 and 24 h of

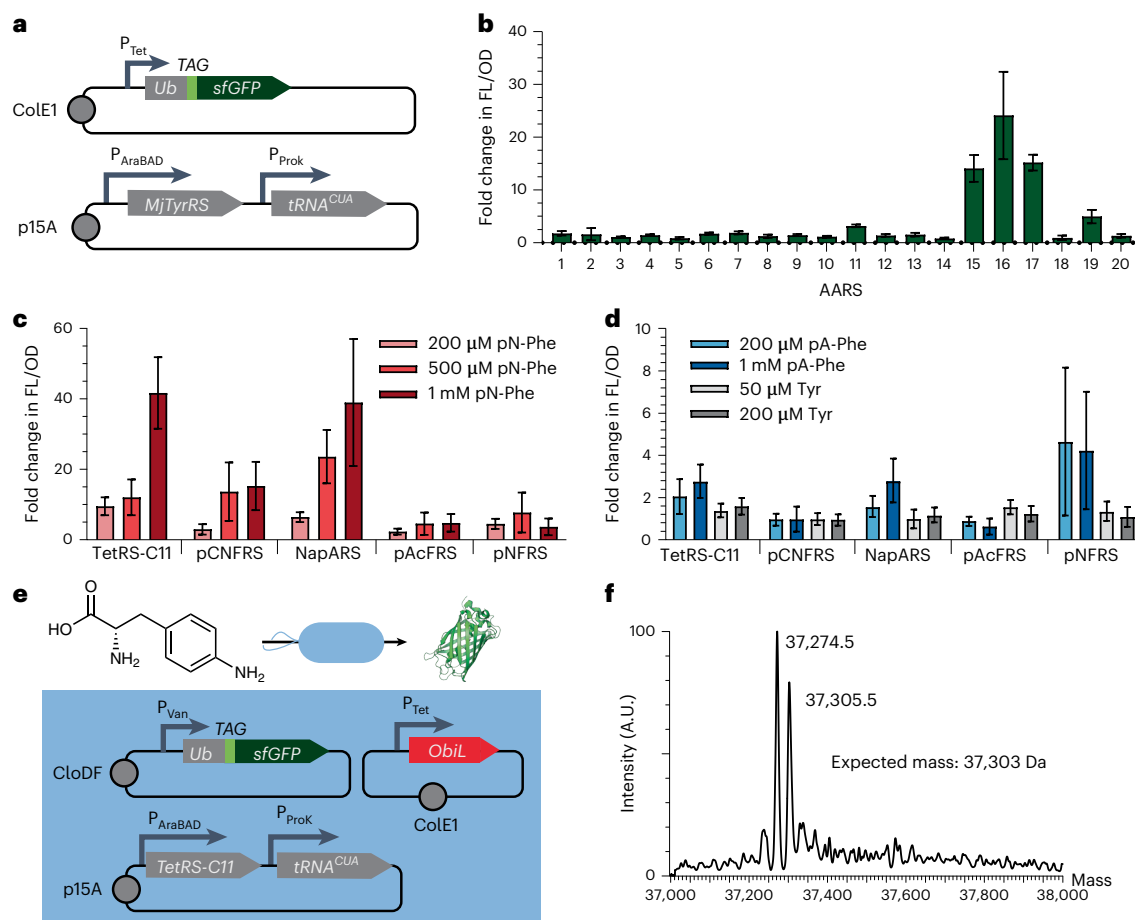


Fig. 5 | Aminoacyl-tRNA synthetase screening for incorporation of pN-Phe.

a, A ubiquitin-fused sfGFP reporter with a TAG codon at the N-terminus of the sfGFP was expressed on a pZE vector with a pEVOL plasmid expressing an aminoacyl *M. jannaschii* tyrosyl-tRNA synthetase and corresponding amber suppressor tRNA. **b**, A library of literature-derived AARS variants was screened for pN-Phe incorporation by comparison of the fold change in fluorescence (ex, 485 nm; em, 510 nm) normalized by OD₆₀₀ between cultures supplemented with 1 mM pN-Phe and cultures with no nonstandard amino acid supplemented.

c,d, Top-performing AARSs were screened for activity at lower pN-Phe concentration (**c**) and off-target incorporation of Tyr or pA-Phe (**d**). **e**, Depiction of the strain plasmid setup used for biosynthesis of pN-Phe from pA-Phe followed by subsequent incorporation in *E. coli* MG1655. **f**, Mass spectra from intact protein MS for Ub-GFP synthesized with pA-Phe supplementation and *N*-oxygenase *ObiL* expression. Samples size is $n = 3$ using biological replicates, and the data shown are mean \pm s.d. with the exception of mass spectra.

$10.9 \pm 0.9 \mu\text{M}$ and $50 \pm 7 \mu\text{M}$, respectively. We sought to improve pN-Phe synthesis in this integrated system by limiting the burden from the additional plasmid. Thus, we created a construct (pRepAARS) that features the AARS and Ub-GFP reporter within an anhydrotetracycline-inducible operon and that includes separate constitutive expression of *Mj*tRNA^{Tyr-_{CUA}}. The top-three performing AARSs in the pRepAARS vector context exhibited comparable fold-change to the system with a separate reporter and AARS/tRNA plasmid. In addition, for pRepAARS, we observed an OD₆₀₀-normalized fold-change in GFP fluorescence of approximately fourfold over a background at pN-Phe concentrations as low as 50 μM (Fig. 6d).

We then evaluated pN-Phe synthesis utilizing pRepTetRS-C11 and the pN-Phe synthesis plasmid in NST37(DE3) Δ *pheA glnX_{CUG}* via fermentation with MOPSEZ Rich glucose media. We induced the pN-Phe synthesis pathway at midexponential phase (OD - 0.5) and induced pRepTetRS-C11 2 h later. Using this protocol, we observed extracellular pN-Phe titers at $28 \pm 4 \mu\text{M}$ after 4 h and $360 \pm 60 \mu\text{M}$ after 24 h (Fig. 6e). These titers were comparable to prior titers despite the burden posed by induction of gene expression for the additional genes. Encouragingly, we observed distinct green fluorescence under these conditions as well. With this result, we investigated if we could further improve titers earlier by inducing pN-Phe synthesis at inoculation. However,

induction at inoculation resulted in a near elimination of pN-Phe production (Supplementary Fig. 15).

Given the promising titers and fluorescence observed when we induced the pN-Phe biosynthesis pathway during mid-exponential phase, we repeated this experiment to determine the fidelity of nitrated reporter protein production. We purified the Ub-GFP reporter and obtained a protein titer of 0.53 mg l^{-1} . To confirm the incorporation of pN-Phe, we performed intact protein MS, which encouragingly revealed a mass peak at the expected mass (Fig. 6f, Supplementary Fig. 16). Additional mass peaks suggested incorporation of pA-Phe/Tyr into some of the proteins. To investigate this, we performed additional fermentations for the production of pN-Phe-containing protein (achieving $3.5 \pm 0.8 \text{ mg l}^{-1}$ purified protein) for tryptic digestion liquid chromatography (LC)-MS/MS. After we ran the protein sample on an SDS-PAGE gel, we excised the band corresponding to the mass of the intact protein of interest for in-gel tryptic digestion (Supplementary Fig. 17). We submitted the digested protein to LC-MS/MS, confirming the presence of nitrated protein product as the plurality product (47% as measured by extracted ion chromatogram absorbance) in addition to byproducts that resulted from misincorporation of pA-Phe (36%) and tyrosine (17%) at the target site (Supplementary Fig. 18). Despite some misincorporation observed, this result demonstrated

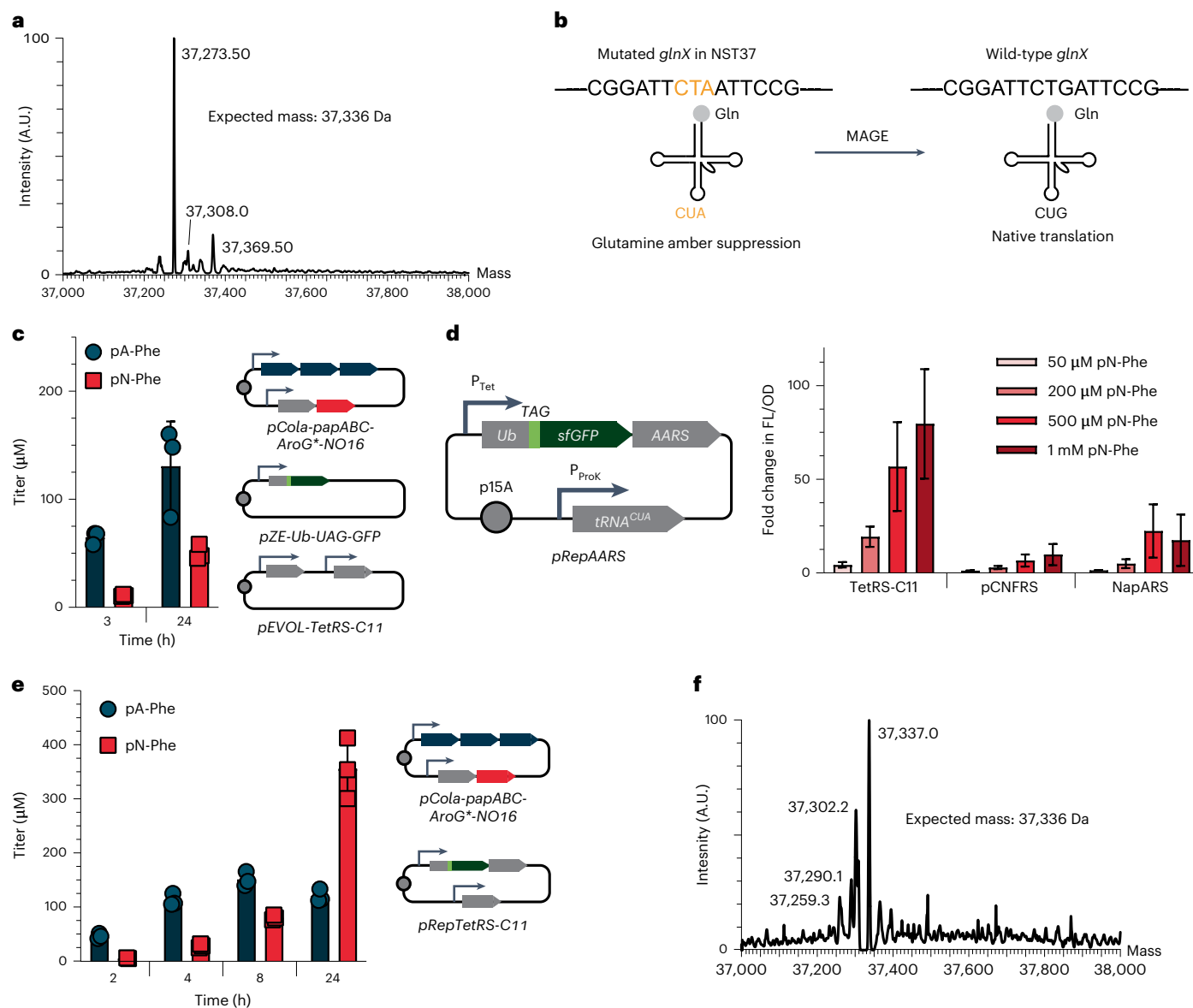


Fig. 6 | Integration of the pN-Phe biosynthesis pathway with orthogonal *MjTyrRS/tRNA* for de novo biosynthesis and incorporation. a, Mass spectra from intact protein MS of Ub-GFP purified from NST37(DE3) Δ *pheA* expressing three plasmids with a primary mass peak corresponding to glutaminylation incorporation. The higher predicted mass prediction for the Ub-GFP construct relative to Fig. 5f is due to an L213F mutation between plasmids in this figure. **b**, Sequencing revealed a mutation of the *glnX* gene leading to glutaminylation amber suppression addressed via MAGE. **c**, Titer of pN-Phe and pA-Phe in NST37(DE3) Δ *pheA* expressing a separate reporter (pZE-Ub-UAG-GFP), AARS/tRNA (pEVOL-TetRS-C11) and pN-Phe synthesis plasmid (pCola-papABC-AroG*-NO16) using MOPSEZ Rich media with glycerol. **d**, Integration of Ub-GFP reporter in an operon with AARS and screening for the fold-change in normalized GFP fluorescence at a range of pN-Phe concentrations. **e**, Titer of pN-Phe and pA-Phe in NST37(DE3) Δ *pheA* expressing pN-Phe synthesis plasmid (pCola-papABC-AroG*-NO16) and AARS/tRNA/reporter plasmid (pRepTetRS-C11) using MOPSEZ Rich media with glucose. **f**, Mass spectra from intact protein MS of Ub-GFP with pN-Phe incorporated from the metabolic synthesis in *E. coli* NST37(DE3) Δ *pheA glnX_{CUG}*. Samples size is $n = 3$ using biological replicates, and the data shown are mean \pm s.d. with the exception of mass spectra.

that our strain is capable of de novo pN-Phe metabolic synthesis and its site-specific incorporation within a target protein.

Discussion

There are few examples of microbes engineered to create and harness nonstandard amino acids that contain chemical functionalities absent from the standard amino acids for translation into proteins. Metabolic synthesis of functional groups that are uncommon to biology can be difficult due to a limited repertoire of known enzymes that perform these chemistries. To achieve biosynthesis of pN-Phe, we integrated a nonheme diiron monooxygenase-type *N*-oxygenase from the pathway

for obafluorin synthesis to perform nitro group formation. Although the in vivo activity of certain *N*-oxygenases has been demonstrated in *E. coli*²⁹, this class of enzymes had not previously been used in a de novo biosynthetic pathway. In this work, we also identified an *N*-oxygenase with improved activity for pN-Phe synthesis through bioprospecting. We found that nonheme diiron monooxygenase-type *N*-oxygenases generally express well in *E. coli*. Given this result, it is reasonable to wonder whether the *E. coli* genome encodes proteins with similar structure or functionality. Currently, no amine-oxidizing nonheme diiron monooxygenases have been identified in the *E. coli* genome, and a BLAST search against

the genome for ObiL does not reveal highly similar proteins. However, other enzymes in the nonheme diiron class have been found and extensively studied in *E. coli*. One example is the ribonucleotide reductase R2 (RNR-R2) that converts ribonucleotides to deoxyribonucleotides using iron-catalyzed O₂ activation similar to ObiL⁴³. As we do not see pA-Phe/pA-Pyr oxidation in the absence of an exogenous *N*-oxygenase, and because RNR-R2 does not bear much sequence similarity to ObiL, it is unlikely that RNR-R2 performs amine oxidation. However, the cytoplasmic environment is known to be capable of reduction of the diiron cores for the activity of enzymes like RNR-R2, which could lend to the robust in vivo activity of the two *N*-oxygenases observed in this study. Generally, broader screening of the substrate scope of *N*-oxygenases in *E. coli* is warranted for use in pathways toward more diverse nitrated products.

Our work highlights the opportunities afforded by genetic engineering to create and harness xenobiological building blocks. Heterologous expression of three genetic modules—biosynthetic genes, an orthogonal AARS/tRNA pair and a target protein for site-specific incorporation of the nonstandard amino acid—at appropriate levels were critical to form nitrated protein. Our work also illustrated the importance of metabolic engineering performance metrics in a new context, especially given the low observable pN-Phe titers during the coexpression of all modules and the accumulation of the nonstandard amino acid pA-Phe. A threshold nonstandard amino acid titer must be met to achieve effective acylation of orthogonal tRNA. Yet, titers may not need to be substantially higher for efficient target protein synthesis. Our work demonstrated that the rate of nonstandard amino acid biosynthesis is a critical variable because the efficient translation of the target protein requires reaching a threshold titer while in favorable regimes for translation, such as the exponential growth phase. In scenarios where biosynthetic pathways toward a desired nsAA feature another nsAA as an intermediate, and if orthogonal translation machinery is polyspecific, then selectivity is an additional critical variable. Here selectivity was vital to improve fidelity during the downstream operations of tRNA acylation and protein translation. We improved the selectivity of pN-Phe incorporation through a combination of altering plasmid configurations, screening *N*-oxygenase orthologs, varying the timing of induction, identifying a highly selective AARS and removing an undesired genomic amber suppressor tRNA. As more complex nonstandard amino acid targets continue to be pursued, we anticipate that strategies pursued in this work will be required to produce organisms that autonomously harbor expanded genetic codes.

Our work is the first proof-of-concept demonstrating that nitrated proteins could be produced using a model bacterium without suppression of nitrated building blocks. Currently, we can achieve low-yield production of proteins nitrated at a desired site, with moderate levels of impurities that contain Tyr or pA-Phe misincorporation at that site. With further optimization, the technology should allow for the production of nitrated proteins in distributed settings with greater yields and purity. An exciting potential future application of the technology, after considerable improvement, could be the development of a live bacterial delivery vehicle that produces nitrated antigens in the body. However, the autonomous incorporation of biosynthesized nitrated amino acids within proteins is currently limited by low levels of pN-Phe titer at early stages of growth, leading to pA-Phe and Tyr misincorporation. In a preclinical or therapeutic setting, the production of these off-target protein products could be acceptable, provided that sufficient nitrated antigens were synthesized. However, higher titers of nitrated antigens would likely be required. Here genetic modifications to further improve pN-Phe intracellular retention and titer at early growth phases or directed evolution for a *Mj*TyrRS variant with higher affinity for pN-Phe could circumvent these issues. The eventual live vector would also need to harbor genetic cassettes on the genome without antibiotic resistance markers, contain alterations in promoter systems for induction within preclinical settings and achieve

nitrated protein synthesis in targeted regions within animal models. Our proof-of-concept serves as an enabling technology for these future lines of inquiry.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-023-01338-x>.

References

1. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126**, 14306–14307 (2004).
2. Neumann, H., Hazen, J. L., Weinstein, J., Mehl, R. A. & Chin, J. W. Genetically encoding protein oxidative damage. *J. Am. Chem. Soc.* **130**, 4028–4033 (2008).
3. Tsao, M. L., Summerer, D., Ryu, Y. & Schultz, P. G. The genetic incorporation of a distance probe into proteins in *Escherichia coli*. *J. Am. Chem. Soc.* **128**, 4572–4573 (2006).
4. Jackson, J. C., Duffy, S. P., Hess, K. R. & Mehl, R. A. Improving nature's enzyme active site with genetically encoded unnatural amino acids. *J. Am. Chem. Soc.* **128**, 11124–11127 (2006).
5. Grünewald, J. et al. Immunochemical termination of self-tolerance. *Proc. Natl Acad. Sci. USA* **105**, 11276–11280 (2008).
6. Grünewald, J. et al. Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids. *Proc. Natl Acad. Sci. USA* **106**, 4337–4342 (2009).
7. Gauba, V. et al. Loss of CD4 T-cell-dependent tolerance to proteins with modified amino acids. *Proc. Natl Acad. Sci. USA* **108**, 12821–12826 (2011).
8. Shorter, J. *PATAI'S Chemistry of Functional Groups* (John Wiley & Sons, 2009). <https://doi.org/10.1002/9780470682531.pat0081>
9. Tian, H. et al. Nitrated T helper cell epitopes enhance the immunogenicity of HER2 vaccine and induce anti-tumor immunity. *Cancer Lett.* **430**, 79–87 (2018).
10. Tian, H. et al. PDL1-targeted vaccine exhibits potent antitumor activity by simultaneously blocking PD1/PDL1 pathway and activating PDL1-specific immune responses. *Cancer Lett.* **476**, 170–182 (2020).
11. Li, F. et al. A new vaccine targeting RANKL, prepared by incorporation of an unnatural amino acid into RANKL, prevents OVX-induced bone loss in mice. *Biochem. Biophys. Res. Commun.* **499**, 648–654 (2018).
12. Völler, J. S. & Budisa, N. Coupling genetic code expansion and metabolic engineering for synthetic cells. *Curr. Opin. Biotechnol.* **48**, 1–7 (2017).
13. Dickey, R. M., Forti, A. M. & Kunjapur, A. M. Advances in engineering microbial biosynthesis of aromatic compounds and related compounds. *Bioresour. Bioprocess.* **8**, 91 (2021).
14. Ding, C., Ma, J., Dong, Q. & Liu, Q. Live bacterial vaccine vector and delivery strategies of heterologous antigen: a review. *Immunol. Lett.* **197**, 70–77 (2018).
15. van Bloois, E., Winter, R. T., Kolmar, H. & Fraaije, M. W. Decorating microbes: surface display of proteins on *Escherichia coli*. *Trends Biotechnol.* **29**, 79–86 (2011).
16. Wang, S. et al. Salmonella vaccine vectors displaying delayed antigen synthesis in vivo to enhance immunogenicity. *Infect. Immun.* **78**, 3969–3980 (2010).
17. Mehl, R. A. et al. Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**, 935–939 (2003).
18. Chen, Y. et al. Unleashing the potential of noncanonical amino acid biosynthesis to create cells with precision tyrosine sulfation. *Nat. Commun.* **13**, 5434 (2022).

19. Kim, S., Sung, B. H., Kim, S. C. & Lee, H. S. Genetic incorporation of L-dihydroxyphenylalanine (DOPA) biosynthesized by a tyrosine phenol-lyase. *Chem. Commun.* **54**, 3002–3005 (2018).
20. Zuo, R. & Ding, Y. Direct aromatic nitration system for synthesis of nitrotryptophans in *Escherichia coli*. *ACS Synth. Biol.* **8**, 857–865 (2019).
21. Chen, Y. et al. Creation of bacterial cells with 5-hydroxytryptophan as a 21st amino acid building block. *Chem* **6**, 2717–2727 (2020).
22. Marchand, J. A. et al. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019).
23. Waldman, A. J., Ng, T. L., Wang, P. & Balskus, E. P. Heteroatom–heteroatom bond formation in natural product biosynthesis. *Chem. Rev.* **117**, 5784–5863 (2017).
24. Masuo, S., Zhou, S., Kaneko, T. & Takaya, N. Bacterial fermentation platform for producing artificial aromatic amines. *Sci. Rep.* **6**, 25764 (2016).
25. Chen, Y. et al. A noncanonical amino acid-based relay system for site-specific protein labeling. *Chem. Commun.* **54**, 7187–7190 (2018).
26. Schaffer, J. E., Reck, M. R., Prasad, N. K. & Wencewicz, T. A. β -Lactone formation during product release from a nonribosomal peptide synthetase. *Nat. Chem. Biol.* **13**, 737–744 (2017).
27. Scott, T. A., Heine, D., Qin, Z. & Wilkinson, B. An L-threonine transaldolase is required for L-threo- β -hydroxy- α -amino acid assembly during obafuorin biosynthesis. *Nat. Commun.* **8**, 15935 (2017).
28. Chanco, E., Choi, Y. S., Sun, N., Vu, M. & Zhao, H. Characterization of the N-oxygenase AurF from *Streptomyces thioletus*. *Bioorg. Med. Chem.* **22**, 5569–5577 (2014).
29. Lu, H., Chanco, E. & Zhao, H. CmlI is an N-oxygenase in the biosynthesis of chloramphenicol. *Tetrahedron* **68**, 7651–7654 (2012).
30. Finnigan, W., Hepworth, L. J., Flitsch, S. L. & Turner, N. J. RetroBioCat as a computer-aided synthesis planning tool for biocatalytic reactions and cascades. *Nat. Catal.* **4**, 98–104 (2021).
31. Hadadi, N., Hafner, J., Shajkofci, A., Zisaki, A. & Hatzimanikatis, V. ATLAS of biochemistry: a repository of all possible biochemical reactions for synthetic biology and metabolic engineering studies. *ACS Synth. Biol.* **5**, 1155–1166 (2016).
32. Rau, J. & Stolz, A. Oxygen-insensitive nitroreductases NfsA and NfsB of *Escherichia coli* function under anaerobic conditions as lawsone-dependent azo reductases. *Appl. Environ. Microbiol.* **69**, 3448–3455 (2003).
33. Onuffer, J. J., Ton, B. T., Klement, I. & Kirsch, J. F. The use of natural and unnatural amino acid substrates to define the substrate specificity differences of *Escherichia coli* aspartate and tyrosine aminotransferases. *Protein Sci.* **4**, 1743–1749 (1995).
34. Kunjapur, A. M., Tarasova, Y. & Prather, K. L. J. Synthesis and accumulation of aromatic aldehydes in an engineered strain of *Escherichia coli*. *J. Am. Chem. Soc.* **136**, 11644–11654 (2014).
35. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
36. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
37. Seong Choi, Y., Zhang, H., Brunzelle, J. S., Nair, S. K. & Zhao, H. *In vitro* reconstitution and crystal structure of p-aminobenzoate N-oxygenase (AurF) involved in aureothin biosynthesis. *Proc. Natl Acad. Sci. USA* **105**, 6858–6863 (2008).
38. Young, D. D. et al. An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894–1900 (2011).
39. Kunjapur, A. M. et al. Engineering posttranslational proofreading to discriminate nonstandard amino acids. *Proc. Natl Acad. Sci. USA* **115**, 619–624 (2018).
40. Blizzard, R. J. et al. Ideal bioorthogonal reactions using a site-specifically encoded tetrazine amino acid. *J. Am. Chem. Soc.* **137**, 10044–10047 (2015).
41. Wang, L., Brock, A. & Schultz, P. G. Adding L-3-(2-naphthyl)alanine to the genetic code of *E. coli*. *J. Am. Chem. Soc.* **124**, 1836–1837 (2002).
42. Wang, L., Zhang, Z., Brock, A. & Schultz, P. G. Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **100**, 56–61 (2003).
43. Wörsdörfer, B. et al. Function of the diiron cluster of *Escherichia coli* class Ia ribonucleotide reductase in proton-coupled electron transfer. *J. Am. Chem. Soc.* **135**, 8585–8593 (2013).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

Methods

Strains and plasmids

E. coli strains and plasmids used in this study are listed in Supplementary Table 1. Molecular cloning and vector propagation were performed in *E. coli* DH5 α . PCR-based DNA replication was performed using KOD XTREME Hot Start Polymerase. Cloning was performed using Gibson Assembly with constructs and oligos for PCR amplification shown in Supplementary Table 7. Genes were purchased as gBlocks or gene fragments from Integrated DNA Technologies (IDT) or Twist Bioscience and were optimized for *E. coli* K12 using the IDT Codon Optimization Tool with sequences shown in Supplementary Tables 6 and 8. The plasmids pOSIP-TH (Addgene, 45978) and pE-FLP (Addgene, 45978) were gifts from D. Endy and K.E. Shearwin⁴⁴. The *papABC* operon was kindly provided by R. Mehl of Oregon State University in plasmid pLASC-lppPW. The pORTMAGE-EC1 recombineering plasmid was kindly provided by T. Wannier⁴⁵.

The chorismate overproducer strain was derived from a commercially available phenylalanine overproducer *E. coli* strain (NST37; ATCC, 31882). To enable the compatibility of this strain with T7-promoter systems, the 4521 bp region of the phage T7 genome that is responsible for T7 polymerase functionality (*lacI*, *lacZ* and T7 RNA polymerase; Supplementary Table 9) was genomically integrated using one-step clonete-gration with the pOSIP-TH plasmid. Following plasmid assembly, NST37 was transformed with the clonete-gration plasmid, and the integration of the region for T7 polymerase activity (DE3) was confirmed via the selection on LB-agar plates containing 9 $\mu\text{g ml}^{-1}$ tetracycline. Following Sanger sequencing-based confirmation of genetic incorporation, genomic tetracycline resistance was removed using pE-FLP. Then, using multiplexed automated genome engineering (MAGE) with the m-toluic acid inducible pORTMAGE-EC1 recombineering plasmid, a TAA and a TGA stop codon were introduced into the genomic sequence for the chorismate mutase/prephenate dehydratase PheA at positions 10 and 12 to serve as a translational knockout. Curing of the pORTMAGE-EC1 plasmid following Sanger sequencing confirmation of genomic knockouts produced the chorismate overproducer strain (NST37(DE3) *ΔpheA*). Further rounds of MAGE were performed downstream to create NST37(DE3) *ΔpheA*glnX_{CUC} to eliminate glutaminyl amber suppression in the strain.

Chemicals

The following compounds were purchased from MilliporeSigma: vanillic acid, hydrogen peroxide, kanamycin sulfate, chloramphenicol, carbenicillin disodium, dimethyl sulfoxide (DMSO), potassium phosphate dibasic, potassium phosphate monobasic, magnesium sulfate, calcium chloride dihydrate, imidazole, glycerol, M9 salts, sodium dodecyl sulfate, lithium hydroxide, boric acid, Tris base, glycine, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), trifluoroacetic acid (TFA) and KOD XTREME Hot Start and KOD Hot Start polymerases. pN-Phe, m-toluic acid and D-glucose were purchased from TCI America. pA-Phe, methanol, agarose, Laemmli SDS sample-reducing buffer and ethanol were purchased from Alfa Aesar. pA-Pyr and pN-Pyr were purchased from abcr GmbH. Anhydrotetracycline and IPTG were purchased from Cayman Chemical. Acetonitrile, sodium chloride, LB Broth powder (Lennox) and LB-Agar powder (Lennox) were purchased from Fisher Chemical. L-Arabinose was purchased from VWR. A MOPS EZ Rich defined medium kit and components were purchased from Teknova. Trace Elements A was purchased from Corning. Taq DNA ligase was purchased from GoldBio. Phusion DNA polymerase and T5 exonuclease were purchased from New England BioLabs. Sybr Safe DNA gel stain and BenchMark His-tagged Protein Standard were purchased from Invitrogen. Horseradish peroxidase (HRP) conjugated 6*His His-Tag Mouse McAB was obtained from Proteintech.

Culture conditions

Cultures for general culturing and *N*-oxygenase protein overexpression were grown in LB-Lennox medium (LB: 10 g l⁻¹ bacto tryptone, 5 g l⁻¹

sodium chloride and 5 g l⁻¹ yeast extract). Cultures to demonstrate de novo pN-Phe synthesis were grown in either LB-Lennox-glucose medium (LB with 1.5% glucose (wt/vol)), M9-glucose minimal media⁴⁶ with Corning Trace Elements A (1.60 $\mu\text{g ml}^{-1}$ CuSO₄ · 5H₂O, 863.00 $\mu\text{g ml}^{-1}$ ZnSO₄ · 7H₂O, 17.30 $\mu\text{g ml}^{-1}$ Selenite · 2Na and 1155.10 $\mu\text{g ml}^{-1}$ ferric citrate) and 1.5% glucose (wt/vol) or MOPS EZ Rich defined media (Teknova M2105) with 1.5% glucose (wt/vol). For cultures of NST37(DE3) *ΔpheA* strains that were grown in M9-glucose minimal media, 0.04 mg ml⁻¹ L-phenylalanine, 0.04 mg ml⁻¹ L-tyrosine and 0.04 mg ml⁻¹ L-tryptophan were added to the media to ensure growth.

For stability testing, a culture of *E. coli* K12 MG1655 (DE3) was inoculated from frozen stock and grown to confluence overnight in 5 ml of LB media. Confluent overnight cultures were then used to inoculate experimental cultures in 300 μl volumes in a 96-deep-well plate (Thermo Fisher Scientific, 260251) at 100 \times dilution. Cultures were supplemented with 0.5 mM of heterologous metabolites (pA-Phe, pA-Pyr and pN-Phe), except in the case of pN-Pyr where 0.25 mM with an additional 15 μl of DMSO (-5% final concentration) were supplemented due to solubility concerns. Cultures were incubated at 37 °C with shaking at 1,000 r.p.m. and an orbital radius of 3 mm. Compounds were quantified from the extracellular broth over a 24 h period using HPLC.

For toxicity testing, cultures were similarly prepared with confluent overnight cultures of MG1655 (DE3) used to inoculate experimental cultures at 100 \times dilution in 200 μl volumes in a Greiner clear bottom 96-well plate (Greiner, 655090) in LB media. Cultures were supplemented with 1 mM of heterologous metabolite and 5% DMSO for metabolite solubility and grown for 24 h in a Spectramax i3x plate reader with medium plate shaking at 37 °C with absorbance readings at 600 nm taken every 5 min to calculate doubling time and growth rate.

For supplementation testing, strains transformed with plasmids expressing pathway genes were prepared with inoculation of 300 μl volumes in a 96-deep-well plate with appropriate antibiotic added to maintain plasmids (34 $\mu\text{g ml}^{-1}$ chloramphenicol, 50 $\mu\text{g ml}^{-1}$ kanamycin, 50 $\mu\text{g ml}^{-1}$ carbenicillin or 95 $\mu\text{g ml}^{-1}$ streptomycin). Cultures were incubated at 37 °C with shaking at 1,000 r.p.m. and an orbital radius of 3 mm until an OD₆₀₀ of 0.5–0.8 was reached. OD₆₀₀ was measured using a Thermo Fisher Scientific BioMate 160 UV-Vis Spectrophotometer. At this point, the pathway plasmids were fully induced with the addition of a corresponding inducer (1 mM IPTG, 1 mM vanillate or 0.2 nM anhydrotetracycline), and the metabolite of interest was supplemented at this time. Cultures were incubated over 24 h at 37 °C with sampling, and metabolite concentration was measured via supernatant sampling and submission to HPLC.

For pN-Phe biosynthesis testing, cultures were inoculated with overnight culture grown in LB-glucose media. Cultures were inoculated at 100 \times dilution from confluent overnight culture in 50 ml of the corresponding media with appropriate antibiotics in 250 ml baffled shake flasks and grown at 37 °C at 250 r.p.m. Expression vectors were fully induced at OD₆₀₀ 0.5–0.8 with 1 mM IPTG and then were cultured at 30 °C. Synthesis of metabolites was quantified via supernatant sampling over 24 h and analysis by HPLC. Compound confirmation was performed via ultra-performance liquid chromatography mass spectrometry (UPLC-MS).

Overexpression and purification of *N*-oxygenases

A strain of *E. coli* BL21 (DE3) harboring a pZE plasmid encoding expression of an *N*-oxygenase with a hexahistidine tag at either the N-terminus or C-terminus (NB01 or NB36) was inoculated from frozen stocks and grown to confluence overnight in 5 ml LB-containing kanamycin. Confluent cultures were used to inoculate 400 ml of the experimental culture of LB supplemented with kanamycin. The culture was incubated at 37 °C until an OD₆₀₀ of 0.5–0.8 was reached while in a shaking incubator at 250 r.p.m. *N*-oxygenase expression was induced by the addition of anhydrotetracycline (0.2 nM), and cultures were incubated at 30 °C for 5 h. Cultures were then grown at 20 °C for an additional

18 h. Cells were centrifuged using an Avanti J-15R refrigerated Beckman Coulter centrifuge at 4 °C at 4,000g for 15 min. The supernatant was then aspirated, and pellets were resuspended in 8 ml of lysis buffer (25 mM HEPES, 10 mM imidazole, 300 mM NaCl, 10% glycerol, pH 7.4) and disrupted via sonication using a QSonica Q125 sonicator with cycles of 5 s at 75% amplitude and 10 s off for 5 min. The lysate was distributed into microcentrifuge tubes and centrifuged for 1 h at 18,213g at 4 °C. The protein-containing supernatant was then removed and loaded into a HisTrap Ni-NTA column using an ÄKTA Pure GE fast protein liquid chromatography (FPLC) system. Protein was washed with three column volumes (CVs) at 60 mM imidazole and four CVs at 90 mM imidazole. *N*-oxygenase was eluted in 250 mM imidazole in 1.5 ml fractions. Selected fractions were denatured in Lamelli SDS reducing sample buffer (62.5 mM Tris-HCl, 1.5% SDS, 8.3% glycerol, 1.5% β -mercaptoethanol and 0.005% bromophenol blue) for 10 min at 95 °C and subsequently run on an SDS-PAGE gel with a Thermo Fisher Scientific PageRuler Prestained Plus ladder to identify protein-containing fractions and confirm their size. Gels were imaged using an Azure c280 imaging system. The *N*-oxygenase-containing fractions were combined and applied to an Amicon column (10 kDa MWCO), and the buffer was diluted 1,000 \times into a 20 mM Tris (pH 8.0) and 5% glycerol buffer.

***N*-oxygenase expression testing**

To test the expression of the *N*-oxygenase library, 5 ml cultures of NB15-NB35 were inoculated in 5 ml cultures of LB containing 50 μ g ml⁻¹ kanamycin and then grown at 37 °C until midexponential phase (OD = 0.5–0.8). At this time, cultures were induced via the addition of 0.2 nM anhydrotetracycline and then grown at 30 °C for 5 h before growing at 20 °C for an additional 18 h. After this time, 1 ml of cells was mixed with 0.05 ml of glass beads and then vortexed using a Vortex-Genie 2 for 15 min. After this time, the lysate was centrifuged at 18,213g at 4 °C for 30 min. Lysate was denatured, as described in ‘Overexpression and purification of *N*-oxygenases,’ and then subsequently run on an SDS-PAGE gel with an Invitrogen BenchMark His-tagged Protein Standard ladder and then analyzed via western blot with an HRP-conjugated 6*His His-Tag Mouse McAB primary antibody at 10,000 \times dilution. The blot was visualized using an Amersham ECL Prime chemiluminescent detection reagent in an Azure c280 imaging system.

In vitro *N*-oxygenase activity assay

Reactions were performed in 1 ml volumes consisting of 25 mM HEPES (pH 7.0), 25 mM NaCl and 1.5% H₂O₂ with 1 mM pA-Phe or pA-Pyr. The reaction mixture was incubated for 3 h at 25 °C with 10 μ M purified *N*-oxygenase, following which the reaction was terminated with the addition of 0.8% TFA. Following 1 h incubation, the mixture was centrifuged and the supernatant was analyzed by HPLC. pN-Phe production was further confirmed via UPLC-MS.

Nonstandard amino acid incorporation assays

Orthogonal AARS/tRNA pairs were cloned within pEVOL plasmids and transformed into an *E. coli* MG1655 (DE3) strain with a pZE plasmid expressing a reporter protein fusion consisting of a ubiquitin domain, followed by an in-frame amber suppression codon, followed by GFP (pZE-Ub-UAG-GFP). Sequences of reporter proteins used in this study can be found in Supplementary Table 10. These transformed strains were cultured at 37 °C in 300 μ l MOPS EZ Rich media with aromatic amino acid (Phe, Trp and Tyr) dropout in deep 96-well plates with a specified concentration of nonstandard amino acid, 34 μ g ml⁻¹ chloramphenicol and 50 μ g ml⁻¹ kanamycin with shaking at 1,000 r.p.m. and an orbital radius of 3 mm. At midexponential growth (OD ~ 0.5), 0.2 nM anhydrotetracycline and 0.2% (wt/vol) L-arabinose were added to induce transcription of mRNA that requires UAG suppression to form full-length GFP. Cultures were grown for 18 h at 30 °C before pelleting them via centrifugation. To eliminate possible fluorescence or absorbance via free nonstandard amino acids in culture media, cultures were

washed in PBS buffer before quantification of both GFP fluorescence at excitation and emission wavelengths of 488 and 528 nm, respectively, and OD₆₀₀ using a Spectramax i3x plate reader with Softmax Pro 7.0.3 software. For each synthetase, we performed this screen in the presence and absence of externally supplied nonstandard amino acid. For experiments using the single reporter/AARS/tRNA plasmid, the same protocol was performed without kanamycin or L-arabinose addition.

Coupled biosynthesis and incorporation

For the conversion of exogenously supplemented pA-Phe for subsequent incorporation of pN-Phe, *E. coli* MG1655 (DE3) was cotransformed with pEVOL-TetRSC-11, a pZE-ObiL construct expressing the *N*-oxygenase ObiL and ubiquitin-fused GFP reporter containing an amber suppression codon and a C-terminal hexahistidine tag encoded on a vanillic acid inducible promoter system (pCDF-Ub-UAG-GFP-C_{term}His_{6x}). This strain was cultured at 37 °C in 50 ml of LB broth in 250 ml baffled shake flasks with 0.2% (wt/vol) arabinose, 1 mM pA-Phe, 25.5 μ g ml⁻¹ chloramphenicol, 37.5 μ g ml⁻¹ kanamycin, 71.3 μ g ml⁻¹ streptomycin and 0.2 nM anhydrotetracycline in a shaking incubator at 250 r.p.m. At an OD₆₀₀ of 0.5–0.8, 1 mM vanillic acid was added to induce transcription of mRNA that requires UAG suppression to form full-length GFP. Cultures were then grown at 37 °C for an additional 18 h. The reporter protein was then lysed and purified using FPLC with a HisTrap column, as previously described. The protein sample was then concentrated using a 10 kDa MWC Amicon column and then diluted 10:1 in 10 mM ammonium acetate buffer and spun down to 1 ml samples three times. Then, the sample was diluted 10:1 in 2.5 mM ammonium acetate buffer and spun down to 1 ml samples three times. Protein in 2.5 mM ammonium acetate buffer was then submitted for whole-protein LC-MS.

For the production of pN-Phe without precursor supplementation, cultures of indicated strains were cultured in 50 ml of MOPS EZ Rich media with 1.5% glucose or glycerol and reduced levels of tyrosine (100 μ M as opposed to 200 μ M) in 250 ml baffled shake flasks in a shaking incubator at 250 r.p.m. and 37 °C. At an OD₆₀₀ of approximately 0.5, 1 mM IPTG was added to induce the pN-Phe synthesis plasmid. Cultures were grown at 30 °C for 2 h; following this time, the reporter and AARS were induced with 0.2 nM anhydrotetracycline and 0.2% (wt/vol) arabinose (if applicable). Cultures were then grown at 30 °C for an additional 18 h, following which the protein sample was prepared for whole-protein LC-MS, as described above.

HPLC analysis

Metabolites of interest were quantified via HPLC using an Agilent 1260 Infinity model equipped with a Zorbax Eclipse Plus-C18 column (part, 959701-902, 5 μ m, 95 Å, 2.1 \times 150 mm). To quantify amine-containing metabolites, an initial mobile phase of solvent A/B = 100/0 was used (solvent A, 20 mM potassium phosphate, pH 7.0; solvent B, acetonitrile/water at 50/50) and maintained for 7 min. A gradient elution was performed (A/B) with gradient from 100/0 to 80/20 for 7–18 min, gradient from 80/20 to 50/50 for 18–19 min, gradient from 50/50 to 100/0 for 19–20 min and equilibration at 100/0 for 20–24 min. A flow rate of 0.5 ml min⁻¹ was maintained, and absorption was monitored at 210, 270, 280 and 300 nm (Supplementary Fig. 19). HPLC chromatograms were collected and analyzed using Agilent ChemStation (Version C.01.10) or Agilent OpenLab (version 2.3) software.

MS

MS measurements for small molecule metabolites were submitted to a Waters Acquity UPLC H-Class coupled to a single quadrupole mass detector 2 with an electrospray ionization source. Metabolite compounds were analyzed using a Waters Cortecs UPLC C18 column (part, 186007092, 1.6 μ m, 90 Å, 2.1 \times 30 mm) with an initial mobile phase of solvent A/B = 95/5 (solvent A, water and 0.1% formic acid; solvent B, acetonitrile and 0.1% formic acid) with gradient elution from

(A/B) 95/5 to 5/95 over 5 min. Flow rate was maintained at 0.5 ml min⁻¹. For samples collected from *E. coli* growth cultures and standards for collection and subsequent UPLC–MS analysis, an initial submission to an Agilent 1100 series HPLC system with a Zorbax Eclipse Plus–C18 column (part, 959701-902) was used to collect pN-Phe elution peaks for enhanced MS resolution. A 100 µl injection was made with an initial mobile phase of solvent A/B = 95/5 (solvent A, water and 0.1% trifluoroacetic acid; solvent B, acetonitrile and 0.1% trifluoroacetic acid) and maintained for 1 min. A gradient elution was then performed (A/B) with gradient from 95/5 to 50/50 over 1–24 min, gradient from 50/50 to 95/5 over 24–25 min and equilibration at 95/5 for 25–27 min. Flow rate was 1 ml min⁻¹ and metabolites were tracked at 270 nm. pN-Phe elution was identified at 7.20 min using a chemical standard, and this peak was collected for submission to UPLC–MS.

For intact protein MS measurements, samples were submitted to a Waters Acquity UPLC H-Class with an Acquity Protein BEH C4 column (part, 186004495; 1.7 µm, 300 Å, 2.1 × 50 mm) coupled to a Xevo G2-XS Quadrupole Time-of-Flight (QToF) mass spectrometer. Protein sample was injected into a Waters Acquity with an initial mobile phase of solvent A/B = 85/15 (solvent A, water and 0.1% formic acid; solvent B, acetonitrile and 0.1% formic acid) held at 85/15 for 1 min followed by a gradient elution from (A/B) 85/5 to 5/95 over 5 min. Flow rate was maintained at 0.5 ml min⁻¹. The QToF mass spectrometer measurements were obtained following positive electrospray ionization (ES+) with a source temperature of 150 °C. Spectrum was analyzed from m/z 500 to 2,000, and the spectra were deconvoluted using the maximum entropy function in MassLynx software. Here an assumed uniform Gaussian distribution (width = half height of 0.60 Da with 30% intensity right and left channels) was used to evaluate a mass range from 37,000 to 38,000 Da (resolution, 0.1 Da/channel) at an elution time of 3.2 min.

For LC–MS/MS measurements, protein samples (20 µg per lane) were first submitted to SDS–PAGE. The gel was then stained with Bio-Safe Coomassie stain (Bio-Rad), and destained with water. Then, protein bands corresponding to the molecular weight of the protein of interest were cut from the polyacrylamide gel, and the gel fragments were digested using an in-gel tryptic digestion kit (Thermo Fisher Scientific, 89871). Following digestion, extracted peptides were then desalted using Pierce pipette tips (Thermo Fisher Scientific) and dried, resuspended in 20 µl of 0.5% acetic acid (pH4.5) and then subjected to LC–MS/MS using a Thermo Fisher Scientific Orbitrap Eclipse Tribrid Mass Spectrometer (MS; Thermo Fisher Scientific) with an Ultimate 3000 nano-LC and a FAIMS Pro Interface (Thermo Fisher Scientific). The LC–MS/MS analysis was performed using an Orbitrap Eclipse MS (Thermo Fisher Scientific). Peptides were first loaded onto a trap column (PepMap C18) and then separated by an analytical column (PepMap C18, 2.0 µm; 15 cm × 75 mm I.D.; Thermo Fisher Scientific) at 300 nl/min flow rate using a binary buffer system (buffer A, 0.1% formic acid in water; buffer B, 0.1% formic acid in acetonitrile) with a 165-min gradient (1%–10% in 8 min; then to 25% buffer B over 117 min; 25%–32% buffer B in 10 min, then to 95% buffer B over 3 min; back to 1% buffer B in 5 min, and stay equilibration at 1% buffer B for 20 min). Multiple CVs (–40, –55 and –75) were applied for FAIMS separation. For all experiments, the survey scans (MS1) were acquired over a mass range of 375–1,500 m/z at a resolution of 60,000 in the Orbitrap. The maximum injection time was set to dynamic, and the AGC target was set to standard. Monoisotopic peak selection was set to peptides, and the charge state filter was set to 2–7. For MS/MS acquisition, precursors were isolated with a width of 1.6 m/z, fragmented with high energy collision dissociation (HCD) using 30% collision energy with a maximum injection time of 100 ms and collected in Orbitrap at 15,000 resolution. The dynamic exclusion was set to 60 s and can be shared across different FAIMS experiments. LC–MS/MS data were collected in *n* = 3 independent biological replicates.

Proteomic analysis was performed in the MaxQuant–Andromeda software suite (version 1.6.3.4) with most of the default parameters⁴⁷.

An *E. coli* reference proteome (strain BL21-DE3; taxonomy_id:469008) was used for the database search. Other parameters include trypsin as an enzyme with maximally two missed cleavage sites; protein N-terminal acetylation, methionine oxidation, and pA-Phe and pN-Phe substitution for tyrosine were entered as variable modifications; cysteine carbamidomethylation as a fixed modification and peptide length was set to a minimum of seven amino acids. The false-discovery rate (FDR) of high-confidence protein and peptide identification was 1%. Peptide intensity values derived from MaxQuant were used for quantification. To obtain the MS/MS spectra of peptides (Supplementary Fig. 18), additional analysis was performed using Proteome Discoverer software (version 3.0; Thermo Fisher Fisher). Raw data were searched against the *E. coli* reference proteome (strain BL21(DE3)) with National Center for Biotechnology Information (NCBI) reference Proteome (469008) using Sequest HT Processor and consensus workflow (CWF) Basic analysis workflows. Iodoacetamide-mediated cysteine carbamidomethylation was set as a static modification, while methionine oxidation and pA-Phe and pN-Phe substitution for tyrosine were entered as dynamic modifications. Precursor mass tolerance was set at 10 ppm while allowing fragment ions to have a mass deviation of 0.02 Da for the HCD data. Validation of peptide-spectrum matches based on *q* value was done using a percolator, with target FDRs of 1% and 5% for stringent and relaxed validations, respectively. The FDR of high-confidence protein and peptide identification was 1%. The raw files and MaxQuant results have been deposited at the MASSIVE repository (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>) with the dataset identifier MSV000091379 (doi: 10.25345/C5BV7B546).

Creation of SSN

Using NCBI BLAST, the 1,000 most closely related sequences, as measured by BLASTP alignment score, were obtained from four characterized diiron monooxygenase-type *N*-oxygenases with activity on aromatic amines—AurF, CmlI, Azoc⁴⁸ and ObiL. After deleting duplicate sequences, 2,134 unique sequences were obtained, which were then submitted to the Enzyme Function Initiative-Enzyme Similarity Tool⁴⁹ to generate a SSN. Sequences exhibiting greater than 95% similarity were grouped into single nodes, resulting in 775 unique nodes, and a minimum alignment score of 100 was selected for node edges. Inspection was performed to identify nodes corresponding to previous literature-verified *N*-oxygenases (PvfB⁵⁰, PsAAO⁵¹, CmlI²⁹, Azoc⁴⁸, AurF(RJ)⁵², AurF³⁷, AlmD⁵³, HamC⁵⁴ and Bez⁵⁵), which were then colored to indicate prior work.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The datasets generated during and/or analyzed during the current study are contained in the published article (and its Supplementary Information) and are publicly accessible via cited repositories or are available from the corresponding author upon reasonable request. Source data are provided with this paper.

References

- Cui, L. & Shearwin, K. E. *Methods in Molecular Biology* Vol. 1472, pp. 139–155 (Humana Press, 2017).
- Wannier, T. M. et al. Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl Acad. Sci. USA* **117**, 13689–13698 (2020).
- Kunjapur, A. M., Hyun, J. C. & Prather, K. L. J. Dereglulation of S-adenosylmethionine biosynthesis and regeneration improves methylation in the *E. coli de novo* vanillin biosynthesis pathway. *Micro. Cell Fact.* **15**, 61 (2016).

47. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
48. Guo, Y. Y. et al. Molecular mechanism of azoxy bond formation for azoxymycins biosynthesis. *Nat. Commun.* **10**, 4420 (2019).
49. Gerlt, J. A. et al. Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **1854**, 1019–1037 (2015).
50. Kretsch, A. et al. Discovery of (dihydro)pyrazine *N*-oxides via genome mining in *Pseudomonas*. *Org. Lett.* **20**, 4791–4795 (2018).
51. Platter, E., Lawson, M., Marsh, C. & Sazinsky, M. H. Characterization of a non-ribosomal peptide synthetase-associated diiron arylamine *N*-oxygenase from *Pseudomonas syringae* *pv.* *phaseolicola*. *Arch. Biochem. Biophys.* **508**, 39–45 (2011).
52. Indest, K., Eberly, J. & Hancock, D. Expression and characterization of an *N*-oxygenase from *Rhodococcus jostii* RHA1. *J. Gen. Appl. Microbiol.* **61**, 217–223 (2015).
53. Cortina, N. S., Revermann, O., Krug, D. & Müller, R. Identification and characterization of the althiomycin biosynthetic gene cluster in *Myxococcus xanthus* DK897. *ChemBioChem* **12**, 1411–1416 (2011).
54. Jenul, C. et al. Biosynthesis of fragin is controlled by a novel quorum sensing signal. *Nat. Commun.* **9**, 1297 (2018).
55. Tsutsumi, H. et al. Unprecedented cyclization catalyzed by a cytochrome P450 in benzastatin biosynthesis. *J. Am. Chem. Soc.* **140**, 6631–6639 (2018).

Acknowledgements

We thank the Mass Spectrometry Facility at the University of Delaware for the mass spectrometry analysis that is supported by the National Institute of General Medical Sciences or the National Institutes of Health under award P20GM104316 and Y. Yu and P.N. Asare-Okai in particular for their assistance. We acknowledge support from the following funding sources: The National Science Foundation (NSF CBET-2032243, to A.M.K.), University of Delaware Start-Up Funds (to A.M.K.), the Mort Collins Foundation (to N.D.B.) and minor research support as part of the Center for Plastics Innovation, an Energy Frontier

Research Center funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences, under award DE-SC0021166 (to A.M.K.). We are also grateful to the American Institute of Chemical Engineers for their support of this concept through the 2021 Langer Prize for Innovation and Entrepreneurial Excellence (to A.M.K.). We also thank M. Jones for valuable experimental troubleshooting suggestions for this work.

Author contributions

A.M.K. conceived and supervised the study; N.D.B. designed and performed all experiments, analyzed data, prepared figures and wrote the manuscript; M.L. aided with molecular cloning; S.S. cloned the *N*-oxygenase library and confirmed expression and L.B.B. cloned the *MjTyrRS* variants tested in this study.

Competing interests

N.D.B. and A.M.K. are co-inventors on a filed patent application related to this work that has now been transferred to a commercial entity cofounded by the authors (Nitro Biosciences). A.M.K. also serves on the Scientific Advisory Board of Wild Microbes. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-023-01338-x>.

Correspondence and requests for materials should be addressed to Aditya M. Kunjapur.

Peer review information *Nature Chemical Biology* thanks Jorge Marchand and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-----|-----------|
| n/a | Confirmed |
|-----|-----------|
- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
 - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
 - The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
 - A description of all covariates tested
 - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
 - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
 - For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
 - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
 - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
 - Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Plate-based fluorescence and optical density measurements were collected using Softmax Pro 7.0.3 software with a Spectramax i3x platereader. HPLC data were collected using Agilent ChemStation (Version C.01.10) or OpenLab (Version 2.3) software. Gels and blots were imaged using an Azure c280 imaging system. LC-MS/MS data was collected using an Orbitrap Eclipse MS (Thermo Scientific).

Data analysis

HPLC data were analyzed using Agilent ChemStation (Version C.01.10) or OpenLab (Version 2.3) software. Microsoft Excel was used to analyze platereader data, and Graphpad Prism was used to generate plots and calculate standard deviations. Whole protein and small molecule LC/MS data was analyzed using MassLynx (Version 4.2) software. Tryptic digestion LC-MS/MS data was analyzed using MaxQuant-Andromeda software. Additional sequences for N-oxygenases were identified using a BLASTp search through NCBI and a sequence similarity network was created using the EFI-Enzyme Similarity Tool from the University of Illinois. Sequence similarity networks were visualized using Cytoscape (Version 3.9.1). Protein structures on N-oxygenases were created using AlphaFold2 via ColabFold and were visualized using PyMOL (Version 2.5.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The datasets generated during and/or analyzed during the current study are contained in the published article (and its Supplementary Information) or are available from the corresponding author on reasonable request. For database searching used in proteomics, an E. coli reference proteome (BL21-DE3; taxonomy_id:469008) was used in addition to fasta file containing the sequence of the ubiquitin-GFP reporter (found in the SI).

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	<input type="text" value="N/A"/>
Population characteristics	<input type="text" value="N/A"/>
Recruitment	<input type="text" value="N/A"/>
Ethics oversight	<input type="text" value="N/A"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were selected based on acceptable convention in the field where 3 replicates is standard practice for plate reader experiments (examples include: (i) doi:10.1038/s41467-021-25691-4 and (ii) doi:10.1038/nbt.3372) and HPLC analysis of metabolites produced via fermentation (examples include: (i) doi:10.1038/ncomms11709 and (ii) doi: 10.1038/s41589-020-00684-4). Single purified protein samples were used for intact LC/MS while 3 replicates were used for tryptic digestion LC-MS/MS. Previous studies investigating nsAA incorporation only report single replicates for quantification/confirmation (examples include: (i) doi:10.1002/cbic.201600448 (ii) doi: 10.1039/c7ob00582b (iii) doi:10.1021/acscchembio.9b01026). No calculation was used to determine sample size, but experiments were repeated to increase replicability.
Data exclusions	No data were excluded.
Replication	All attempts to replicate were successful, and the number of replicates is reported for each experiment. Some full experiments were performed in replicates on separate days, exhibiting similar results with a single day's set of replicates chosen here. Plate reader-based experiments were normally replicated at least twice and most fermentation-based production samples were replicated on 1-2 separate days.
Randomization	Experimental samples using the same strain but coming from different groups were generated from the same source. If one strain was exposed to +nsAA and -nsAA conditions, a single culture of that strain was split immediately before nsAA was added. For this study, randomization was not relevant, as all bacterial cells were analyzed equally.
Blinding	All experiments were unblinded as there would be difficulty fully blinding studies to ensure proper growth and media conditions, but is unlikely researcher bias in would result in replicated effects.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

- n/a | Involved in the study
- Antibodies
 - Eukaryotic cell lines
 - Palaeontology and archaeology
 - Animals and other organisms
 - Clinical data
 - Dual use research of concern

- n/a | Involved in the study
- ChIP-seq
 - Flow cytometry
 - MRI-based neuroimaging

Antibodies

Antibodies used

Horseradish peroxidase-conjugated 6*His, His-tag monoclonal antibody was used from Proteintech (catalog number: HRP-66005; clone number: 21006227)

Validation

Validation was performed by vendor: "Recombinant protein were subjected to SDS PAGE followed by western blot with HRP-66005 (6*His, His-Tag antibody) at dilution of 1:10000 incubated at room temperature for 1.5 hours."