Improved pyrrolysine biosynthesis through continuous directed evolution of the complete pathway

Joanne Ho
Rice University https://orcid.org/0000-0002-5512-6992
Corwin Miller
Rice University
Jacob Mattia
Rice University
Matthew Bennett (matthew.bennett@rice.edu)
Rice University https://orcid.org/0000-0002-4975-8854

Article

Keywords: Pyrrolysine, Pyl, Pyl biosynthesis, Escherichia coli, continuous directed evolution

Posted Date: September 21st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-78440/v1

Version of Record: A version of this preprint was published at Nature Communications on June 24th, 2021. See the published version at https://doi.org/10.1038/s41467-021-24183-9.

Improved pyrrolysine biosynthesis through continuous directed evolution of the complete pathway

Joanne M. L. Ho[1,*], Corwin A. Miller[1,*], Jacob R. Mattia[1], and Matthew R. Bennett[1, 2,†]

[1] Department of Biosciences, Rice University MS-140, 6100 Main St., Houston, TX 77005, USA

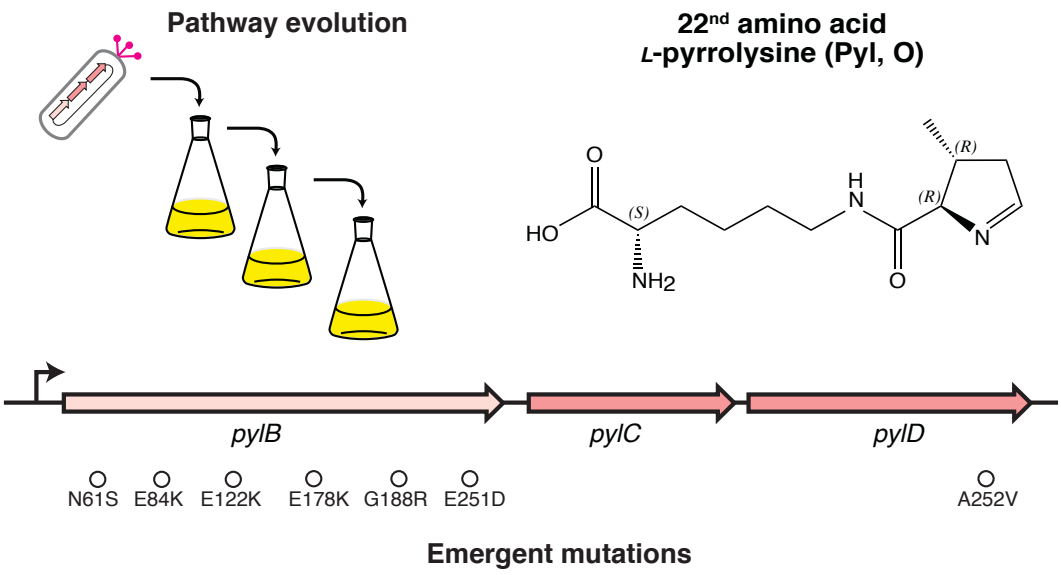
[2] Department of Bioengineering, Rice University MS-140, 6100 Main St. Houston, TX 77005, USA

* These authors contributed equally.

† Correspondence: matthew.bennett@rice.edu

Abstract

Pyrrolysine (Pyl, O) exists in nature as the 22nd proteinogenic amino acid. Despite being a fundamental building block of proteins, studies of Pyl have been hindered by the difficulty and inefficiency of both its chemical and biological syntheses. Here, we improved Pyl biosynthesis via rational engineering and directed evolution of the entire biosynthetic pathway. To accommodate toxicity of Pyl biosynthetic genes in *Escherichia coli*, we devised an approach termed Alternating Phage Assisted Non-Continuous Evolution (Alt-PANCE) that alternates mutagenic and selective phage growths. The evolved pathway exhibited a 32-fold improved yield of Pyl-containing super-folder green fluorescent protein (sfGFP) compared to the rationally engineered ancestor, whereas the WT pathway produced no detectable quantities of Pyl-containing sfGFP. This study demonstrates that Alt-PANCE provides a general approach for evolving proteins exhibiting toxic side effects, and further provides an improved pathway capable of producing substantially greater quantities of Pyl-proteins in *E. coli*.



Graphical Abstract Figure

Main text

Introduction

Pyrrolysine (Pyl, O) exists in nature as the 22nd proteinogenic amino acid[1]. Pyl represents an ancient addition to the genetic code, believed to have been present in the last universal common ancestor[2]. Today, Pyl is found in numerous bacterial and archaeal species but not eukaryotes. Although Pyl has been found in several classes of proteins[3], it is best known for its essential role in a unique class of methanogenic enzymes[1, 4]. Pyl has a remarkably distinct structure compared to other proteinogenic amino acids and is noteworthy for its reactive electrophilic moiety[1]—a feature absent in all other proteinogenic amino acids.

The genetic components required for Pyl incorporation are encoded in a single operon, *pylSTBCD*[4], which mediates Pyl biosynthesis and protein incorporation through nonsense suppression of amber (UAG) codons[5]. Within the operon, *pylS* encodes pyrrolysyl-tRNA synthetase (PylRS), which catalyzes the ligation of Pyl to tRNA, while *pylT* encodes the corresponding transfer RNA ($tRNA^{Pyl}$)[4]. Genes *pylB*, *pylC*, and *pylD* encode enzymes that biosynthesize pyrrolysine from lysine (Figure 1A)[4]. To date, numerous genetic code expansion studies have utilized PylRS and $tRNA^{Pyl}$ to incorporate synthetic amino acids into proteins, as these genes provide an aminoacyl tRNA synthetase (aaRS)-tRNA pair that does not exhibit cross-reactivity with the existing *E. coli* translation system[6-10].

Despite being a fundamental building block of proteins in nature, studies of Pyl proteins have been hindered by the poor supply of the amino acid. In contrast to most synthetic amino acids, Pyl is naturally recognized by PylRS and is ligated to $tRNA^{Pyl}$ with high efficiency. Improving production of Pyl proteins provides an unusual challenge, as genetic code expansion studies have typically focused on improving aaRS-tRNA pairs for better recognition of synthetic amino acids. To date, Pyl protein production has been severely limited by the poor activity of the archaeal biosynthetic *pylBCD* pathway. When this pathway is expressed heterologously in laboratory strains (such as *E. coli*), Pyl proteins are produced at a very low yield[4, 11]. Supplying cells with an exogenous source of

synthetically produced Pyl provides an alternative solution[12]. However, organic synthesis of Pyl is known for its difficulty[12-14] and thus remains commercially unavailable.

We expected that Phage Assisted Non-Continuous Evolution (PANCE[15]) would provide an effective method for improving production of Pyl proteins in *E. coli*, as Pyl biosynthesis is conducted by three genes across which problematic regions are difficult to identify. In phage-assisted evolution, the activity of an evolving gene of interest is linked to the life cycle of M13 bacteriophage, allowing each generation of phage growth to effectively serve as a cycle of directed evolution[16]. PANCE, which relies on serial flask transfers, and its chemostat-based counterpart PACE have previously been used to rapidly evolve increased activity in a wide variety of individual enzymes, including RNA polymerases[16], proteases[17], and aminoacyl-tRNA synthetases[18]. Of note, PANCE is categorized as an *in vivo* continuous directed evolution method[19]; 'non-continuous' within its name denotes its use of serial culture transfers, in contrast to the continuous flow machinery used for PACE[16]. Like other methods of continuous directed evolution[20], PANCE allows an entire genetic region to be quickly evolved without focusing on specific regions of interest.

Here, we detail the improvement of the *pylBCD* pathway for increased production of Pyl proteins in *E. coli*, performed using a two-step process. Our first step entailed the rational addition of a solubility tag to *pylB*, resulting in reduced toxic protein aggregation within the cell and also facilitating detectable levels of Pyl-containing sfGFP production. We next devised a version of PANCE that we term "Alternating Phage Assisted Non-Continuous Evolution" (Alt-PANCE), designed to reduce cellular toxicity during evolution. We used this method to evolve *pylBCD* for increased activity across numerous selection conditions. This process resulted in an additional 32-fold increase in Pyl-sfGFP production mediated by our most active mutant. Our evolutionary characterization found that the majority of mutations occured within *pylB*, and served to increase cellular production of this protein by ~6-fold. This work provides both a new procedure to enable continuous directed evolution of proteins exhibiting toxic side effects, and further provides a substantially improved biosynthetic pathway for bacterial production of Pyl proteins.

Results

Devising Alt-PANCE and improving PylB solubility

We initially attempted to use PANCE to evolve a codon-optimized variant of the *M. acetivorans pylBCD* pathway, and the poor initial activity of these genes led us to perform additional optimization before beginning evolution. Following overexpression of the *pylBCD* pathway, we observed formation of inclusion bodies within each cell (Figure S1). After noting that cells expressing only *pylCD* did not form inclusion bodies, we rationally fused a SUMO tag to the N-terminus of PylB to improve its solubility[21]. The addition of a SUMO tag has previously been shown to improve PylB solubility, enabling purification and crystallization of this protein[22]. Following the addition of a SUMO tag to *pylB*, we observed that expression of SUMO-*pylBCD* resulted in healthy cells without inclusion bodies, indicating improved PylB solubility *in vivo* and reduced toxic side effects (Figure S1). We next cloned SUMO-*pylBCD* into an M13 selection phage (SP) vector, termed SP.BCD (see Methods).

As expression of SUMO-*pylBCD* still exhibited a moderate toxic effect on *E. coli* cells, we next developed an alternating version of PANCE[15] that we termed Alt-PANCE (Figure 1C) to enable evolution of this pathway for improved activity. Typically, PANCE exposes evolving phage to simultaneous selection and mutagenesis, both of which lead to a high fitness cost and reduced phage titers. We further observed that expression of *pylBCD* in *E. coli* results in an additional fitness cost, the cumulative effect of which precludes simultaneous selection and mutagenesis. The Alt-PANCE procedure was thus developed to reduce the fitness cost associated with continuous evolution of mild to moderately toxic genes. Further, the Alt-PANCE approach can be used to simultaneously co-evolve multiple genes, in this case the entire Pyl biosynthetic pathway.

To mediate genetic selection, we expressed *pylS* and *pylT* within cells to link Pyl production to translation of an amber mutant of the essential phage gene, *gIII*[16]. Following Pyl biosynthesis, PylRS ligates Pyl to $tRNA^{Pyl}$, which leads to Pyl incorporation at 1–3 amber codons within *gIII* and expression of functional PIII (Figure 1B). Besides mediating selection, covalent ligation of Pyl to

$tRNA^{Pyl}$ also limits cell-to-cell diffusion of Pyl, which facilitates evolution by reducing evolutionary "cheating". A similar selection system was used in previous work to evolve *pylS* for improved incorporation of a Pyl analog—*N*ε-Boc-L-lysine (BocK)[15, 18]. Like Pyl, this synthetic amino acid mediates amber suppression[15].

Continuous directed evolution of Pyl biosynthesis pathway

Following our engineered improvement of PylB solubility, the *pylBCD* pathway exhibited sufficient activity to initiate Alt-PANCE. We began evolution at low selection stringency by supplementing the media with a starting concentration of BocK (200 μM) high enough to ease phage propagation yet low enough that phage variants producing more Pyl had a selective advantage (Figure S2B). During selection growths, we began by using accessory plasmid (AP) JH61, which encodes *gIII* containing one amber codon to facilitate Pyl insertion and also encodes a highly expressed PylRS (Table S1). During mutagenic growths, phage were propagated using strain S1059[23], a permissive strain that expresses *gIII* following phage infection without imposing a selection. Mutagenic growths also included the presence of mutagenesis plasmid MP6, previously shown to increase mutation rate by five orders of magnitude following arabinose induction[24]. SP.BCD was capable of separately propagating across both the selective and mutagenic conditions described above, thereby enabling Alt-PANCE initiation.

We performed Alt-PANCE of SP.BCD across three independent lineages, termed lineages A, B, and C. As phage continued to evolve, we steadily increased selection strength by varying four separate parameters, growing cells across a total of 11 stringency conditions (Table S1). After each round of Alt-PANCE, we tested phage growth under stronger selection conditions; we then either maintained selection strength at the prior level in the subsequent round, or increased it if growth was possible. Initially, we increased selection strength by decreasing BocK supplementation, thereby forcing evolving phage to produce additional Pyl to fill the gap. Once the evolving phage could propagate in the absence of BocK, we increased selection strength by altering the AP vector via three

approaches: (1) we reduced expression of *pylS* by mutating its promoter and ribosome binding site (RBS), (2) we increased the number of amber codons in *gIII*, and (3) we substituted mutant *pylS* variants with reduced Pyl affinity. Each of these approaches required evolving phage to produce greater amounts of Pyl in order to maintain comparable levels of *gIII* translation. We halted evolution upon reaching the final stringency condition after 34–40 rounds of Alt-PANCE (Figures S2C and S3).

Analysis of evolved *pylBCD* mutations

For each lineage, we isolated 10-15 phage plaques; for each isolate, we sequenced the *pylBCD* insert, its upstream promoter, and ribosome binding sites. We identified 5–8 mutations in each lineage (Figure 2A; Table S2), 11 convergent high-frequency mutations (Figure 2A), and a total of 16 unique mutations (Table S2). All mutations were found within protein coding sequences and the majority were distal to the active site of each enzyme (Figure 2B). We identified six distinct subpopulations—36A_sub-pop1, 36A_sub-pop2, 34B_sub-pop3, 34B_sub-pop4, 34B_sub-pop5, and 40C_sub-pop6—that contained representative combinations of convergent mutations (Table S2). The *pylBCD* cassettes from these subpopulations were cloned into expression vectors to measure the activity of each variant via a coupled super-folder green fluorescent protein (sfGFP) reporter assay, wherein biosynthesized Pyl is incorporated into sfGFP though amber suppression (see Methods).

While most evolved mutants appeared to exhibit higher activity than the ancestral variant SUMO-*pylBCD*, the highest activity levels were observed in 36A_sub-pop1, 36A_sub-pop2, and 34B_sub-pop3 (Figure S4). Next, we rationally combined mutations originating from separate lineages to produce five combinatorial variants (Table S3) and observed that 3f2 and JM10.1 exhibited the highest activity (Figure S5). Although ancestral variant SUMO-*pylBCD* did not produce detectable sfGFP signal under other assay conditions, activity was observed when mediating suppression of three amber codons in *E. coli* strain C321.ΔA.exp[25] (see Methods). Under these conditions, variant 3f2 exhibited 4.9-fold greater activity compared to variant 36A_sub-pop2, and 32-fold greater activity compared to ancestral variant SUMO-*pylBCD* (Figure 3). The WT *pylBCD* variant

did not produce detectable sfGFP activity under any condition tested; however, luminescent signal was observed following expression of this variant in a separate coupled luciferase activity assay (Figure S6, see Methods). His-tagged sfGFP containing two amber codons was produced using variant 3f2, purified, and analyzed by LC-MS/MS to confirm the identity of Pyl at positions 39 (Figure S7 and Table S4) and 151 (Figure S8 and Table S5, see Methods).

Assessing the 16 unique mutations identified within the *pylBCD* operon at the endpoint, the majority (9) were found within the *pylB* coding sequence, and two of those were within the SUMO tag (Table S2). Each isolated subpopulation contained significantly more mutations in *pylB* than in *pylC* or *pylD*, with two subpopulations (36A_sub-pop1 and 40C_sub-pop6) containing no mutations in *pylC* or *pylD*. This disparity is also observed in the highly active combinatorial variants 3f2 and JM10.1. Those two variants each contained only a single mutation in *pylD*, no mutations in *pylC*, and either six or eight mutations in *pylB* (Table S3). These observations suggest that improved Pyl production primarily stemmed from mutations in *pylB*, consistent with prior biochemical evidence and quantum mechanical simulations indicating this enzyme catalyzes the rate-limiting step of Pyl synthesis[11, 26].

Characterization of PylB mutants

Notably, *pylB* mutations were remarkably convergent in character, resulting in increased cationic protein surface charge. Each Alt-PANCE evolved subpopulation contained mutations that increased the charge of SUMO-PylB by +2 to +5. Combinatorial variants 3f2 and JM10.1 exhibited an even greater change, with cationic charge increasing by +7 and +9, respectively. Mutations changing anionic glutamic acid to cationic lysine residues were prevalent, with four separate instances observed (E84K, E122K, E175K, and E178K). Analysis of the crystal structure of PylB from a closely related organism *Methanosarcina barkeri*[22] revealed that all four E-to-K mutations occurred at solvent-exposed regions of the protein surface (Figure 2B). This pattern suggests that, instead of directly improving catalytic properties of PylB, these mutations may confer a benefit towards the

biophysical properties of the protein, thereby increasing Pyl production by reducing protein turn-over within the cell.

To better characterize evolved mutations in PylB, we next overexpressed and purified two evolved SUMO-PylB variants (from 3f2 and JM10.1 cassettes) as well as the ancestral SUMO-PylB for further analysis (see Methods). Consistent with prior work, we were unable to measure activity from any purified PylB samples, owing to the extreme lability of its SAM cofactor[22]. As altered protein surface charge is associated with changes in solubility[27], and given that activity improvements were previously achieved through addition of a SUMO solubility tag, we performed precipitation experiments to evaluate the solubility of our purified PylB samples. While no precipitation was observed following exposure to the precipitant ammonium sulfate, exposure to PEG-8,000 induced aggregation in each PylB sample (see Methods). Noting that protein surface charge has also been associated with adaptation to different salt conditions[27], we performed separate precipitation experiments under both high salt (250 mM NaCl) and low salt (10 mM NaCl) conditions. Following exposure to varying amounts of PEG-8,000, both evolved mutants exhibited a greater propensity to aggregate compared to the SUMO-PylB ancestor, indicative of reduced solubility under both high and low salt conditions (Figure S9). We further confirmed that solubility is not improved within evolved mutants 3f2 and JM10.1 by cloning SUMO tag deletion variants, which produced inclusion bodies following pathway induction (Figure S10).

We next examined the effects of evolved PylB mutations on protein thermostability, as protein stability is a key determinant of steady-state protein concentration[28]. For these experiments we used differential scanning fluorimetry (DSF, also known as the Themofluor assay)[29], a method in which a fluorescent probe is used to monitor protein unfolding as samples are slowly heated (see Methods). This assay was performed using each of the aforementioned purified PylB samples, both under low and high salt conditions (Figure S11). Under both salt conditions, evolved mutants exhibited similar melting temperature (Tm) values to one another, while the ancestral SUMO-PylB exhibited a higher

Tm. Compared to low salt conditions, higher amounts of salt led to an increased Tm value for the ancestral SUMO-PylB variant, but did not significantly affect the evolved variants.

During purification, we noted that protein yields of both evolved PylB variants were were substantially greater than that of the ancestral SUMO-PylB (Figure S12). This result was confirmed upon repeating each purification in triplicate (Figure S13), with variants 3f2 and JM10.1 producing 6.0-fold and 5.6-fold greater yields compared to the ancestral variant, respectively. These findings indicate that the evolved mutants exhibit increased steady-state protein concentration compared to the ancestral variant, consistent with our expectation that these evolved mutations confer a benefit. Since PylB is an enzyme, such an increase in protein concentration will produce an even larger fold increase in product formation. Indeed, the ~6-fold increase in PylB concentration within evolved mutants may largely or entirely account for the ~32-fold increase in Pyl-sfGFP production (Figure 3).

Discussion

In light of our above results, we thus conclude that *pylBCD* mutations evolved during PANCE mediate improved Pyl protein production largely by increasing steady-state concentrations of PylB. As noted in the 'Results' section, evolved PylB mutations convergently mediated increased cationic protein surface charge, which provide a clue towards the mechanism mediating increased protein concentration. While prior studies have linked altered protein surface charge to both altered solubility and adaptation to different salt conditions[30-32], our biophysical studies of purified PylB variants found that evolved mutants exhibited reduced protein solubility and showed similar properties under both high and low salt conditions. Thermostability assays revealed a similar result, with evolved mutants showing lower stability compared to their ancestor under both high and low salt conditions. Thus, while the mechanism mediating increased cellular concentrations of evolved PylB mutants in *E. coli* remains unclear, our biophysical experiments nonetheless add to the existing understanding of the protein properties of PylB.

We note here that we also attempted multiple strategies to extract and quantify free Pyl following its biosynthesis (see Supplementary Methods). Although our approach was guided by prior work[33], we were unable to detect Pyl in our numerous extracts. While the reasons underpinning our inability to reproduce this protocol are unclear, we postulate that our efforts may have been complicated by chemical modification or degradation of the free amino acid within the complex cellular mileu. Indeed, similar instability is observed with the 21st amino acid selenocysteine, as this residue and its biosynthetic intermediates are highly labile; selenocysteine becomes stabilized only upon incorporation into proteins[34, 35].

The Alt-PANCE procedure developed here establishes a general methodology for rapidly evolving proteins for improved function while reducing toxic side effects. Our results indicate this method can improve protein stability within the producing cell by altering biophysical parameters that are difficult to address via rational or targeted methods[36]. While prior PACE and PANCE experiments have primarily evolved single proteins[19, 37], here we apply these techniques to a complete multi-gene biosynthetic pathway. A persistent challenge lay in diffusion of biosynthesized products between competing cells[37], which enables evolutionary escape. Our results demonstrate that ligation of the biosynthetic product to tRNA sufficiently reduces diffusion to enable successful continuous directed evolution. As numerous naturally occurring ncAAs cannot be biosynthesized *in E. coli* at high yield[38], our selection system could be applied to evolve other valuable amino acid biosynthetic pathways.

Our improved *pylBCD* pathway enables production of useful Pyl proteins at substantially greater yields. As the 22nd amino acid, Pyl is fundamental to the origins of life, yet has long been difficult to study. In nature, Pyl is a critical residue in methanogenic enzymes—*mttB*, *mtbB, mtmB*—that produce methane from methylamines[1]. Methane is increasingly viewed as a substitute for fossil fuels, but microbial methanogenesis is only known to occur in archaea[39], which is recalcitrant to genome engineering. Heterologous expression of these Pyl-containing methanogenic enzymes in *E. coli*, a chassis much easier to engineer than archaea, is now feasible and can provide a viable strategy toward the production of industrially relevant quantities of methane.

The unusual chemical properties of Pyl further enable other exciting bioengineering applications. While electrophilic moieties are notably absent in other proteinogenic amino acids[40], the imine group found within the Pyl pyrrole ring is both electrophilic and highly reactive. This moiety has been previously shown to react with 2-amino-benzaldehyde (2-ABA) and 2-amino-acetophenone (2-AAP) groups to form a bio-orthogonal crosslink[11], which provides a bioconjugation method with unique chemistry distinct from other approaches. The positive charge carried by Pyl also enables its use in antimicrobial peptides whose mechanisms of action require a cationic charge[41]. Additionally, considering the atypical structure of Pyl, peptides containing this residue are likely to resist proteolytic digestion[42]. Pyl is thus unusually well suited for use in antimicrobial peptides, as degradation remains the primary hurdle for their therapeutic application. This work achieves the facile overproduction of Pyl-containing proteins, enables use of this remarkable chemical in future studies, and provides insight on the biophysical properties of the biosynthetic enzymes that produce this understudied naturally-occurring non-canonical amino acid.

Acknowledgments

We are grateful to TuKiet Lam and Jean Kanyo at the Yale/NHLBI Proteomics Center for collecting and processing the LC-MS/MS data. We thank Yousif Shamoo (Rice) and Zachary Ball (Rice) for useful discussions related to this work. This work was funded by the National Institutes of Health and the National Sciences Foundation through the joint NSF-National Institutes of General Medical Sciences Mathematical Biology Program grant nos. R01GM104974 and DMS-1662290; the National Institutes of Health grant no. R01GM117138; and the Robert A. Welch Foundation grant no. C-1729 (M.R.B.).

Data Availability Statement

Raw data is available from the corresponding author upon request (Figs. 3, S2, S4-S6, S9, S10). Relevant plasmids (Table S6) will be made available through Addgene.

References

1. Rother, M. & Krzycki, J.A. Selenocysteine, Pyrrolysine, and the Unique Energy Metabolism of Methanogenic Archaea. *Archaea* (2010).
2. Nozawa, K. et al. Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163-1167 (2009).
3. Heinemann, I.U. et al. The appearance of pyrrolysine in tRNAHis guanylyltransferase by neutral evolution. *Proc Natl Acad Sci U S A* **106**, 21103-21108 (2009).
4. Gaston, M.A., Zhang, L., Green-Church, K.B. & Krzycki, J.A. The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. *Nature* **471**, 647-650 (2011).
5. Ambrogelly, A. et al. Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc Natl Acad Sci U S A* **104**, 3141-3146 (2007).
6. Wan, W., Tharp, J.M. & Liu, W.R. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Bba-Proteins Proteom* **1844**, 1059-1070 (2014).
7. Fan, C., Xiong, H., Reynolds, N.M. & Soll, D. Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res* **43**, e156 (2015).
8. Fischer, E.C. et al. New codons for efficient production of unnatural proteins in a semisynthetic organism. *Nat Chem Biol* (2020).
9. Willis, J.C.W. & Chin, J.W. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat Chem* **10**, 831-837 (2018).
10. Wang, Y.S., Fang, X., Wallace, A.L., Wu, B. & Liu, W.R. A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J Am Chem Soc* **134**, 2950-2953 (2012).
11. Ou, W.J. et al. Site-specific protein modifications through pyrroline-carboxy-lysine residues. *P Natl Acad Sci USA* **108**, 10437-10442 (2011).
12. Hao, B. et al. Reactivity and chemical synthesis of L-pyrrolysine - the 22(nd) genetically encoded amino acid. *Chem Biol* **11**, 1317-1324 (2004).
13. Wong, M.L., Guzei, I.A. & Kiessling, L.L. An asymmetric synthesis of L-pyrrolysine. *Org. Lett.* **14**, 1378-1381 (2012).
14. Han, M.Y. et al. A concise synthesis of L-pyrrolysine. *Chemistry* **19**, 8078-8081 (2013).
15. Suzuki, T. et al. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* **13**, 1261-1266 (2017).
16. Esvelt, K.M., Carlson, J.C. & Liu, D.R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-U550 (2011).
17. Packer, M.S., Rees, H.A. & Liu, D.R. Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nature Communications* **8** (2017).
18. Bryson, D.I. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253-1260 (2017).
19. Badran, A.H. & Liu, D.R. In vivo continuous directed evolution. *Curr Opin Chem Biol* **24**, 1-10 (2015).
20. d'Oelsnitz, S. & Ellington, A. Continuous directed evolution for strain and protein engineering. *Curr Opin Biotechnol* **53**, 158-163 (2018).
21. Marblestone, J.G. et al. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci* **15**, 182-189 (2006).
22. Quitterer, F., List, A., Eisenreich, W., Bacher, A. & Groll, M. Crystal Structure of Methylornithine Synthase (PylB): Insights into the Pyrrolysine Biosynthesis. *Angew Chem Int Edit* **51**, 1339-1342 (2012).
23. Hubbard, B.P. et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat Methods* **12**, 939-942 (2015).
24. Badran, A.H. & Liu, D.R. Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).
25. Lajoie, M.J. et al. Genomically recoded organisms expand biological functions. *Science* **342**, 357-360 (2013).

26. Zhu, W.Y., Liu, Y.J. & Zhang, R. QM/MM study of the conversion mechanism of lysine to methylornithine catalyzed by methylornithine synthase (PylB). *Theor Chem Acc* **132** (2013).
27. Kramer, R.M., Shende, V.R., Motl, N., Pace, C.N. & Scholtz, J.M. Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. *Biophys J* **102**, 1907-1915 (2012).
28. Yen, H.C., Xu, Q., Chou, D.M., Zhao, Z. & Elledge, S.J. Global protein stability profiling in mammalian cells. *Science* **322**, 918-923 (2008).
29. Vollrath, F., Hawkins, N., Porter, D., Holland, C. & Boulet-Audet, M. Differential Scanning Fluorimetry provides high throughput data on silk protein transitions. *Sci Rep* **4**, 5625 (2014).
30. Siglioccolo, A., Paiardini, A., Piscitelli, M. & Pascarella, S. Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Struct Biol* **11**, 50 (2011).
31. Britton, K.L. et al. Insights into the molecular basis of salt tolerance from the study of glutamate dehydrogenase from Halobacterium salinarum. *J Biol Chem* **273**, 9023-9030 (1998).
32. DasSarma, S. & DasSarma, P. Halophiles and their enzymes: negativity put to good use. *Curr Opin Microbiol* **25**, 120-126 (2015).
33. Longstaff, D.G. et al. A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1021-1026 (2007).
34. Gonzalez-Flores, J.N., Shetty, S.P., Dubey, A. & Copeland, P.R. The molecular biology of selenocysteine. *Biomol Concepts* **4**, 349-365 (2013).
35. Miller, C. et al. A synthetic tRNA for EF-Tu mediated selenocysteine incorporation in vivo and in vitro. *Febs Letters* **589**, 2194-2199 (2015).
36. Rigoldi, F., Donini, S., Redaelli, A., Parisini, E. & Gautieri, A. Review: Engineering of thermostable enzymes for industrial applications. *APL Bioeng* **2**, 011501 (2018).
37. Roth, T.B., Woolston, B.M., Stephanopoulos, G. & Liu, D.R. Phage-Assisted Evolution of Bacillus methanolicus Methanol Dehydrogenase 2. *Acs Synth Biol* **8**, 796-806 (2019).
38. Almhjell, P.J., Boville, C.E. & Arnold, F.H. Engineering enzymes for noncanonical amino acid synthesis. *Chem Soc Rev* **47**, 8980-8997 (2018).
39. Enzmann, F., Mayer, F., Rother, M. & Holtmann, D. Methanogens: biochemical background and biotechnological applications. *AMB Express* **8**, 1 (2018).
40. Gaston, M.A., Jiang, R. & Krzycki, J.A. Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr Opin Microbiol* **14**, 342-349 (2011).
41. Mahlapuu, M., Hakansson, J., Ringstad, L. & Bjorn, C. Antimicrobial Peptides : An Emerging Category of Therapeutic Agents. *Front Cell Infect Mi* **6** (2016).
42. Baumann, T. et al. Prospects of In vivo Incorporation of Non-canonical Amino Acids for the Chemical Diversification of Antimicrobial Peptides. *Frontiers in Microbiology* **8** (2017).

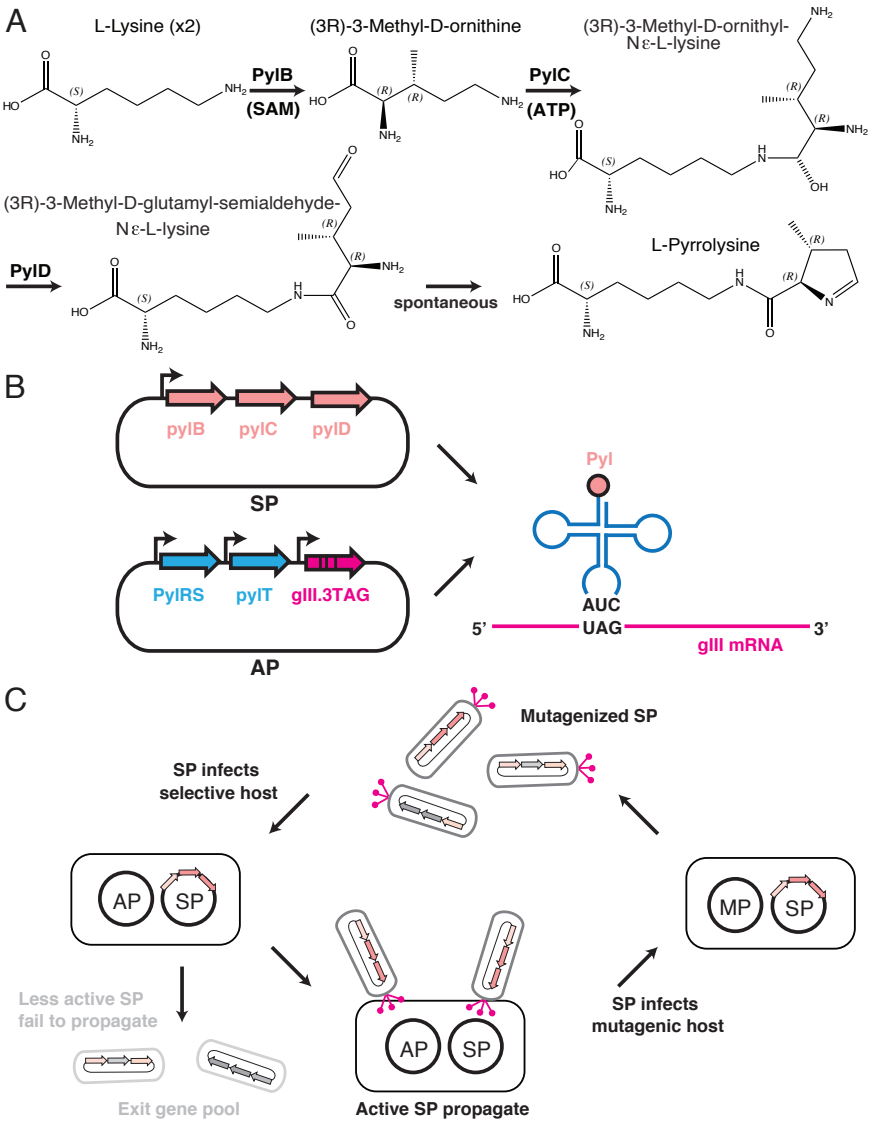


Figure 1. Pyl biosynthetic pathway evolution via Alt-PANCE. (A) Pyl biosynthesis involves condensation of two lysine molecules by three enzymes—the radical SAM enzyme PylB, ATP-dependent PylC, and PylD. (B) Biosynthetic *pylBCD* operon was cloned into a selection phage (SP), while the constitutively expressed *pylST* and phage shock promoter-controlled *gIII* were cloned into Accessory Plasmid (AP) vectors. (C) Each Alt-PANCE round entailed two phage passages: Selective passage (left) entails operon activity-dependent expression of PIII (pink rods with circular tips); Mutagenic passage (right) entails Mutagenesis Plasmid (MP)-dependent mutagenesis of the SP.

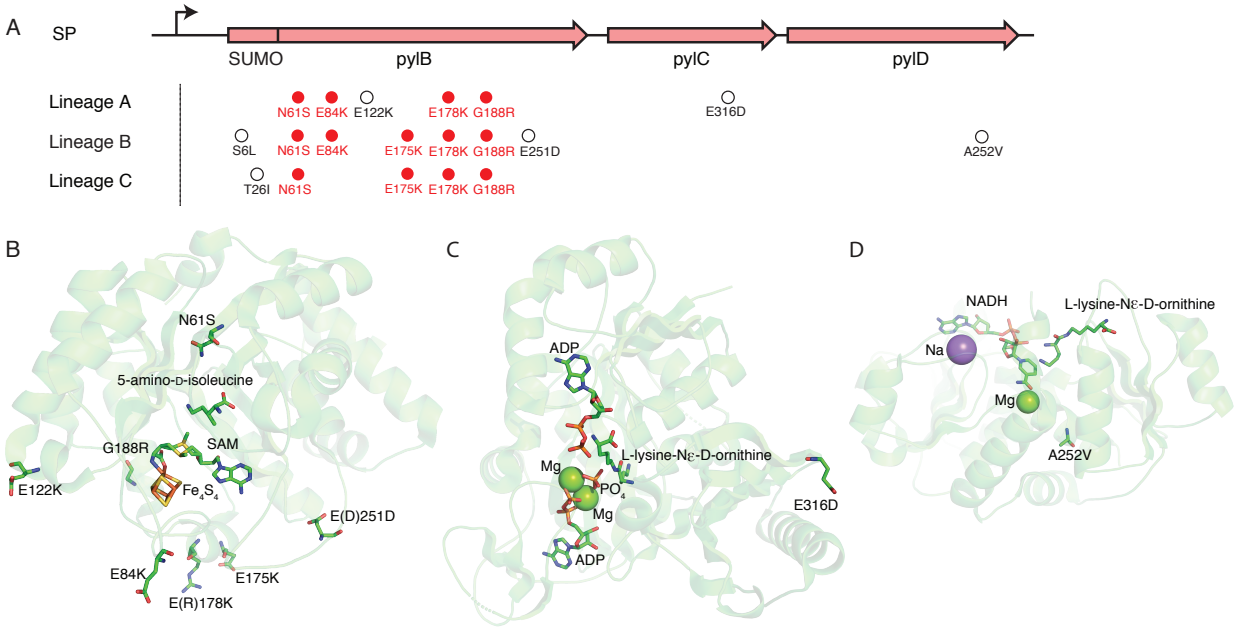


Figure 2. (A) Convergent mutations in the co-evolved *pylBCD* genes. Locations of critical mutations in the previously solved structures of *M. barkeri* (B) PylB (PDB: 3T7V), (C) PylC (PDB: 4FFP), and (D) PylD (PDB: 4J4B); amino acid substitutions were all found outside the enzyme active sites, with many PylB mutations resulting in increased cationic surface charge. Numbering scheme is based on the wild-type protein sequences of PylB, PylC, and PylD in *Methanosarcina acetivorans*. Specifically, numerical labels for PylB refer to the WT PylB protein sequence and do not include the 104 amino acids added by the SUMO tag. Brackets indicate residues naturally occurring in *M. barkeri* proteins that have diverged from the *M. acetivorans* sequence.

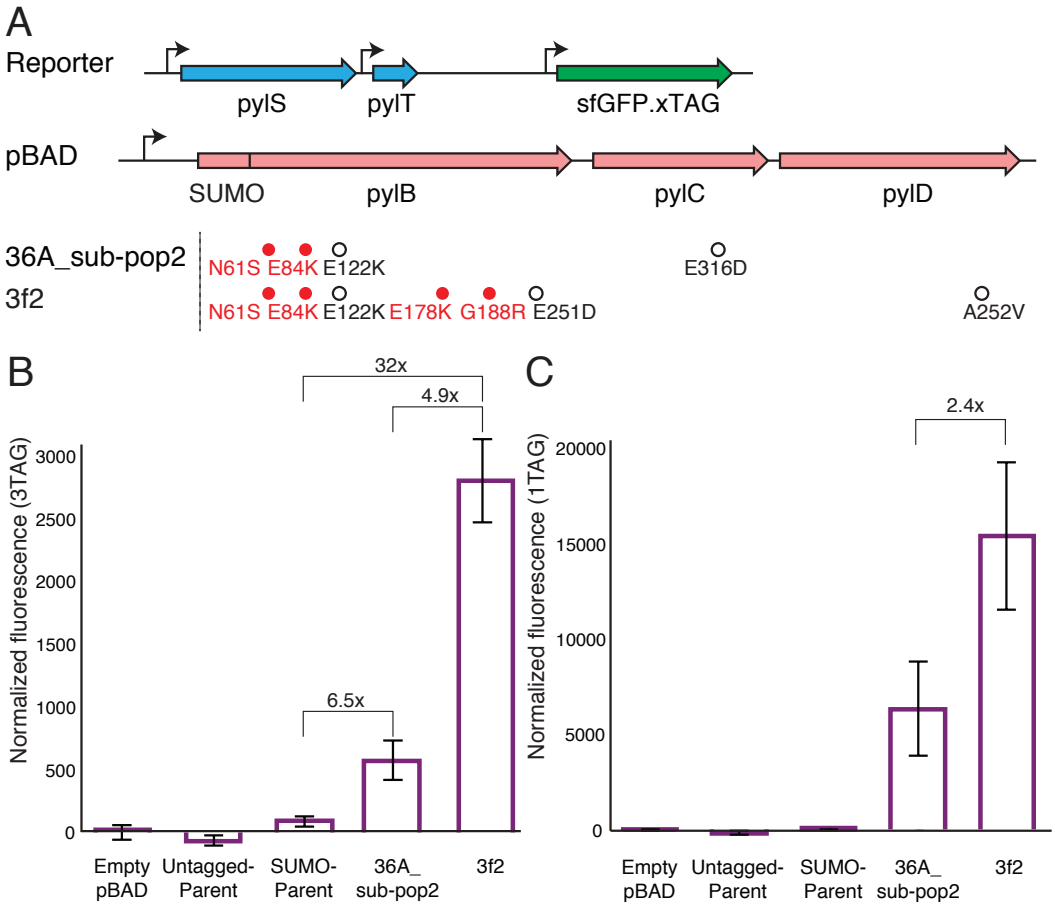


Figure 3. Evolved biosynthetic pathway mediates improved Pyl-protein yield. (A) The best evolved variant (36A_sub-pop2) and best combinatorial variant (3f2) contained a total of 4 and 7 mutations, of which 2 and 4 were convergent mutations (red), respectively. (B) When tested in E. coli strain C321.ΔA.exp, read-through of three amber codons in sfGFP is improved 32-fold in variant 3f2 combinatorial variant and 6.5-fold in variant 36A_sub-pop2, relative to the parental variant SUMO-*pylBCD*. The WT *pylBCD* pathway (termed 'Untagged Parent') did not have detectable activity. (C) Read-through of one amber codon in sfGFP is improved 2.4-fold in variant 3f2, relative to variant 36A_sub-pop2. The WT *pylBCD* variant and the SUMO-*pylBCD* did not have detectable activity in this assay. In panels B and C, fluorescence intensity shown (excitation 488 nm, emission 509 nm) was normalized by optical density (A_{600}) (see Methods). Samples were tested in biological triplicate; data shown represents mean values ± s.d.

Figures

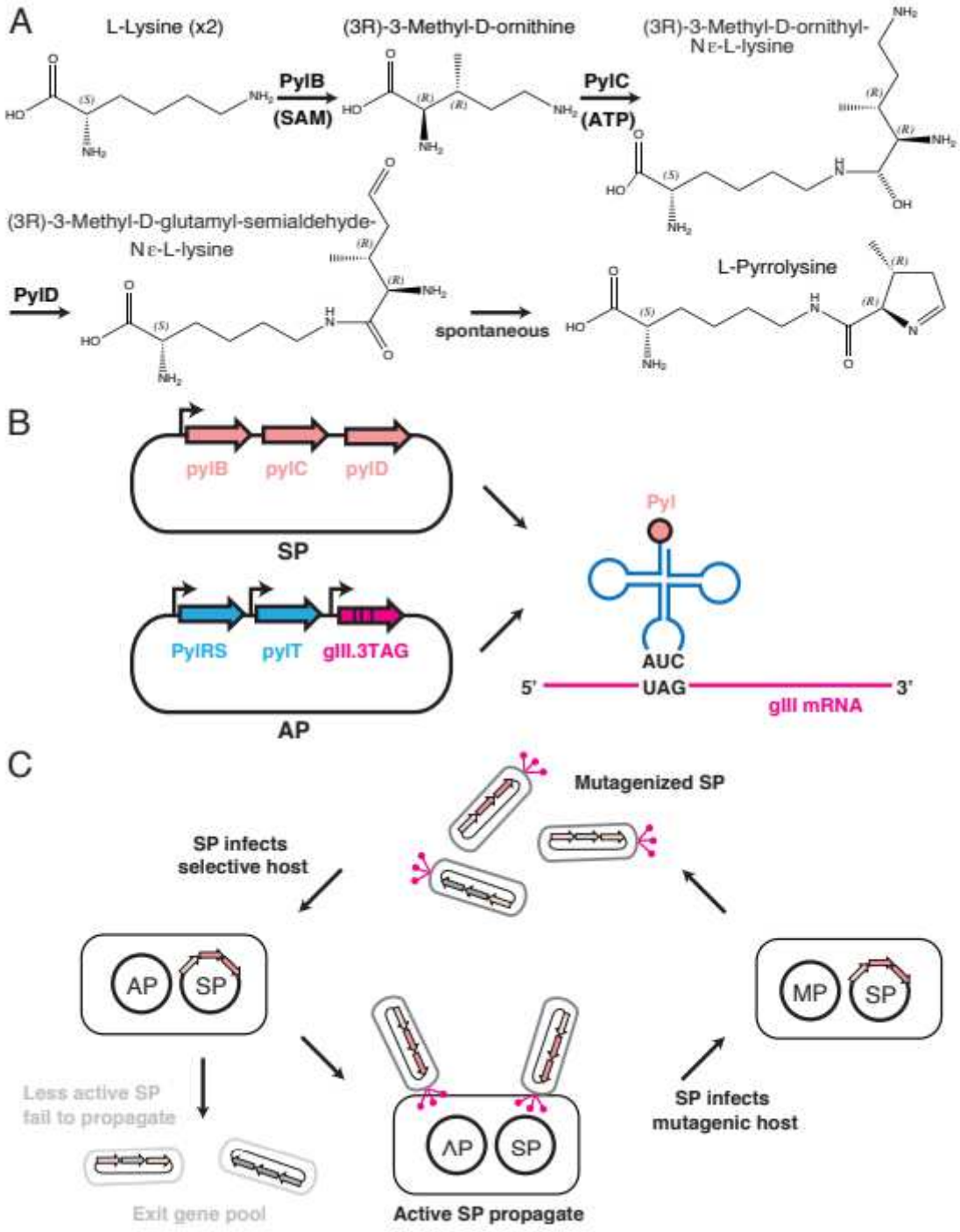


Figure 1

Pyl biosynthetic pathway evolution via Alt-PANCE. (A) Pyl biosynthesis involves condensation of two lysine molecules by three enzymes—the radical SAM enzyme PylB, ATPdependent PylC, and PylD. (B) Biosynthetic pylBCD operon was cloned into a selection phage (SP), while the constitutively expressed pylST and phage shock promoter-controlled gIII were cloned into Accessory Plasmid (AP) vectors. (C)

Each Alt-PANCE round entailed two phage passages: Selective passage (left) entails operon activity-dependent expression of PIII (pink rods with circular tips); Mutagenic passage (right) entails Mutagenesis Plasmid (MP)-dependent mutagenesis of the SP.

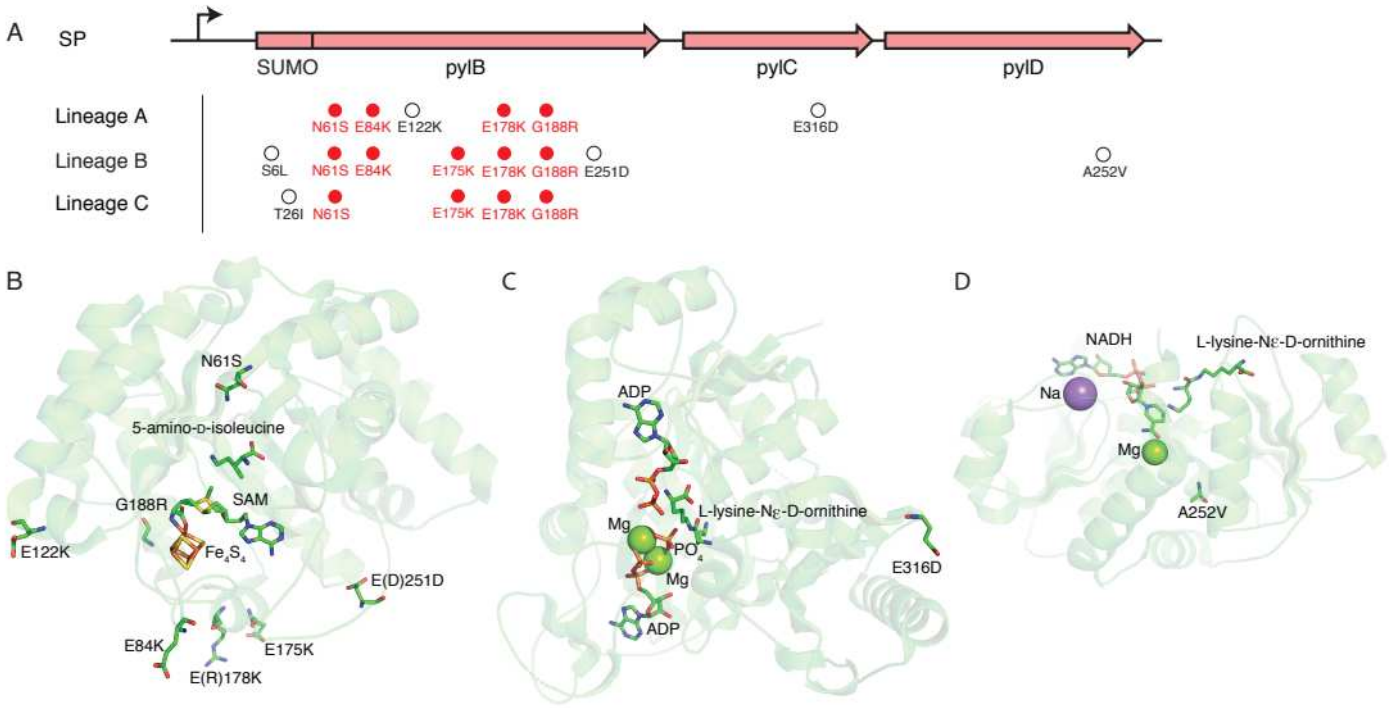


Figure 2

(A) Convergent mutations in the co-evolved pylBCD genes. Locations of critical mutations in the previously solved structures of M. barkeri (B) PylB (PDB: 3T7V), (C) PylC (PDB: 4FFP), and (D) PylD (PDB: 4J4B); amino acid substitutions were all found outside the enzyme active sites, with many PylB mutations resulting in increased cationic surface charge. Numbering scheme is based on the wild-type protein sequences of PylB, PylC, and PylD in Methanosarcina acetivorans. Specifically, numerical labels for PylB refer to the WT PylB protein sequence and do not include the 104 amino acids added by the SUMO tag. Brackets indicate residues naturally occurring in M. barkeri proteins that have diverged from the M. acetivorans sequence.

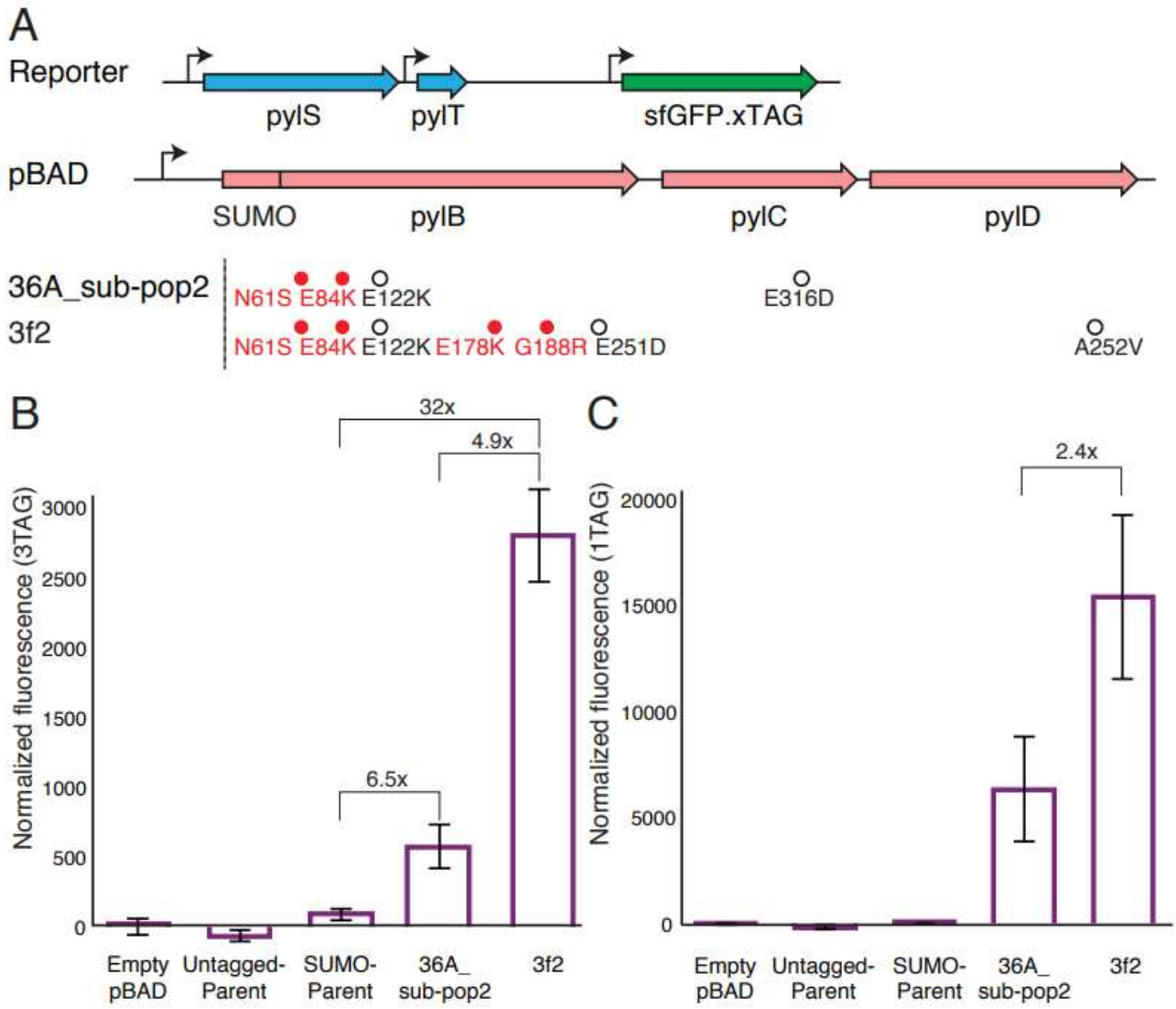


Figure 3

Evolved biosynthetic pathway mediates improved Pyl-protein yield. (A) The best evolved variant (36A_sub-pop2) and best combinatorial variant (3f2) contained a total of 4 and 7 mutations, of which 2 and 4 were convergent mutations (red), respectively. (B) When tested in E. coli strain C321.ΔA.exp, read-through of three amber codons in sfGFP is improved 32-fold in variant 3f2 combinatorial variant and 6.5-fold in variant 36A_sub-pop2, relative to the parental variant SUMOpylBCD. The WT pylBCD pathway (termed 'Untagged Parent') did not have detectable activity. (C) Read-through of one amber codon in sfGFP is improved 2.4-fold in variant 3f2, relative to variant 36A_sub-pop2. The WT pylBCD variant and the SUMO-pylBCD did not have detectable activity in this assay. In panels B and C, fluorescence intensity shown (excitation 488 nm, emission 509 nm) was normalized by optical density (A600) (see Methods). Samples were tested in biological triplicate; data shown represents mean values ± s.d.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- NCommsSupplementaryInformation.pdf