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

Neil D. Butler[1], Sabyasachi Sen[1], Lucas B. Brown[2], Minwei Lin[1], and Aditya M. Kunjapur[1]*

[1] *Department of Chemical & Biomolecular Engineering, University of Delaware, Newark, DE 19716, United States*

[2] *Present Address: Systems, Synthetic, and Physical Biology Program, Rice University, Houston, TX 77005, United States*

***Corresponding Author**

**Aditya Kunjapur** - *Department of Chemical & Biomolecular Engineering, University of Delaware, Newark, DE 19716, United States*; orcid.org/0000-0001-6869-9530
Email: kunjapur@udel.edu

**Abstract:**

The incorporation of the non-standard amino acid (nsAA) *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 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 first-time biosynthesis of pN-Phe in *Escherichia coli* by creating a pathway that features a previously uncharacterized non-heme diiron *N*-monooxygenase, which resulted in pN-Phe titers of 820 ± 130 μM after optimization. After we identified an orthogonal translation system that exhibited selectivity towards 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 towards distributed and autonomous production of nitrated proteins.

## Introduction

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 twenty standard amino acids contain nitroaromatic functionality nor exhibit similar levels of electron-withdrawing potential. One approach to overcome this limitation is to incorporate nitroaromatic non-standard amino acids (nsAAs), also referred to as non-canonical amino acids or unnatural amino acids, at specified protein residues in live cells using genetic code expansion technology[1–3]. Across these studies, the nitroaromatic nsAA *para*-nitro-L-phenylalanine (pN-Phe) has been of heightened interest given its diversity of applications (**Fig. 1A**), including as a peptide distance marker[3], an enhancer of activity for nitroreductase enzymes[4], 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 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 murine 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 systems[8] and the recognition of these neoepitopes by $CD4^+$ T cells during peptide:MHC-II recognition[7]. The T cells subsequently activate B cells that produce antibodies capable of binding the wild-type protein. A broad range of proteins have been targeted using this strategy, including TNF-α, C5α, RANKL, HER2, and PD-L1[5–7,9–11]. Compared to other nsAAs, pN-Phe proved the most effective at eliciting an immune response and has been reported in the largest number of investigations[7]. However, one major drawback of the approach of administering nitrated proteins was the need for aggressive immunization schedules ranging from 4 injections over 28 days[7] to 8 injections over 17 days[5].

Microbial biosynthesis of nsAAs presents a promising methodology to solve many limitations of biomanufacturing proteins that contain an expanded genetic code by forming the nsAA intracellularly and on-demand prior to incorporation within proteins[12,13]. In addition, the generation of a bacterial strain that can autonomously access an expanded genetic code could 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 disease[14], high antigen loading potential (potentially 100,000 copies displayed/cell in *E. coli*)[15], and adjuvanticity arising from the bacterial cell[16].

Toward the on-demand production of proteins with non-standard functionalities more generally, a handful of nsAAs have been metabolically biosynthesized and incorporated in distinct engineering efforts, including *para*-amino-L-phenylalanine (pA-Phe)[17], L-sulfotyrosine[18], L-dihydroxyphenylalanine[19], 4-nitro-L-tryptophan[20], 5-hydroxy-L-tryptophan[21], and L-propargylglycine[22]. While significant advancements, most of these nsAAs 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 identified[23]. As such, pN-Phe biosynthesis would require *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 non-heme 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 obafluorin. 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 AurF[28] (with a native substrate of para-aminobenzoic acid) and CmlI[29] (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 *Escherichia coli* which utilized four heterologous enzymes, including an *N*-oxygenase enzyme for nitro group formation and the *papABC* operon from *S. venezuelae*. Through chassis engineering and *N*-oxygenase bioprospecting, we were able improve pN-Phe titers to near millimolar levels and decrease accumulation of the nsAA intermediate pA-Phe. We then identified an 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 and discussion

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 - RetroBioCat[30] and ATLAS of Biochemistry[31] - 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**): (i) Three heterologous enzymes from the chloramphenicol biosynthesis pathway in *Streptomyces venezuelae* that generate pA-Phe from chorismate; (ii) the putative *N*-oxygenase ObiL; (iii) 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 activity[32] 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. 2**). 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 reported polyspecificity of native aromatic aminotransferases measured *in vitro*[33].

Next, we sought to identify an *N*-oxygenase that could fully oxidize pA-Phe or pA-Pyr. Recent work identified the putative diiron monooxygenase ObiL in the synthesis of obafluorin[26]. In that study, the characterization of ObiL featured a colorimetric assay to measure oxidation of pA-Pyr rather than direct chromatographic or spectrometric confirmation. To directly assay ObiL activity on pA-Pyr using 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. 3**). We then performed *in vitro* characterization on ObiL using $H_2O_2$ as a reductant to recycle the diiron core of the enzyme[28]. Here, we demonstrated that

ObiL was active on both pA-Pyr and pA-Phe, although only 10% conversion of 1 mM substrate was achieved after 3 *in vitro* (**Fig. 2B**). We next investigated heterologous activity of ObiL in MG1655 using metabolite supplementation to fermentations. Once again, we observed formation of nitroaromatic products upon supplementation of 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 implemented 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 synthesis of pA-Phe using the previously established *papABC* operon from *S. venezuelae*. We integrated the native operon into a single 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 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 DAHP synthase[34] (*aroG*_D146N, or *aroG**) on a separate plasmid from *papABC* genes (strain NB06), resulting in significant 24 h post-induction pA-Phe titer of 624 ± 79 μM.

We next integrated the *N*-oxygenase *obiL* into the pathway in an operon with *aroG** to investigate whether co-expression of these genes with the *papABC* operon would achieve pN-Phe biosynthesis. We were encouraged to observe pN-Phe, albeit with fairly low 24 h titer of 83 ± 13 μM (strain NB07). To reduce metabolic burden and limit the number of plasmids for downstream site-specific incorporation steps, we moved the *obiL-aroG** operon into the pCola plasmid (**Fig. 3B**). This strain achieved a doubling of pN-Phe titer to 153 ± 22 μM (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 (Δ*tyrA* and Δ*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)*ΔpheA* strain. (**Fig. 3C**). Using this strain in M9-glucose media with aromatic amino acids supplemented (0.04 mg/ml L-Phe, 0.04 mg/ml L-Tyr, and 0.04 mg/mL L-Trp), we achieved our highest measured pN-Phe titer of 260 ± 11 μM with the single plasmid system (strain NB14), though final cell density for the dual plasmid system was lower than that of the single plasmid system (**Supplementary Table 6**). Despite promising pN-Phe titer, a higher titer of pA-Phe (454 ± 28 μM) was measured using the engineered strain with the single plasmid system. Given that pA-Phe is a structurally related nsAA 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 conversion of pA-Phe to pN-Phe. Previously characterized non-heme diiron monooxygenases have demonstrated activity on diverse aromatic amines and compatibility with expression in *E. coli*. Thus, while 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**). 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 supplementation to transformed and metabolically active *E. coli* MG1655. 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 beta-lactam obafluorin (**Supplementary Table 8**). 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]. 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 (pdb: 3cht[37]) showed binding pocket differences that could influence pA-Phe binding (**Supplementary Fig. 9**).

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 two-fold higher yield of pN-Phe from pA-Phe compared to the yield catalyzed by ObiL, converting 1 mM pA-Phe to 196 ± 23 μM 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 co-expressing NO16 (330 ± 75 μM), as compared to ObiL (197 ± 25 μM) (**Fig. 4E**). 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 ± 130 μM after 24 h using NO16 compared to a titer of 435 ± 23 μM using ObiL (**Fig. 4F**). The use of LB-glucose media resulted in no distinguishable difference in activity between ObiL and NO16 (**Supplementary Fig. 10**). To further investigate these results, we repeated *in vivo* supplementation experiments for ObiL and NO16 in LB media compared to M9 (**Supplementary Fig. 11).** While ObiL and NO16 appear to perform comparably in most conditions, in M9-minimal 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. Furthermore, given that genetic code expansion efforts often supplement nsAAs at 1 mM concentration, after achieving pN-Phe titers greater than 0.8 mM, 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 exhibits sufficient activity at sub-millimolar 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* utilized an evolved *Mj*TyrRS (referred to as pNFRS) paired with *Mj*tRNA$^{Tyr}_{CUA}$[3]. However, the activity of this system on other nsAAs has not been reported, and many *Mj*TyrRS derivatives exhibit substrate polyspecificity[38,39]. Hence, we evaluated activity and selectivity of a panel of 20 *Mj*TyrRS derivatives that were reported to have high demonstrated activity on one or more nsAAs (**Supplementary Table 4**). To test site-specific incorporation of pN-Phe within a target protein, we used a Ubiquitin-sfGFP fusion reporter system 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 nsAA is added to the growing polypeptide at that position. In addition, to improve signal to noise, we performed screening in an aromatic amino acid dropout variation of MOPS EZ Rich medium (**Fig. 5A)**. In this initial screen, we identified five AARSs which achieved over five-fold enhancement in $OD_{600}$ normalized GFP fluorescence after supplementation of pN-Phe: (i) TetRS-C11[40], (ii) pCNFRS[38], (iii) NapARS[41], (iv) pAcFRS[42], and (v) pNFRS (**Fig. 5B**). 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. 13**). We then investigated the dose-response to pN-Phe at concentrations from 200 to 1000 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. Given that TetRS-C11 exhibited the most efficient incorporation of pN-Phe within proteins at lower concentration and had low fold-change enhancement at relevant pA-Phe concentrations, we chose to investigate whether we could engineer this AARS for improved selectivity and titer (**Supplementary Figs. 14-15**). 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 nor activity, we continued with TetRS-C11.

To test the coupling of nsAA biosynthesis and incorporation in stages, we first investigated whether we could couple an *N*-oxygenase with the orthogonal translation system and reporter. We co-expressed the TetRS-C11 AARS/*Mj*tRNA$^{Tyr}_{CUA}$ pair along with ObiL and a C-terminal heptahistidine-tagged version of the ubiquitin-fused 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 MS, where we saw two major masses, one at 37274.5 Da and one at 37305 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.

<u>Coupled *de novo* biosynthesis and incorporation of pN-Phe</u>

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 ubiquitin-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 ubiquitin-GFP 2 h later to allow accumulation of pN-Phe prior to induction of the AARS. Surprisingly, we

obtained a mass at ~63 Da less than predicted, corresponding to incorporation of either glutamine or lysine (**Fig. 6A**). While introduction of glutaminyl amber suppression tRNA is common when genetic engineering using phage, the listed phenotype for the NST37 strain did not note this mutation. However, sequencing of this region indicated that *glnX* tRNA had a CUA anticodon. Thus, we modified the anticodon to the wild-type CUG (**Fig. 6B**).

Next, we evaluated how pN-Phe titer would be affected by the burden of expressing biosynthetic genes, incorporation machinery, and ubiquitin-GFP reporter using separate plasmids. Here, we measured titer at 3 h and 24 h from fermentative growth using MOPS EZ Rich-glycerol media to enable induction of the $P_{araBAD}$ promoter for AARS translation (**Fig. 6C**). We obtained low pN-Phe titers at both 3 h and 24 h of 10.9 ± 0.9 μM and 50 ± 7 μM 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 ubiquitin-GFP reporter within an aTc-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 separate reporter and AARS/tRNA plasmid. In addition, for pRepAARS, we observed an $OD_{600}$-normalized fold-change in GFP fluorescence of approximately 4-fold over background at pN-Phe concentrations as low as 50 μM (**Fig. 6D**).

We then evaluated pN-Phe synthesis utilizing pRepTetRS-C11 with the pN-Phe synthesis plasmid in NST37(DE3)Δ*pheA glnX*$_{CUG}$ via fermentation with MOPS EZ Rich-glucose media. We induced the pN-Phe synthesis pathway at mid-exponential phase (OD~0.5) and followed with induction pRepTetRS-C11 2 h later. Using this protocol, we observed extracellular pN-Phe titers at 28 ± 4 μM after 4 h and 360 ± 60 μM after 24 h (**Fig. 6E**), which was 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. 16**).

Given the robust incorporation efficiency of TetRS-C11, even at the lowest levels of pN-Phe supplementation, and observed green fluorescence, we chose to investigate whether mid-exponential phase induction of the biosynthetic pathway would result in reporter protein that contained pN-Phe at the target site. Thus, we induced the pN-Phe synthesis pathway at OD~0.5 and the pRepTetRS-C11 plasmid 2 h later. Then, we purified the ubiquitin-GFP reporter and obtained a protein titer of 0.53 mg/L. To confirm incorporation of pN-Phe, we performed intact protein MS, which encouragingly revealed a mass peak for our protein matching that expected from the incorporation of pN-Phe (**Fig. 6F**). Additional mass peaks suggested incorporation of pA-Phe/Tyr into some of the proteins. To investigate this, subsequent fermentations were performed for production of pN-Phe containing protein (achieving 3.5 ± 0.8 mg/L purified protein) for tryptic digestion LC-MS/MS. The protein sample was then run on an SDS-PAGE gel and the band corresponding to the intact protein of interest was excised for in-gel tryptic digestion (**Supplementary Fig. 18**). Submission of the digested protein to LC-MS/MS confirmed 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. 19**). Despite some misincorporation observed, this result demonstrated 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 nsAAs 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 which perform these chemistries. Here, to achieve biosynthesis of pN-Phe, we integrated a non-heme diiron monooxygenase-type *N*-oxygenase from the pathway for obafluorin synthesis to perform nitro group formation. While the *in vivo* activity of certain *N*-oxygenases has been demonstrated in *E. coli*[29], this class of enzymes had not previously been utilized 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 non-heme 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 non-heme 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 non-heme diiron class have been found and extensively studied in *E. coli*. One example is the ribonucleotide reductase R2 (RNR-R2) which converts ribonucleotides to deoxyribonucleotides using iron-catalyzed $O_2$ activation similar to ObiL[43]. Given we do not see pA-Phe/pA-Pyr oxidation in the absence of an exogenous *N*-oxygenase and RNR-R2 does not bear much sequence similarity to ObiL, it is unlikely RNR-R2 performs amine oxidation. However, the cytoplasmic environment is known to be capable of reduction of the diiron cores for 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 nsAA - 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 co-expression of all modules and the accumulation of the nsAA pA-Phe. A threshold nsAA 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 nsAA biosynthesis is a critical variable, because efficient translation of the target protein requires reaching a threshold titer while in favorable regimes for translation, such as exponential growth phase. In scenarios where biosynthetic pathways to produce an nsAA feature another nsAA 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 selectivity of pN-Phe incorporation through a combination of altering plasmid configurations, screening *N*-oxygenase orthologs, varying timing of induction, identifying a highly selective AARS, and removing an undesired genomic amber suppressor tRNA. As more complex nsAA 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 that nitrated proteins could be produced using a model bacterium without supplementation 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 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, 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 an *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.

**Acknowledgements**

We thank the Mass Spectrometry Facility in the University of Delaware for the mass spectrometry analysis which is supported by National Institute of General Medical Sciences or the National Institutes of Health under Award Number P20GM104316 and Yanbao Yu and PapaNii Asare-Okai in particular for their assistance. We acknowledge support from the following funding sources: The National Science Foundation (NSF CBET #2032243, A.M.K), University of Delaware Start-Up Funds (A.M.K.), the Mort Collins Foundation (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 No. # DE-SC0021166 (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 (A.M.K.). We also thank Michaela Jones for valuable experimental troubleshooting suggestions to this work.

## Author Statements

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; L.B.B. cloned the *Mj*TyrRS variants tested in this study.

## Declaration of Competing Interest

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

**Figure 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 utilized in several applications. Those depicted are: **(i)** Incorporation into self-proteins for immunochemical termination of tolerance toward the 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 of our *de novo* synthesis pathway for pN-Phe in *E. coli* which 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 *in vivo* integration); **(ii)** an *N*-oxygenase with activity on pA-Phe or pA-Pyr (shown in red to indicate a lack of prior *in vivo* integration); **(iii)** an endogenous, promiscuous aminotransferase capable of converting aromatic keto acids to amino acid (shown in gray to indicate it is an endogenous enzyme).

**Figure 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 mid-exponential 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 within fermentative cultures in LB media expressing ObiL. Samples sizes are n=3 using biological replicates. Data shown are mean ± standard deviation.

**Figure 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 ± standard deviation. **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) *ΔpheA* 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 a star (*).GIK, glucokinase; AroG, phenylalanine-sensitive Phospho-2-dehydro-3-deoxyheptonate (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.

**Figure 4. Bioprospecting of non-heme diiron monooxygenases for activity on pA-Pyr and pA-Phe.** **a**, Sequence similarity network generated using 2134 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 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 ºC. **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 minimal media. **f**, Synthesis of pN-Phe in MOPS EZ-Rich media. Samples sizes are n=3 using biological replicates. Data shown are mean ± standard deviation.

**Figure 5. Aminoacyl-tRNA synthetase screening for incorporation of pN-Phe. a**, A ubiquitin-fused GFP reporter with a TAG codon at the N-terminus of the GFP 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 were screened for pN-Phe incorporation by comparison of the fold change in fluorescence (ex: 485 nm, em: 510 nm) normalized by $OD_{600}$ between cultures supplemented with 1 mM pN-Phe and cultures with no nsAA supplemented. **c**, Top-performing AARS's were screened for activity at lower pN-Phe concentration and **d,** off-target incorporation of Tyr or pA-Phe. 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 ubiquitin-GFP synthesized with pA-Phe supplementation and *N*-oxygenase ObiL expression. Samples size is n=3 using biological replicates and data shown are mean ± standard deviation with the exception of mass spectra.

**Figure 6. Integration of the pN-Phe biosynthesis pathway with orthogonal *Mj*TyrRS/tRNA for *de novo* biosynthesis and incorporation. a**, Mass spectra from intact protein MS of Ub-GFP purified from NST37(DE3) *ΔpheA* expressing three plasmids with primary mass peak corresponding to glutaminyl incorporation. The higher predicted mass prediction for the Ub-GFP construct relative to Figure 5F is due to an L213F mutation between plasmids in this figure. **b**, Sequencing revealed mutation of *glnX* gene leading to glutaminyl amber suppression addressed via MAGE. **c**, Titer of pN-Phe and pA-Phe in NST37(DE3) *ΔpheA* expressing a separate reporter (pZE-Ub-UAG-GFP), AARS/tRNA (pEVOL-TetRS-C11), and pN-Phe synthesis plasmid (pCola-papABC-AroG*-NO16) using MOPS EZ Rich media with glycerol. **d**, Integration of ubiquitin-GFP reporter in 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) *ΔpheA* expressing pN-Phe synthesis plasmid (pCola-papABC-AroG*-NO16) and AARS/tRNA/reporter plasmid (pRepTetRS-C11) using MOPS EZ Rich media with glucose. **f**, Mass spectra from intact protein MS of Ub-GFP with pN-Phe incorporated from metabolic synthesis in *E. coli* NST37(DE3) *ΔpheA* $glnX_{CUG}$. Samples size is n=3 using biological replicates and data shown are mean ± standard deviation with the exception of mass spectra.

**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 U S A* **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 U S A* **106**, 4337–4342 (2009).

7. Gauba, V. *et al.* Loss of CD4 T-cell-dependent tolerance to proteins with modified amino acids. *Proceedings of the National Academy of Sciences* **108**, 12821–12826 (2011).

8. Shorter, J. Electronic Effects of Nitro, Nitroso, Amino and Related Groups. in *PATAI'S Chemistry of Functional Groups* (John Wiley & Sons, Ltd, 2009). doi: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 for Creation of Cells with Site-Specific Tyrosine Sulfation. *bioRxiv* 2022.03.25.485857 (2022) doi:10.1101/2022.03.25.485857.

19. Kim, S., Sung, B. H., Kim, S. C. & Lee, H. S. Genetic incorporation of l -dihydroxyphenylalanine (DOPA) biosynthesized by a tyrosine phenol-lyase. *Chemical Communications* **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. *Chemical Communications* **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 obafluorin 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 Science* **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 2021 596:7873* **596**, 583–589 (2021).

36. Mirdita, M. *et al.* ColabFold - Making protein folding accessible to all. *bioRxiv* 2021.08.15.456425 (2021) doi:10.1101/2021.08.15.456425.

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 U S A* **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 U S A* **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 U S A* **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).

## Materials and Methods

### Strains and plasmids

*Escherichia coli* strains and plasmids used are listed in **Supplementary Table 1**. Molecular cloning and vector propagation were performed in DH5α. Polymerase chain reaction (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 2**. 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 Table 3** and **Table 4**. The plasmids pOSIP-TH (Addgene plasmid # 45978) and pE-FLP (Addgene plasmid # 45978) were gifts from Drew Endy & Keith Shearwin[44]. The *papABC* operon was kindly provided by Professor Ryan Mehl of Oregon State University in plasmid pLASC-lppPW. The pORTMAGE-EC1 recombineering plasmid was kindly provided by Timothy Wannier[45].

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 5**) was genomically integrated using one-step clonetegration with the pOSIP-TH plasmid. Following plasmid assembly, NST37 was transformed with the clonetegration plasmid, and the integration of the region for T7 polymerase activity (DE3) was confirmed via selection on LB-agar plates containing 9 μg/mL 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*$_{CUG}$ 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, HEPES, 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 (aTc) and isopropyl β-D-1-thioglactopyranoside (IPTG) were purchased from Cayman Chemical. Acetonitrile, sodium chloride, LB Broth powder (Lennox), LB Agar powder (Lennox), were purchased from Fisher Chemical. L-Arabinose was purchased from VWR. A MOPS EZ rich defined medium kit and components for was 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 (NEB). Sybr Safe DNA gel stain and BenchMark™ His-tagged Protein Standard were purchased from Invitrogen. HRP-conjugated 6*His His-Tag Mouse McAB was obtained from Proteintech.

Culture conditions

Cultures for general culturingand *N*-oxygenase protein overexpression were grown in LB-Lennox medium (LB: 10 g/L bacto tryptone, 5 g/L sodium chloride, 5 g/L 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 media[46] with Corning® Trace Elements A (1.60 µg/mL $CuSO_4 \cdot 5H_2O$, 863.00 µg/mL $ZnSO_4 \cdot 7H_2O$, 17.30 µg/mL Selenite • 2Na, 1155.10 µg/mL 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 L-phenylalanine, 0.04 mg/ml L-tyrosine, and 0.04 mg/mL L-tryptophan were added to the media to ensure growth.

For stability testing, a culture of *E. coli* K12 MG1655 (DE3) was inoculated from a 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 Scientific™ 260251) at 100x dilution. Cultures were supplemented with 0.5 mM of heterologous metabolites (pA-Phe, pA-Pyr, 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 ºC with shaking at 1000 RPM 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 100x 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 ºC 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 µg/mL chloramphenicol (Cm), 50 µg/mL kanamycin (Kan), 50 µg/mL carbenicillin (Carb), or 95 µg/mL streptomycin (Str)). Cultures were incubated at 37 ºC with shaking at 1000 RPM and an orbital radius of 3 mm until an $OD_{600}$ of 0.5-0.8 was reached. $OD_{600}$ was measured using a Thermo Scientific™ BioMate™ 160 UV-Vis Spectrophotometer. At this point, the pathway plasmids were fully induced with addition of corresponding inducer (1 mM IPTG, 1 mM vanillate, or 0.2 nM aTc), and the metabolite of interest was supplemented at this time. Cultures were incubated over 24 h at 37 ºC with sampling and metabolite concentration 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 100x 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 ºC at 250 RPM. Expression vectors were fully induced at $OD_{600}$ 0.5-0.8 with 1 mM IPTG, and then were cultured at 30 ºC. Synthesis of metabolites was quantified via supernatant sampling over 24 h and analysis by HPLC. Compound confirmation was performed via 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 experimental culture of LB supplemented with kanamycin. The culture was incubated at 37 ºC until an $OD_{600}$ of 0.5-0.8 was reached while in a shaking incubator at 250 RPM. *N*-oxygenase expression was induced by addition of anhydrotetracycline (0.2 nM), and cultures were incubated at 30 ºC for 5 h. Cultures were then grown at 20 ºC for an additional 18 h. Cells were centrifuged using an Avanti J-15R refrigerated Beckman Coulter centrifuge at 4 ºC at 4,000 *g* for 15 min. 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 minutes. The lysate was distributed into microcentrifuge tubes and centrifuged for 1 h at 18,213 x *g* at 4 ºC. The protein-containing supernatant was then removed and loaded into a HisTrap Ni-NTA column using an ÄKTA Pure GE FPLC system. Protein was washed with 3 column volumes (CV) at 60 mM imidazole and 4 CV 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% beta-mercaptoethanol, 0.005% bromophenol blue) for 10 minutes at 95 ºC and subsequently run on an SDS-PAGE gel with a Thermo Scientific PageRuler$^{TM}$ 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 applied to an Amicon column (10 kDa MWCO) and the buffer was diluted 1,000x into a 20 mM Tris pH 8.0, 5% glycerol buffer.

*N*-oxygenase expression testing

To test expression of the *N*-oxygenase library, 5 mL cultures of NB15-NB35 were inoculated in 5 mL cultures of LB containing 50 µg/mL kanamycin and then grown at 37 ºC until mid-exponential phase (OD = 0.5-0.8). At this time, cultures were induced via addition of 0.2 nM aTc and then grown at 30 ºC for 5 h before growing at 20 ºC 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 minutes. After this time, the lysate was centrifuged at 18,213 *g* at 4 ºC for 30 minutes. Lysate was denatured as described in Section 2.5 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 10,000x 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% $H_2O_2$ with 1 mM pA-Phe or pA-Pyr. The reaction mixture was incubated for 3 h at 25 ºC with 10 µM purified *N*-oxygenase, following which the reaction was terminated with addition of 0.8% TFA. Following 1h incubation, the mixture was centrifuged, and supernatant was analyzed by HPLC. pN-Phe production was further confirmed via UPLC-MS.

nsAA incorporation assays

Orthogonal AARS/tRNA pairs were cloned within pEVOL plasmids and transformed into a *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). These

transformed strains were cultured at 37 ºC in 300 µL MOPS EZ Rich media with aromatic amino acid (Phe, Trp, Tyr) dropout in deep 96-well plates with specified concentration of nsAA, 34 µg/mL chloramphenicol, and 50 µg/mL kanamycin with shaking at 1000 RPM and an orbital radius of 3 mm. At mid-exponential growth (OD ~0.5), 0.2 nM ATC and 0.2% (wt/v) 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 ºC before pelleting them via centrifugation. To eliminate possible fluorescence or absorbance via free nsAAs 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 $OD_{600}$ using a Spectramax i3x platereader with Softmax Pro 7.0.3 software. For each synthetase, we performed this screen in the presence and absence of externally supplied nsAA. 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 co-transformed with pEVOL-TetRSC11, 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 ºC in 50 mL of LB broth in 250 mL baffled shake flasks with 0.2% (wt/v) arabinose, 1 mM pA-Phe, 25.5 µg/mL chloramphenicol, 37.5 µg/mL kanamycin, 71.3 µg/mL streptomycin, and 0.2 nM ATC in a shaking incubator at 250 RPM. At an $OD_{600}$ 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 ºC for an additional 18 h. The reporter protein was then lysed and purified using FPLC with a His-Trap 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 µM as opposed to 200 µM) in 250 mL baffled shake flasks in a shaking incubator at 250 RPM and 37ºC. At an $OD_{600}$ of approximately 0.5, 1 mM IPTG was added to induce the pN-Phe synthesis plasmid. Cultures were grown at 30ºC for 2h, following which time, the reporter and AARS were induced with 0.2 nM ATC and 0.2% (wt/v) arabinose (if applicable). Cultures were then grown at 30 ºC 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 high-performance liquid chromatography (HPLC) using an Agilent 1260 Infinity model equipped with a Zorbax Eclipse Plus-C18 column (part number: 959701-902, 5 µm, 95Å, 2.1 x 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, 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 (**Fig. S1**). HPLC chromatograms were collected and analyzed using Agilent ChemStation (Version C.01.10) or Agilent OpenLab (Version 2.3) software.

Mass Spectrometry

Mass spectrometry (MS) measurements for small molecule metabolites were submitted to a Waters Acquity UPLC H-Class coupled to a single quadrupole mass detector 2 (SQD2) with an electrospray ionization source. Metabolite compounds were analyzed using a Waters Cortecs UPLC C18 column (part number: 186007092, 1.6 µm, 90Å, 2.1 x 30 mm) with an initial mobile phase of solvent A/B = 95/5 (solvent

A, water, 0.1% formic acid; solvent B, acetonitrile, 0.1% formic acid) with a 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 number: 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, 0.1% trifluoroacetic acid; solvent B, acetonitrile, 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, equilibration at 95/5 for 25-27 min. Flow rate was 1 mL/min 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 number: 186004495, 1.7 µm, 300Å, 2.1 x 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, 0.1% formic acid; solvent B, acetonitrile, 0.1% formic acid) held at 85/15 for 1 minute 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 source temperature of 150 $^{o}$C. Spectrum was analyzed from m/z 500 to 2000 and the spectra was 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 37000 to 38000 Da (resolution: 0.1 Da/channel) at an elution time of 3.2 minutes.

For LC-MS/MS measurements, protein samples (20 µg/lane) were first submitted to SDS-PAGE. The gel was then stained with Bio-Safe$^{TM}$ Coomassie stain (Bio-Rad), and de-stained 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 Scientific: p/n 89871). Following digestion, extracted peptides were then desalted using Pierce™ pipette tips (ThermoFisher) and dried, resuspended in 20 µL of 0.5% acetic acid (pH4.5) and then subjected to LC-MS/MS using an Thermo Scientific Orbitrap Eclipse Tribrid Mass Spectrometer (MS; Thermo Scientific) with an Ultimate 3000 nano-liquid chromatography (nano-LC) and a FAIMS Pro Interface (Thermo Scientific). The LC-MS/MS analysis was performed using an Orbitrap Eclipse MS (Thermo Scientific). Peptides were first loaded onto a trap column (PepMap C18) and then separated by an analytical column (PepMap C18, 2.0 um; 15 cm x 75mm I.D.; Thermo 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% to 10% in 8 min; then to 25% buffer B over 117 min; 25% to 32% buffer B in 10 min, then to 95% buffer B over 3 min; back to 1% B in 5 min, and stay equilibration at 1% 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 375-1500 m/z at a resolution of 60,000 in the Orbitrap. The maximum injection time was set to Dynamic, and 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 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 was 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 parameters[47]. An E. coli reference proteome (strain BL21-DE3; taxonomy_id:469008) was used for 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 7 amino acids. The false-discovery rate 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. 19**), additional analysis was performed using Proteome Discoverer software (version 3.0; Thermo Fisher). Raw data was searched against the *E. coli* reference proteome (strain BL21(DE3)) with NCBI reference Proteome (469008) using Sequest HT Processor and 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 (PSM) based on q value was done using Percolator, with target false-discovery rates (FDR) of 1% and 5% for stringent and relaxed validations, respectively. The false-discovery rate 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 Sequence Similarity Network (SSN)

Using NCBI BLAST, the 1000 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, AzoC[48], and ObiL. After deleting duplicate sequences, 2134 unique sequences were obtained which were then submitted to the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST)[49] to generate a sequence similarity network (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 (PvfB[50], PsAAO[51], CmlI[29], AzoC[48], AurF(RJ)[52], AurF[37], AlmD[53], HamC[54], and BezJ[55]) which were then colored to indicate prior work.

**Data Availability Statement**

The datasets generated during and/or analyzed during the current study are contained in the published article (and its Supplementary Information), are publicly accessible via cited repositories, or are available from the corresponding author on reasonable request.

**Methods-only References**


44. Cui, L. & Shearwin, K. E. Clonetegration using OSIP plasmids: One-step DNA assembly and site-specific genomic integration in bacteria. in *Methods in Molecular Biology* vol. 1472 139–155 (Humana Press, New York, NY, 2017).

45. Wannier, T. M. *et al.* Improved bacterial recombineering by parallelized protein discovery. *Proc Natl Acad Sci U S A* **117**, 13689–13698 (2020).

46. Kunjapur, A. M., Hyun, J. C. & Prather, K. L. J. Deregulation of S-adenosylmethionine biosynthesis and regeneration improves methylation in the *E. coli de novo* vanillin biosynthesis pathway. *Microb Cell Fact* **15**, 1–17 (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**, 1–9 (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 Proteins Proteom* **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 RHAI*. *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. *Nature Communications 2018 9:1* **9**, 1–13 (2018).

55. Tsutsumi, H. *et al.* Unprecedented cyclization catalyzed by a cytochrome P450 in benzastatin biosynthesis. *J Am Chem Soc* **140**, 6631–6639 (2018).