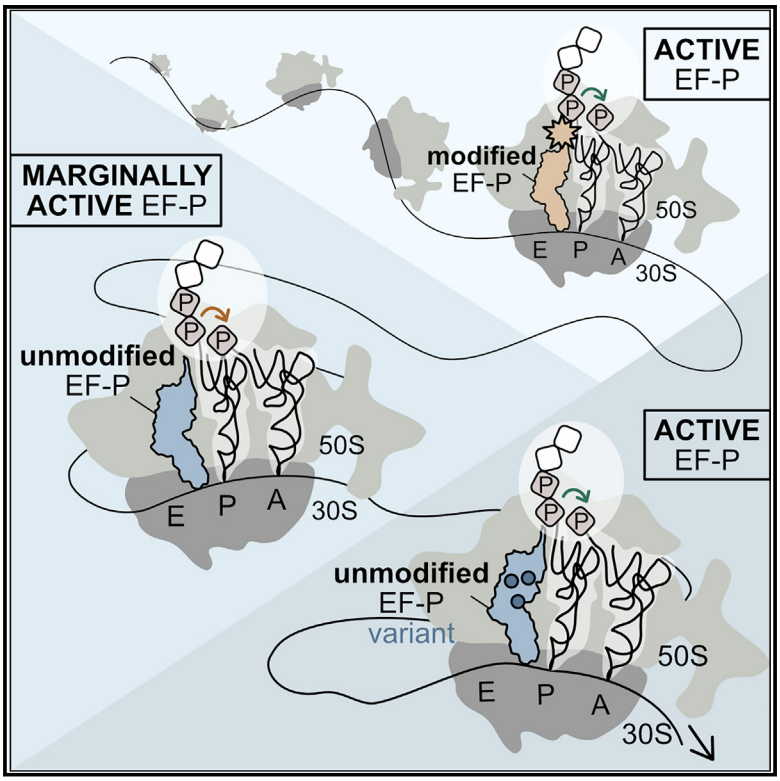
Cell Reports

Article

Decrypting the functional design of unmodified translation elongation factor P

Graphical abstract



Highlights

- The naturally unmodified EF-P of *Rhodomicrobium vannielii* is active in *E. coli*
- Three amino acid positions are crucial to maintain the functionality of unmodified EF-P
- Marginally active EF-Ps can be engineered in fully functional unmodified variants in *E. coli*
- A heterologous expression platform for proteins with polyproline motifs is available

Authors

Urte Tomasiunaite, Pavel Kielkowski, Ralph Krafczyk, Ignasi Forné, Axel Imhof, Kirsten Jung

Correspondence

jung@lmu.de

In brief

Tomasiunaite et al. provide insights into the functional design of unmodified translation elongation factor P. They identify critical residues that determine the functionality of unmodified EF-Ps in *E. coli*. The results indicate that the activity of EF-P is not exclusively dependent on the post-translational modification.

Tomasiunaite et al., 2024, Cell Reports *43*, 114063
May 28, 2024
https://doi.org/10.1016/j.celrep.2024.114063

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Decrypting the functional design of unmodified translation elongation factor P

Urte Tomasiunaite,[1] Pavel Kielkowski,[2] Ralph Krafczyk,[1] Ignasi Forné,[3] Axel Imhof,[3] and Kirsten Jung[1,4,*]
[1]Faculty of Biology, Microbiology, Ludwig-Maximilians-Universität München, 82152 Martinsried, Germany
[2]Department of Chemistry, Institut für Chemische Epigenetik (ICEM), Ludwig-Maximilians-Universität München, 81375 Munich, Germany
[3]Zentrallabor für Proteinanalytik, Biomedical Center Munich, Ludwig-Maximilians-Universität München, 82152 Martinsried, Germany
[4]Lead contact
*Correspondence: jung@lmu.de
https://doi.org/10.1016/j.celrep.2024.114063

SUMMARY

Bacteria overcome ribosome stalling by employing translation elongation factor P (EF-P), which requires post-translational modification (PTM) for its full activity. However, EF-Ps of the PGKGP subfamily are unmodified. The mechanism behind the ability to avoid PTM while retaining active EF-P requires further examination. Here, we investigate the design principles governing the functionality of unmodified EF-Ps in *Escherichia coli*. We screen for naturally unmodified EF-Ps with activity in *E. coli* and discover that the EF-P from *Rhodomicrobium vannielii* rescues growth defects of a mutant lacking the modification enzyme EF-P-(*R*)-β-lysine ligase. We identify amino acids in unmodified EF-P that modulate its activity. Ultimately, we find that substitution of these amino acids in other marginally active EF-Ps of the PGKGP subfamily leads to fully functional variants in *E. coli*. These results provide strategies to improve heterologous expression of proteins with polyproline motifs in *E. coli* and give insights into cellular adaptations to optimize protein synthesis.

INTRODUCTION

The translation of mRNAs to proteins at the ribosome is a highly conserved process in all domains of life. The translational efficiency is dynamic and depends on codon bias, tRNA accessibility, and the chemical nature of amino acids to be incorporated into the polypeptide chain.[1] Among the canonical amino acids, proline is an imino acid. This imino group cannot be optimally oriented for nucleophilic attack onto the carbonyl carbon atom of the P-site substrate during the transpeptidation reaction. In addition, proline is a poor donor during protein synthesis due to the rigidity of the pyrrolidine ring.[2–5] Consequently, the presence of consecutive proline codons leads to a slowdown in protein synthesis and can cause ribosome stalling.[6–9] Despite the hindrance, polyproline motifs play essential roles in the catalytic activity of enzymes, contribute to protein-protein interactions, and regulate protein copy numbers,[7,10–13] rendering them indispensable components of the proteome of all organisms.

To facilitate the synthesis of proteins containing polyproline motifs, organisms from all kingdoms have evolved specialized translation factors. In bacteria, elongation factor P (EF-P) plays a crucial role in this process,[6,7] while eukaryotes and archaea use initiation factor 5A (e/aIF5A).[14] These translation factors bind to the ribosome and enhance peptide bond formation during protein synthesis by stabilizing the P-site tRNA.[15,16] However, to become fully functional, these translation factors must undergo post-translational modification (PTM).[6,7,9,14] Notably, in eukaryotes and archaea, e/aIF5A undergoes a PTM known as hypusination.[17] On the other hand, bacteria use diverse and unusual substrates and enzymes to modify their EF-Ps. For instance, in *Escherichia coli* and *Salmonella enterica*, a conserved lysine is β-lysinylated by the EF-P-(R)-β-lysine ligase (EpmA) and L-lysine 2,3-aminomutase (EpmB), as well as hydroxylated by the EF-P hydroxylase (EpmC).[18–22] Other modifications include rhamnosylation of an arginine by the arginine rhamnosyltransferase (EarP) in bacteria like *Pseudomonas putida* and *Shewanella oneidensis*[23,24] and amino-pentanolylation by the EF-P modification enzyme YmfI in *Bacillus subtilis*[25,26] (Figure 1A).

Interestingly, there is a group of bacteria, particularly from the Actinobacteria phylum, that possesses functional EF-Ps without the need for any PTM[27] (Figure 1A). We have shown that EF-Ps from the genera *Corynebacterium*, *Mycobacterium*, and *Streptomyces* function without PTM.[27] These EF-Ps have the otherwise modified lysine at the tip of a β-hairpin, flanked by two proline residues. It is suggested that the presence of this palindromic sequence Pro-Gly-Lys-Gly-Pro (PGKGP) within the hairpin confers rigidity and enables proper positioning of the protruding lysine 32, thereby stabilizing the acceptor arm of the tRNA. The PGKGP sequence has become a signature motif for this subfamily leading to its designation as the PGKGP subfamily of EF-P[27] (Figure 1A).

While it has been shown that a differentially modified EF-P variant can complement the deletion of *efp* in *E. coli* when co-expressed with the modification machinery,[23] unmodified EF-P variants were found to be inactive.[27] This observation suggests that certain features are responsible for conveying functionality without PTM. The current study was conducted to investigate the underlying design principles that enable the evolution of

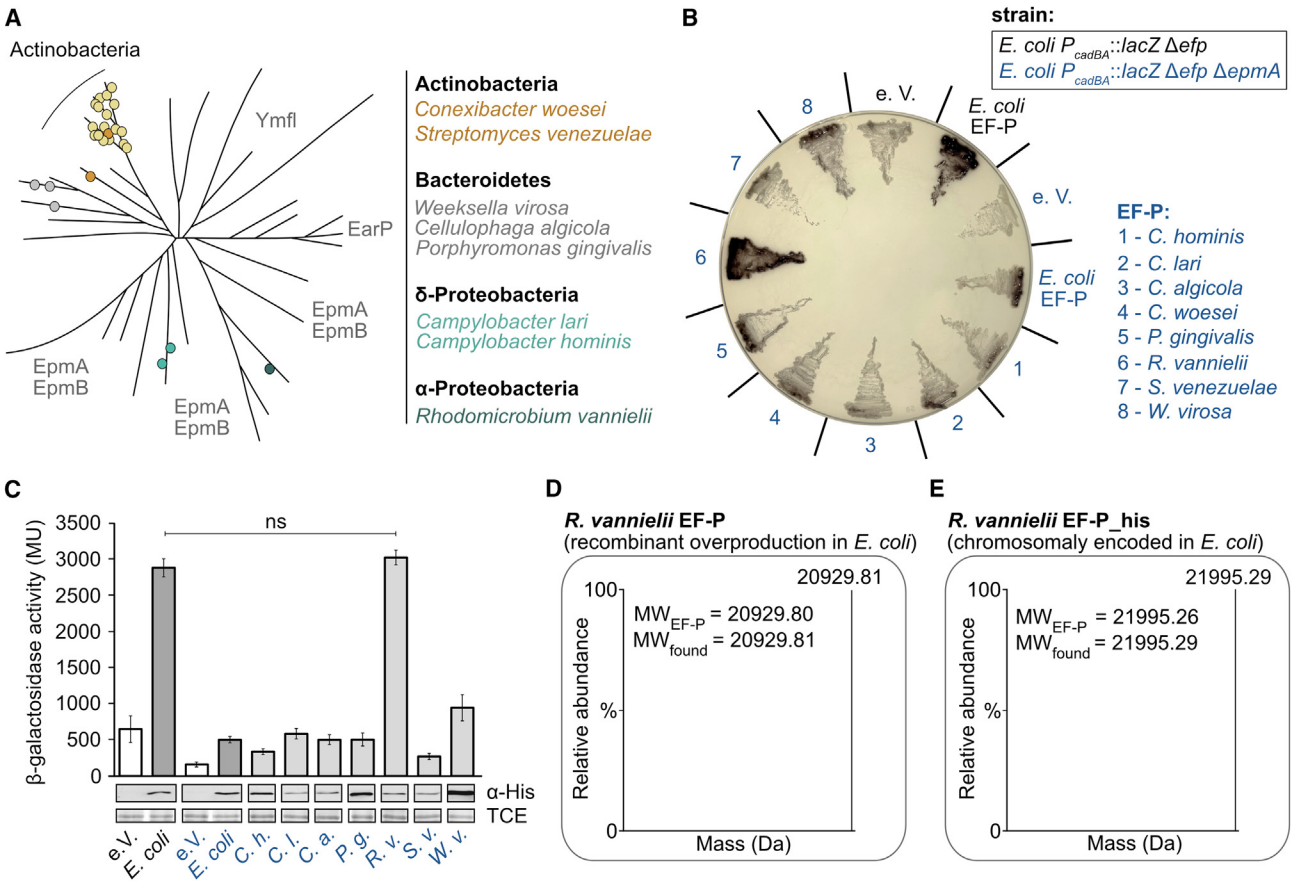


Figure 1. *R. vannielii* EF-P complements Δ*efp E. coli* mutants without the necessity of a PTM

(A) Schematic representation of the phylogenetic tree of bacterial EF-Ps, adapted from Pinheiro et al., 2020.[27] PGKGP subfamily EF-Ps used in this study are marked with colored dots (Actinobacteria – orange, Bacteroidetes – gray, δ-Proteobacteria – light green, α-Proteobacteria – dark green). Yellow dots represent the remaining PGKGP-type EF-Ps from Actinobacteria.

(B and C) EF-P activity measurements using the P_{cadBA}::*lacZ*-based reporter assay with S-Gal® or *o*-nitrophenyl-β-D-galactopyranoside (*o*-NPG) as substrates for the β-galactosidase activity. Color code in (C) corresponds to the strains used in (B). The β-galactosidase activities are given in Miller units (MU). EF-P production was confirmed by western blot analysis using antibodies against the His-tag. Protein bands corresponding to a 72-kDa protein after staining with 2,2,2-trichloroethanol (TCE) were used as loading controls. The blot is shown split as the samples were originally loaded onto the gel in a different order. The bands were cut and rearranged to achieve the desired order in the graph. Error bars indicate the standard deviation (SD) of three independent biological replicates. Statistics: Student's unpaired two-sided t test (ns, $p = 0.2808$).

(D) Deconvoluted MS spectra of intact *R. vannielii* EF-P, recombinantly overproduced in *E. coli.*

(E) Deconvoluted MS spectra of intact His-tagged *R. vannielii* EF-P (chromosomally encoded), produced *in E. coli*. See also Table S1; Figures S1–S4.

functional EF-Ps without the need for PTMs. We were able to identify three crucial amino acid positions with pivotal roles in maintaining the functionality of unmodified EF-P. Using this knowledge, we then succeeded in engineering a functional unmodified EF-P in the bacterial workhorse *E. coli*. These findings pave the way for approaches to optimize synthesis of proteins containing polyproline motifs at lower metabolic cost in other laboratory and industrially relevant bacteria.

RESULTS

Screen for naturally unmodified EF-Ps of the PGKGP subfamily that are functional in *E. coli*

Previous study has demonstrated that EF-Ps, belonging to the PGKGP subfamily of various Actinobacteria, are functional without PTM in their respective hosts, but they are non-functional in *E. coli.*[27] Notably, EF-Ps containing the PGKGP loop are also found in other bacterial phyla, which do not encode enzymes required for EF-P modification.[27] This observation has led to the hypothesis that these EF-Ps may also be functional in an unmodified state. Here, we screened several EF-Ps of the PGKGP subfamily of species from different phyla for their activity in *E. coli*. To this end, we chose *Campylobacter hominis* and *Campylobacter lari* as representatives of the δ-proteobacteria, *Cellulophaga algicola, Porphyromonas gingivalis,* and *Weeksella virosa* as representatives of the Bacteroidetes, *Conexibacter woesei* and *Streptomyces venezuelae* belonging to the Actinobacteria, and *Rhodomicrobium vannielii* as representatives of the α-Proteobacteria (Figure 1A). Despite an overall sequence identity of only 44% (Table S1), the PGKGP activation loop remains conserved.[27]

To elucidate the activity of these EF-Ps in *E. coli*, a well-established reporter system was used.[7] This reporter system is based on the finding that the transcriptional regulator CadC of *E. coli* is a polyproline protein that requires modified EF-P for its translation. In Δ*efp* mutants, the copy number of CadC is too low to induce the *cadBA* promoter (tested as P_{cadBA}::*lacZ*). CadC is activated when *E. coli* cells are exposed to low pH in the presence of lysine. To rule out that heterologously produced EF-Ps are modified by the *E. coli*-specific modification system EpmA, a reporter strain with an additional deletion of *epmA* encoding the β-lysine ligase was used. This reporter strain (*E. coli* P_{cadBA}::*lacZ* Δ*efp* Δ*epmA*) was transformed with plasmids expressing EF-Ps of different representatives of the PGKGP subfamily. *E. coli* was grown at pH 5.8 and β-galactosidase activities were determined. On agar plates containing S-Gal® and ferric ions, cells producing β-galactosidase can easily be detected by black precipitates. As a control, *E. coli efp* was expressed both in *E. coli* P_{cadBA}::*lacZ* Δ*efp* and *E. coli* P_{cadBA}::*lacZ* Δ*efp* Δ*epmA*. Only cells expressing both *efp* and the modification machinery showed black precipitates (Figure 1B). Of the eight selected EF-Ps of the PGKGP subfamily, only EF-P of *R. vannielii* was active. In contrast to *E. coli* EF-P, *R. vannielii* EF-P produced black precipitates in the double deletion Δ*efp* Δ*epmA* reporter strain (Figure 1B). To obtain quantitative results, β-galactosidase activity was determined using a colorimetric assay. *E. coli* EF-P was able to rescue the production of CadC and induce high β-galactosidase activity in the presence of the PTM machinery (Figure 1C). With the exception of *R. vannielii* EF-P, none of the other EF-Ps was able to complement the *E. coli* Δ*efp* mutant. Remarkably, the activity of the reporter strain producing the *R. vannielii* EF-P was comparable to the activity of the strain with the modified *E. coli* EF-P (Figures 1C and S1).

We then used mass spectrometry-based proteomics (MS) to test whether *R. vannielii* EF-P undergoes PTM in *E. coli*. We first purified recombinantly overproduced *R. vannielii* EF-P using the *E. coli* BL21/pET_SUMO system, which had the advantage that the His-tag could be easily cleaved from the protein after purification. For the detection of the *R. vannielii* EF-P without a His-tag, a polyclonal antibody against this protein was generated (α-625) and used for western blot analysis. The calculated mass of *R. vannielii* EF-P (20,929.80 Da) was consistent with the measured mass of the intact protein (20,929.81 Da), indicating the absence of a modification (Figures 1D and S2A). This observation was confirmed by analysis of peptides after chymotrypsin digestion (Data S1). Overproduction of EF-P in *E. coli* may result in a high level of unmodified EF-P[19] due to an imbalance between EF-P and its modifying enzymes. To avoid any artifact, we inserted *R. vannielii efp* (encoding a C-terminal *his*-tag) into the *E. coli* genome downstream of the native *efp* promoter and purified this protein (Figures S3A). The calculated mass of the His-tagged *R. vannielii* EF-P (21,995.26 Da) was consistent with the measured mass of the intact protein (21,995.29 Da) (Figures 1E and S2B), confirming that *R. vannielii* EF-P is unmodified in *E. coli*. This observation was further verified by analyzing the peptides after chymotrypsin digestion (Data S1). It has to be noted that the sequence available online (GenBank: NC_014664.1) for *R. vannielii* EF-P differs slightly from the sequence we determined for the strain used (DSM 162/ATCC 17100) (Figure S4).

Overall, these data demonstrate that an unmodified EF-P variant can support the synthesis of a protein containing polyproline motifs in *E. coli.*

Unmodified *R. vannielii* EF-P is fully functional in *E. coli*

Next, we investigated the functionality of *R. vannielii* EF-P in *E. coli* in more detail. First, we conducted growth tests, including the *E. coli* wild type (WT) and the Δ*efp*, Δ*epmA*, and Δ*efp* Δ*epmA* mutants, compared to the complemented Δ*efp*::*R. v. efp* and Δ*efp*::*R. v. efp* Δ*epmA* mutants. In the latter two strains, *R. v. efp* is chromosomally encoded under control of the native *E. coli efp* promoter (Figure 2A). After 12 h of growth in complex (LB) medium, the cell density of the mutants Δ*efp* and Δ*efp* Δ*epmA* was much lower compared to the WT, which is in line with previous observations[19,28] (Figure 2B). The mutant producing unmodified *E. coli* EF-P (Δ*epmA*) showed slightly higher cell densities but still exhibited limited growth compared to the WT (Figure 2B). However, expression of *R. v. efp* in both the Δ*epmA* and the Δ*efp* mutants suspended the growth defect (Figure 2B). To validate growth differences observed with the spot assay, we monitored the growth of the strains in LB medium at 37°C over time (Figure 2C). The growth defect of the mutant producing the unmodified *E. coli* EF-P (Δ*epmA*) was overcome by the chromosomal insertion of the *R. v efp* (Figure 2C). Only marginal differences in doubling times were detected between the WT and the mutants harboring the *R. v. efp* independent of the absence or presence of *epmA* (Figure 2C), confirming our hypothesis of a functional *R. v.* EF-P without the need for a PTM in *E. coli*. No significant differences were found in the amount of EF-P in *E. coli*, regardless of whether *epmA* was present or not (Figures S3A–S3D). These results emphasize that the observed growth defect of the Δ*epmA* mutant is due to the absence of the EF-P modification and is not a consequence of possible reduced EF-P levels. Compared to native EF-P, the level of *R. vannielii* EF-P was significantly lower in early exponential growth phase of *E. coli* (Figures S3C and S3D).

The absence of EF-P or its modifying enzymes in *E. coli* is associated with downregulation of proteins containing polyproline motifs and alterations of the proteome.[9] Therefore, we tested whether unmodified *R. vannielii* EF-P inserted into these mutants could restore the WT proteome. Cells were grown to late exponential growth phase and prepared for proteomic analysis. The changes in protein levels were determined by liquid chromatography-tandem mass spectrometry proteomics using data dependent acquisition (DDA) and label-free quantification (LFQ).[29] Identified peptide fragments mapped to a total of 2,378 proteins (Data S2). The differences in proteome comparison between the tested strains can be seen in the principal component analysis (PCA) (Figure 2D). The analyzed proteomes cluster into three populations with large distances between them: (cluster i) WT together with the two mutants producing *R. vannielii* EF-P (Δ*efp*::*R. v. efp*; Δ*efp*::*R. v. efp* Δ*epmA*), (cluster ii) mutants lacking EF-P (Δ*efp;* Δ*efp* Δ*epmA*), and (cluster iii) the mutant producing unmodified *E. coli* EF-P (Δ*epmA*)

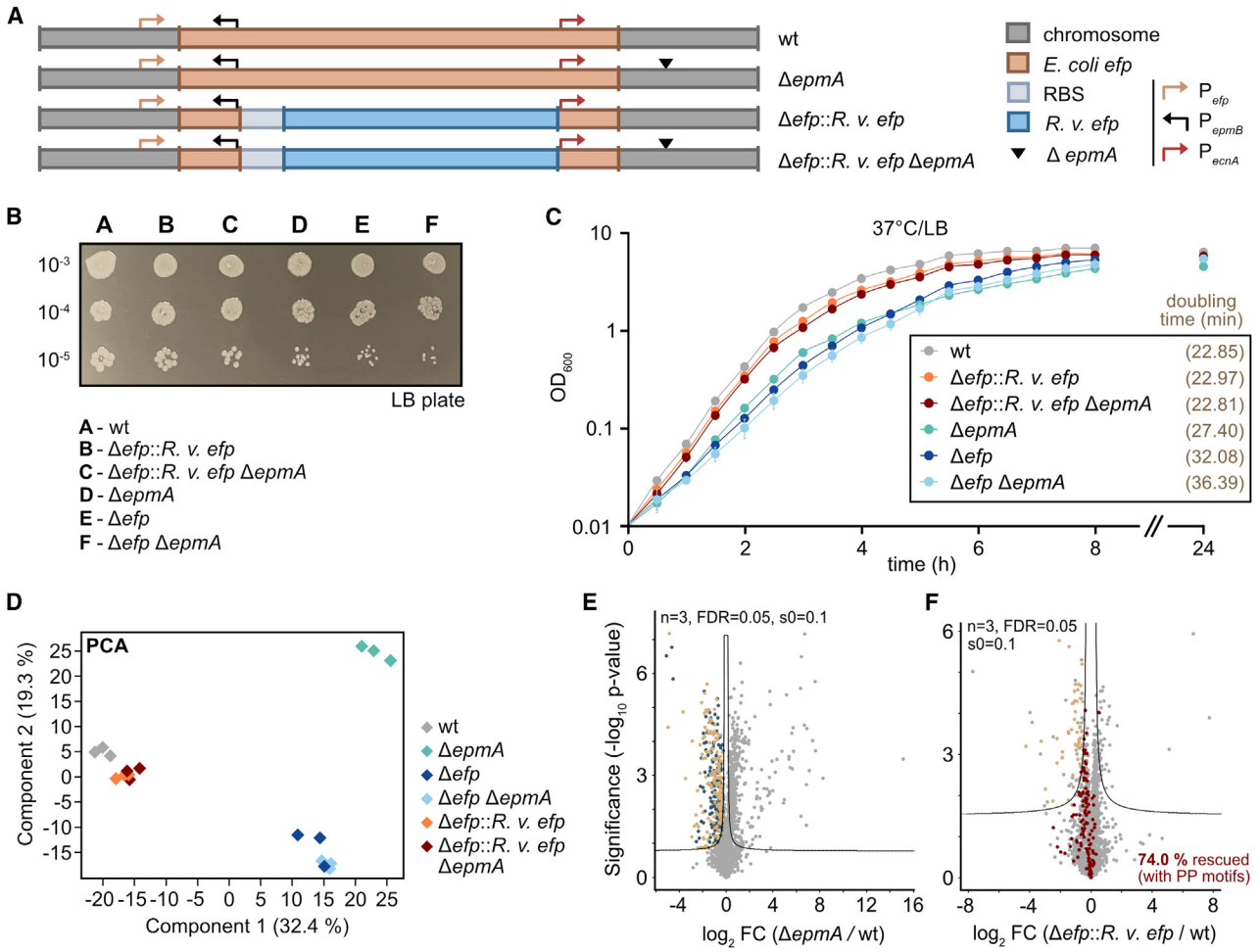


Figure 2. Phenotypic characterization of *E. coli* mutants complemented with *R. vannielii efp*

(A) Schematic overview of *E. coli* BW25113 *efp* deletion mutants with incorporated *R. vannielii efp*. Gene insertions are depicted in colored boxes and promoter locations as colored arrows.

(B) Spot assay of the *E. coli* mutants on an agar LB plate after incubation in LB overnight at 37°C.

(C) Growth of *E. coli* mutants in LB at 37°C. Data are mean values with error bars representing the standard deviation (SD) of four independent biological replicates. The doubling times of the individual strains are given in minutes (min).

(D) Proteome comparison of WT, Δ*epmA*, Δ*efp*, Δ*efp* Δ*epmA*, Δ*efp*::*R. v. efp*, and Δ*efp*::*R. v. efp* Δ*epmA* mutants using principal component analysis (PCA). Proteomes of three independent biological replicates were analyzed and are represented in rhombi with the same color.

(E and F) Volcano plot analysis, highlighting the downregulation of proteins with/without polyproline motifs (PP motifs) in *E. coli* mutant Δ*epmA* (E) and their rescue in mutant Δ*efp*::*R. v. efp* (F). Downregulated proteins without PP motifs are marked in dark blue dots; with PP motifs in light brown dots; rescued proteins in bordeaux red; and remaining identified proteins in gray. The x axes show the fold change (FC) of the mean value of the $\log_2$ protein intensity for each protein (LFQ) between two strains. The y axes show the significance level of the observed difference between the two strains ($-\log_{10} p$ value of the t test). See also Figures S3 and S5.

(Figure 2D). These data suggest that the proteome of mutants carrying the *R. v.* EF-P show the highest correlation with the WT *E. coli* strain.

We analyzed the proteomes in more detail with a specific focus on proteins with polyproline motifs. Comparisons between the proteomes of the Δ*epmA* mutant and the WT (Figure 2E) as well as between the Δ*efp* and Δ*efp* Δ*epmA* mutants and the WT (Figures S5A and S5B) revealed a strong protein scattering in the volcano plot indicating strong differences in the proteomes. Complementation of the Δ*efp* mutant with *R. v. efp* (Figure 2F) or the replacement of *E. c. efp* with *R. v. efp* in the Δ*epmA* mutant (Figure S5D) reduced the scattering, suggesting high similarity to the WT proteome.

The *E. coli* Δ*efp* and Δ*efp* Δ*epmA* mutants were characterized by downregulation of proteins with polyproline motifs compared to the WT (Data S2). Replacement of *E. c. efp* with *R. v. efp* allowed translational rescue of 130 (75.6%) and 152 (78.4%) proteins containing polyproline motifs, downregulated in the Δ*efp* and Δ*efp* Δ*epmA* mutants, respectively (Data S2, Figures S5C and S5D).

It has been previously reported that unmodified *E. coli* EF-P is still able to rescue the translation of polyproline motifs but with

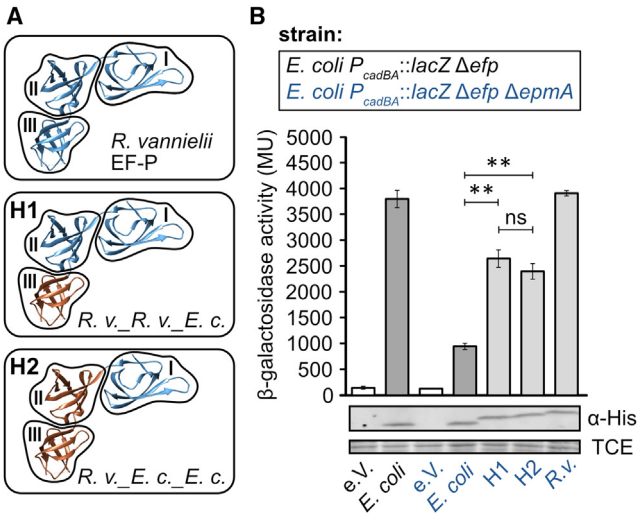


Figure 3. Activity of *E. coli* and *R. vannielii* EF-P hybrids

(A) Schematic overview of the constructed EF-P hybrids H1 (*R. v._R. v._E. c.*) and H2 (*R. v._E. c._E. c.*). Domains of *R. vannielii* EF-P (*R. v.*) are colored in blue and those of *E. coli* EF-P (*E. c.*) in brown.

(B) Activity measurements of the EF-P hybrids in *E. coli* using the P_{cadBA}::*lacZ*-based reporter assay. The β-galactosidase activities are given in Miller units (MU). EF-P production was confirmed by western blot analysis using antibodies against the His-tag. Protein bands corresponding to a 72-kDa protein after staining with 2,2,2-trichloroethanol (TCE) were used as loading controls. 3D protein structures for simplified EF-P domain representation are taken and modified from PDB: 3A5Z. Error bars indicate the standard deviation (SD) of three independent biological replicates. Statistics: Student's unpaired two-sided t test (****$p < 0.0001$; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; ns $p > 0.05$). Δ*efp* Δ*epmA* (*E. coli* vs. *R. v.*_H1, **$p = 0.0023$; *E. coli* vs. R. v._H2, **$p = 0.0021$; *R. v.*_H1 vs. *R. v.*_H2, ns $p = 0.2023$). See also Figure S6.

lower efficiency compared to the modified EF-P.[6,9,30] Here, we also found that of all the downregulated proteins with polyproline motifs in the Δ*efp* mutant (172 proteins) and the Δ*efp* Δ*epmA* mutant (194 proteins), 47 (27.3%) and 59 (30.4%), respectively, were not downregulated in cells producing unmodified *E. coli* EF-P (Δ*epmA* mutant) (Data S2). Importantly, *R. v.* EF-P was able to outperform the rescue potential of unmodified EF-P in *E. coli* by translationally rescuing 128 (74%) proteins with polyproline motifs (Figure 2F; Data S2).

Elucidation of the functional design principles of unmodified EF-P in *E. coli*

Our next aim was to uncover the design principles governing the full functionality of unmodified EF-Ps in *E. coli.* Accordingly, we sought to understand why not all EF-Ps of the PGKGP subfamily are active in *E. coli*. Previous reports showed that the bacterial EF-P consists of three domains, forming a shape similar to a tRNA.[31,32] To identify the essential domains for EF-P, we constructed *E. coli* and *R. vannielii* EF-P hybrids (Figure 3A) and measured their activities using the P_{cadBA}::*lacZ* reporter system. To determine the boundaries of each domain, we considered predicted secondary structures and 3D protein structures. Linker regions connecting the three domains were selected as fusion points to minimize effects on the EF-P 3D structure (Figures S6A–S6D; adapted from Uniprot: P0A6N4 and AlphaFold: AF-P0A6N4-F1[33–35]). The exchange of domains I and II of *E. c.* EF-P with the respective domains of *R. v.* EF-P (H1: *R. v._R. v._E. c*) more than doubled the activity of *E. coli* EF-P in the Δ*epmA* background (Figure 3B). When only domain I of *E. coli* EF-P was replaced by the corresponding *R. v.* EF-P domain (H2: *R. v._E. c._E. c*), the same increase in activity was observed. These observations suggest that domain I is of high importance for an active unmodified EF-P in *E. coli*.

To identify these amino acids, we used a synthetic molecular engineering approach, starting with multiple sequence alignments of EF-Ps of the PGKGP subfamily and *E. coli* (Figure 4A), followed by the construction of EF-P variants using site-directed mutagenesis and screening of those with highest activity using the P_{cadBA}::*lacZ* reporter assay (Figure 4B). The binding of EF-P to the negatively charged ribosomal parts (rRNA) is mediated mainly by charged amino acid side chains.[15] Therefore, we focused on amino acids that differ in charge between functional EF-Ps (*E. c.* EF-P and *R. v.* EF-P) and those that are unfunctional (remaining PGKGP subfamily EF-Ps) in *E. coli*, in particular positions 27, 28, 50, 56, 60, and 65 (Figure 4A).

S. venezuelae EF-P, a representative of the unmodified actinobacterial EF-Ps with low activity in *E. coli*,[27] was chosen for molecular engineering (Figure 4C). Out of all substitutions, Q27E, H28F, V50R, N56K, K60S, and T65D (numbering according to the *S. venezuelae* EF-P sequence), only the replacement of valine at position 50 to arginine (V50R) significantly increased the activity of *S. venezuelae* EF-P in *E. coli* (Figure 4C). This result suggested that a single amino acid substitution can already impact the overall functionality of EF-P. In the multiple sequence alignment of EF-Ps from the PGKGP subfamily, we detected a pattern for position 50 (Figure 4A). Strikingly, in almost all EF-P members of the PGKGP subfamily examined, only uncharged and nonpolar amino acids are present at position 50, with the exception of *R. vannielii* EF-P, which contains a positively charged lysine (K). The EF-P of *E. coli* also has a positively charged amino acid (R, arginine) at this position (Figure 4A). We tested the importance of the positively charged amino acid at position 50. Replacement of lysine by a neutral amino acid (K50V) reduced the activity of *R. vannielii* EF-P in *E. coli* to only 15%, whereas substitution by another positively charged amino acid, arginine (K50R), allowed 70% of the original activity (Figure 4D). Based on these results, we investigated the significance of a positively charged amino acid at position 50 in other potentially unmodified EF-Ps from the PGKGP subfamily. Thus, we constructed EF-P variants of *Cellulophaga algicola*, *Campylobacter hominis*, *Campylobacter lari*, *Conexibacter woesei*, *Porphyromonas gingivalis*, and *Weeksella virosa* EF-Ps with substitutions of valine (V) or isoleucine (I) by positively charged arginine (R) and analyzed their activities using the P_{cadBA}::*lacZ* reporter assay. Indeed, these amino acid substitutions significantly increased the activity of the EF-P variants of *C. hominis*, *P. gingivalis*, and *W. virosa* (Figure 4E), being in line with results for *S. venezuelae* EF-P (Figure 4C) and highlighting the importance of a positively charged amino acid at position 50 for the activity of unmodified EF-P in *E. coli*. Nevertheless, the introduction of a positively charged arginine (R) at position 50 was not sufficient to increase the activity of EF-P of *C. algicola* and *C. lari* although the proteins were synthesized (Figure 4E).

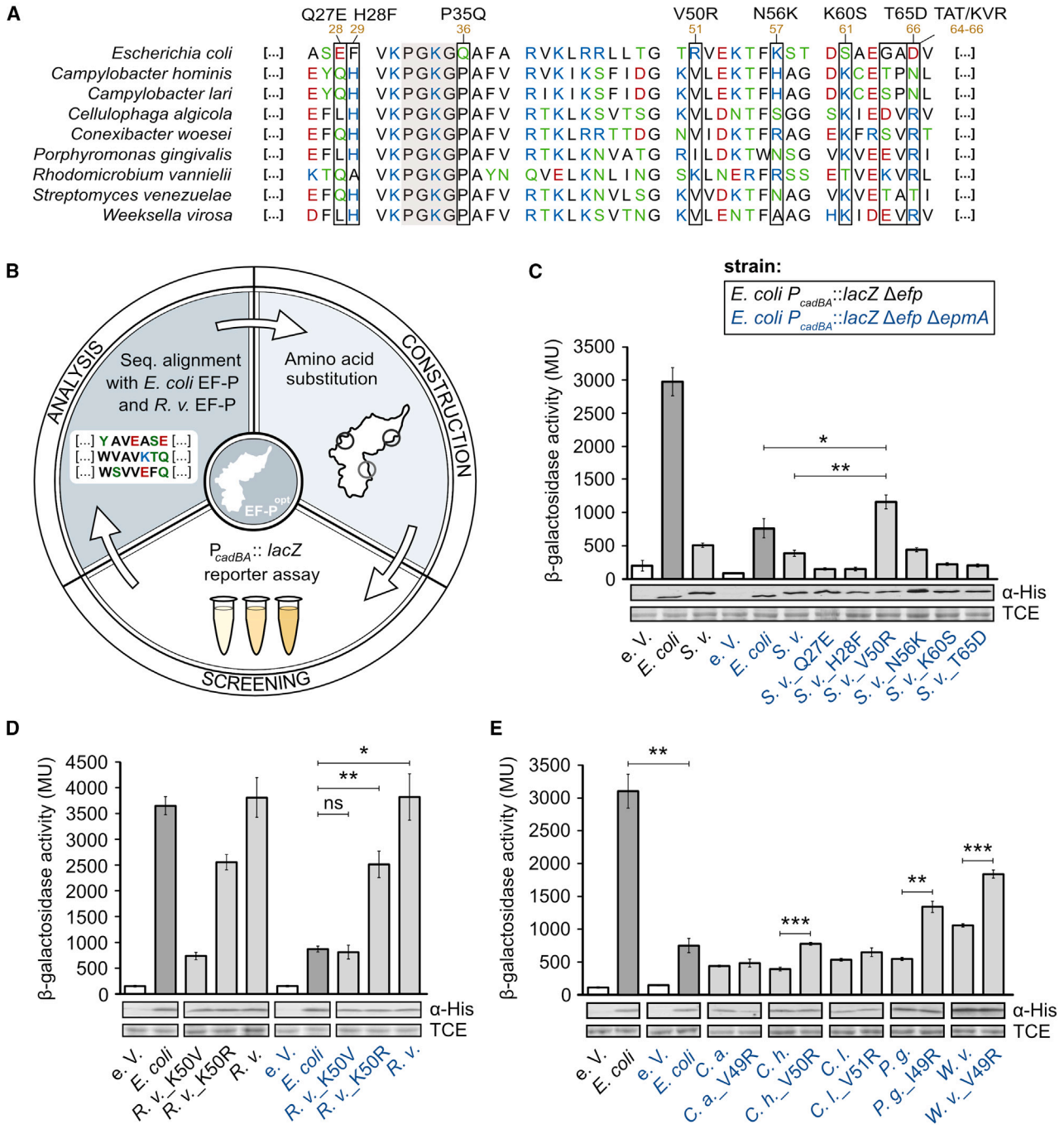


Figure 4. A positively charged amino acid at position 50 has an impact on the activity of PGKGP subfamily EF-Ps in *E. coli*

(A) Multiple sequence alignment of representative EF-Ps of the PGKGP subfamily and *E. coli*. Positions selected for substitution are highlighted and numbered according to *S. venezuelae* EF-P. The brown numbers indicate the positions in *E. coli*.

(B) Workflow for synthetic engineering of unmodified EF-P.

(C–E) Activity measurements of *S. venezuelae* EF-P (C), *R. vannielii* EF-P (D), and PGKGP subfamily EF-P variants (E) in *E. coli* using the P_{cadBA}::*lacZ*-based reporter assay. Color code in (D) and (E) corresponds to the strains used in (C). The β-galactosidase activities are given in Miller units (MU). EF-P production was confirmed by western blot analysis with antibodies against the His-tag. Protein bands corresponding to a 72-kDa protein after staining with 2,2,2-trichloroethanol (TCE) were used as loading controls. *C. a.* – *Cellulophaga algicola*, *C. h.* – *Campylobacter hominis*, *C. l.* – *Campylobacter lari*, *E. c.* – *E. coli*, *P. g.* – *Porphyromonas gingivalis*, *S. v.* – *Streptomyces venezuelae*, *W. v.* – *Weeksella virosa*, EF-P^{opt} – optimized EF-P. Error bars indicate the standard deviation (SD) of three

(legend continued on next page)

Overall, these data show that a single amino acid substitution has an effect on the activity of originally non-functional EF-Ps in *E. coli*.

A few amino acid changes are sufficient to generate a fully functional, unmodified EF-P in *E. coli*

The placement of a positively charged arginine at position 50 resulted in increased activity of *S. venezuelae* EF-P (variant A1) in *E. coli*, although not to the level observed for the modified *E. coli* EF-P (Figures 4C and 5A). Thus, we continued with the synthetic molecular engineering approach (Figure 4B) and investigated other amino acid positions. In particular, we focused on the role of proline 34 (P34), which—together with proline 30 (P30)—is part of the β-hairpin of the PGKGP subfamily EF-Ps.[27] Remarkably, substitution of P34 by glutamine (Q) enhanced the activity of the *S. venezuelae* EF-P variant (variant A2) (Figure 5A).

Analyzing all positions selected for substitution (Figure 4A), we observed a striking difference between amino acids at position 65 (numbering according to *S. venezuelae* EF-P) in *E. coli* and *R. vannielii*: there is a negative charge in *E. coli* EF-P (aspartic acid, D66) and a positive charge in *R. vannielii* EF-P (arginine, R65). Substitution T65D did not yield active *S. venezuelae* EF-P in *E. coli* (Figure 4C), which prompted us to replace amino acids of the region 63–65_TAT (T63-A64-T65, threonine – alanine – threonine) with those found in *R. vannielii* EF-P, 63–65_KVR (K63-V64-65R, lysine – valine – arginine). This substitution significantly improved the functionality of *S. venezuelae* EF-P in *E. coli* (variant A3) (Figure 5A).

We further investigated whether the combination of these substitutions could have an additive effect to increase the activity of *S. venezuelae* EF-P (Figure 4B). Indeed, the activity of variant A1_A2 (V50R_P35Q) was significantly higher than that of variant A1 (Figure 5A). Similarly, variants A1_A3 (V50R_63–65_KVR) and A2_A3 (P35Q_63–65_KVR) showed increased activities. Finally, the activity of the variant A1_A2_A3 (P35Q_V50R_63–65_KVR) was almost 6-fold higher than that of the original WT protein (*S. v.*) (Figure 5A). Nonetheless, the activity of *S. venezuelae* EF-P_A1_A2_A3 remained significantly lower than that of the modified *E. coli* EF-P (Figure 5A).

We found weaker signals in the western blot for *S. venezuelae* EF-P_A1_A2_A3 compared to *E. coli* EF-P (Figure 5A), which indicates lower production and could explain the lower activity. This observation prompted us to search for other EF-P PGKGP family members that are better expressed in *E. coli.* Since the EF-Ps derived from *P. gingivalis* and *W. virosa* were produced very well in *E. coli* (Figure 4E), we decided to continue synthetic molecular engineering with them. Variants with P34Q or E62K substitutions in both EF-Ps had significantly higher activities than the corresponding native proteins (Figure 5B). It should be noted that in both proteins, only the first amino acid (E62K) had to be substituted to obtain the crucial KVR motif (Figure 4A). Remarkably, the *W. virosa* EF-P variant with all three amino acid substitutions (*W. virosa* EF-P_S1_S2_S3; EF-P_P34Q_V49R_E62K) reached the activity level of the modified *E. coli* EF-P (Figure 5C). The activity of *W. virosa* EF-P_S1-S3 in *E. coli* was also not affected by the presence of the modifying enzyme EpmA (Figure 5D).

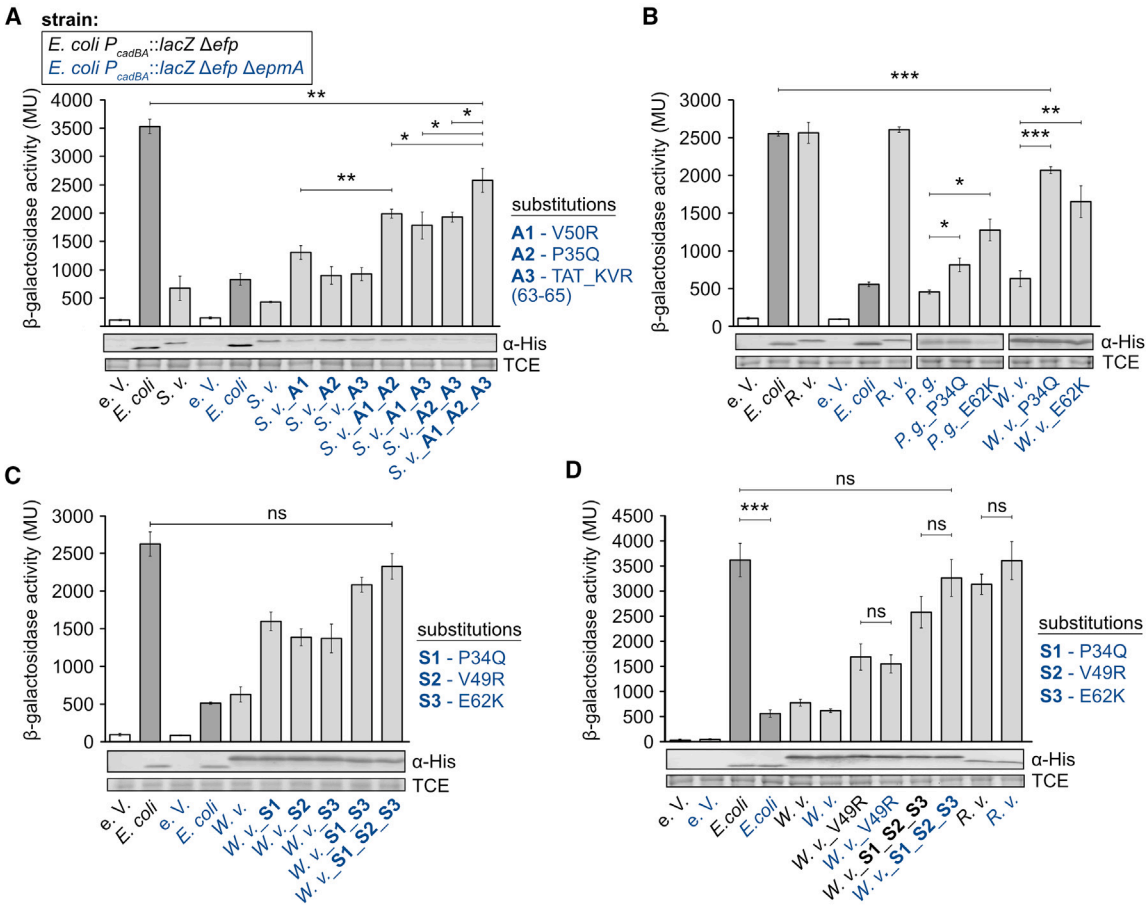
To this end, we characterized *E. coli* strains expressing *W. virosa* EF-P instead of their native modified or unmodified EF-P. For this purpose, *E. coli* mutants with chromosomally inserted *W. virosa efp*_S1-S3 were constructed (Figure S3E), and western blot analysis revealed sufficiently high production of this variant (Figures S3F and S3G). *W. virosa* EF-P_S1-S3 was able to rescue the growth phenotypes observed for the Δ*epmA E. coli* mutant independent of the presence of *epmA* (Figure 5E). The chromosomally encoded *W. virosa* EF-P_S1-S3 was purified by affinity chromatography. MS analysis of the intact protein confirmed that this EF-P variant was unmodified. The calculated protein mass of the *W. virosa* EF-P_S1-S3 variant was consistent with the measured intact protein mass, regardless of the presence of EpmA (calculated: 22,226.21 Da; measured: Δ*epmA* 22,226.33 Da; *epmA*$^+$ 22,226.21 Da) (Figures 5F, 5G, S2C, and S2D). It should be noted that the measured protein masses are given without the first methionine, a phenomenon frequently observed in previous studies.[36,37] To test whether the amino acid substitutions affected possible modification events, we also analyzed the intact protein mass of purified *W. virosa* WT EF-P, and the calculated protein mass (22,139.12 Da) was consistent with the measured intact protein mass, whether or not the cells expressed EpmA (Δ*epmA:* 22,139.26 Da; *epmA*$^+$: 22,139.09 Da) (Figures 5H, 5I, S2E, and S2F). We observed two peaks in the mass spectrometry measurements of *W. virosa* WT EF-P (Figure 5I). These peaks correspond to the mass of *W. virosa* EF-P with (22,266.18 Da) and without (22,139.09 Da) methionine.

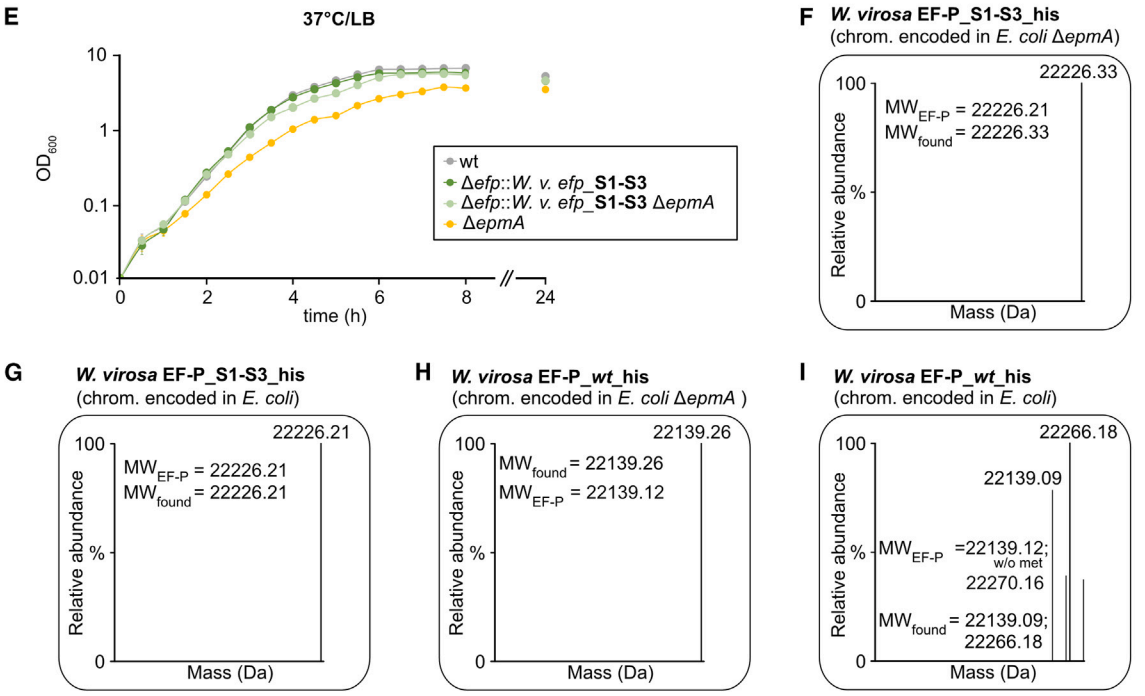
Overall, the finding that the EF-Ps of *S. venezuelae* and *W. virosa* can be significantly activated by a small number of substitutions demonstrates that certain design principles found in *R. vannielii* EF-P can be transferred to other isoforms within the same group. Our data show that three amino acid substitutions are sufficient to convert an inactive EF-P into a fully functional unmodified variant. Moreover, an *E. coli* mutant is now available as a host for heterologous production of proteins containing polyproline motifs with lower metabolic and energy costs, since EF-P no longer requires PTM.

DISCUSSION

For optimal synthesis of proteins with polyproline stretches, the EF-P of *E. coli* requires a PTM catalyzed by the modifying enzymes EpmA, EpmB, and EpmC.[18–21] EpmB converts the precursor (*S*)-α-lysine to (*R*)-β-lysine, which is then ligated by EpmA in an ATP-dependent condensation reaction to the ε-amino group of Lys34 in EF-P. EpmC subsequently catalyzes the hydroxylation of Lys34. Except for hydroxylation, none of

independent biological replicates. Statistics: Student's unpaired two-sided t test (****p < 0,0001; ***p < 0.001; **p < 0.01; *p < 0.05; ns p > 0.05). Δ*efp* Δ*epmA* (*E. coli* vs. *S. v.*_V50R, *p = 0.0396; *S. v.* vs. *S. v.*_V50R, **p = 0.0035) (C); Δ*efp* Δ*epmA* (*E. coli* vs. *R. v.*_K50V, ns p = 0.6099; *E. coli* vs. *R. v.*_K50R, **p = 0.0092; *E. coli* vs. *R. v.*, *p = 0.0104) (D); Δ*efp* Δ*epmA* (*C. h.* vs. *C. h.*_V50R, ***p = 0.0001; *P. g.* vs. *P. g.*_I49R, **p = 0.0037; *W. v.* vs. *W. v.*_V49R, ***p = 0.00099), *E. coli* (in Δ*efp*) vs. *E. coli* (in Δ*efp* Δ*epmA*), **p = 0.0022 (E). See also Figure S1.





(legend on next page)

the modification steps can be bypassed without noticeable effects on translational rescue of proteins with polyproline motifs. Even modification of EF-P with α-lysine instead of β-lysine decreases its activity.[38] Notably, the absence of modification in EF-P impairs the synthesis of proteins with polyproline motifs.[9] This study revealed that only one-third of proteins with polyproline motifs found to be downregulated in the Δ*efp* mutant were rescued by unmodified E-FP (Δ*epmA* mutant) (Data S2). Thus, while the lack of PTM does not completely prevent EF-P from its original mode of action, it impacts its efficiency. The PTM of *E. coli* EF-P is associated with metabolic (β-lysine) and energetic costs (ATP) as well as the production of the modification enzymes EpmA, EpmB, and EpmC.

The relatively high abundance of diproline (XPPX) motif-containing proteins in *E. coli* (2,101 motifs, 0.49 motif/protein),[13] shown to cause ribosome stalling,[9] underscores the need for an EF-P with maximal functionality to maintain proteome homeostasis, especially at high growth rates.[28] Many representatives of Actinobacteria have an even higher number of XPPX motifs within their proteomes.[27] In *Streptomyces coelicolor* and *Mycobacterium tuberculosis*, the number of XPPX motifs even surpasses the number of encoded proteins (1.08 motif/protein and 1.17 motif/protein, respectively).[27] Accordingly, one would expect PTMs to be crucial in these bacteria. However, the actinobacterial EF-Ps are not modified.[27] Both *S. coelicolor* and *M. tuberculosis* grow very slowly.[27,39–41] In these slowly growing bacteria, an unmodified, and therefore metabolically and energetically less demanding, EF-P could provide a selective advantage when coping with the production of a large number of proteins with polyproline motifs.

However, actinobacterial unmodified EF-Ps could not rescue the translation of PP proteins in *E. coli*.[27] To find out why certain EF-Ps exhibit functionality in their native host without PTM, but are non-functional in *E. coli*, we analyzed the activity of EF-Ps belonging to the PGKGP subfamily from different phyla (Figure 1A). We found that of all eight tested PGKGP subfamily EF-Ps, the EF-P from *R. vannielii* was able to rescue the translation of proteins with polyproline motifs in *E. coli* (Figures 1B and 1C). Importantly, the EF-P from *R. vannielii* retained its functionality in an unmodified state in *E. coli* (Figures 1D and 1E). *E. coli* strains expressing *R. vannielii efp* rescued growth defects observed in *E. coli* mutants lacking *efp* or *epmA*. Moreover, the proteome profiles of *R. vannielii* EF-P-producing strains closely resemble those of the WT (Figures 2B–2F).

EF-Ps have conserved regions required for ribosomal contact and adopt a tRNA-like structure, hinting at the evolutionary origin of these proteins.[31,42] The analogy of EF-P to tRNA, which is one of the building blocks of the translational machinery, likely increased the ability of this protein to bind between the ribosomal E- and P-sites[42] (Figure S7A). Therefore, we hypothesized that inactive EF-Ps might essentially be functional in *E. coli* but hindered from contributing to the translation process by weak binding to the ribosome or insufficient contact to the tRNA. To investigate this, we developed a synthetic molecular engineering workflow involving the substitution of amino acids based on the sequence of the functional *R. vannielii* EF-P (Figure 4B). By replacing single amino acids at positions 35, 50, and 63–65 (numbered according to *S. venezuelae* EF-P, Figure 4A), we successfully converted initially non-functional EF-Ps of *S. venezuelae*, *P. gingivalis*, and *W. virosa* into functional ones in *E. coli* (Figures 4C, 4E, and 5A–5D). Previous studies have reported the importance of certain amino acids at position 35 for EF-P activity. Pinheiro et al. showed that a proline at position 34 (position 35 in *S. venezuelae*) is necessary for EF-P to function with highest activity in *Corynebacterium glutamicum*, as substitutions to alanine, glutamine, glycine, and asparagine decreased activity of the corresponding EF-P variants.[27] In *E. coli*, a glutamine (Q) at the corresponding position has been shown to be required for proper function of EF-P.[30] These observations are in line with our results, as substitution of this proline to glutamine significantly increased the activity of the corresponding *S. venezuelae*, *P. gingivalis*, and *W. virosa* variants in *E. coli* (Figures 5A–5C). It is suggested that the amino acid at position 35 is essential for the correct positioning of the conserved lysine (either modified or non-modified) in the β-hairpin to get contact with P-site tRNA.[15,27] The presence of a glutamine at this particular position could be important to

Figure 5. Optimized unmodified EF-P variants are active in *E. coli*

(A–D) Activity measurements of EF-P variants from *S. venezuelae* (A), *P. gingivalis* (B), and *W. virosa* (B, C, and D) in *E. coli* using the P_{cadBA}::*lacZ*-based reporter assay. Color code in (B), (C), and (D) corresponds to the strains used in (A). The β-galactosidase activities are given in Miller units (MU). EF-P production was confirmed by western blot analysis using antibodies against the His-tag. Protein bands corresponding to a 72-kDa protein after staining with 2,2,2-trichloroethanol (TCE) were used as loading controls.

(E) Growth of *E. coli* mutants expressing *W. virosa*_S1-S3 *efp* in LB at 37°C.

(F and G) Deconvoluted mass spectra of the chromosomally encoded His-tagged *W. virosa* EF-P _S1-S3, heterologously produced in *E. coli* Δ*epmA* or *E. coli* WT, respectively. It should be noted that the calculated mass of *W. virosa* corresponds to a protein without the first methionine.

(H and I) Deconvoluted mass spectra of the chromosomally encoded His-tagged *W. virosa* EF-P WT, heterologously produced in *E. coli* Δ*epmA* or *E. coli* WT, respectively. Calculated masses correspond to *W. virosa* EF-P_WT_6xhis = 22270.16 Da and *W. virosa* EF-P_WT_6xhis without the first methionine = 22139.12 Da. Error bars in (A)–(D) indicate the standard deviation (SD) of at least three independent biological replicates. Data in (E) are mean values with error bars representing the standard deviation (SD) of four independent biological replicates. Statistics: Student's unpaired two-sided t test (****p < 0,0001; ***p < 0.001; **p < 0.01; *p < 0.05; ns p > 0.05). Calculated p values: Δ*efp* Δ*epmA* (*S. v.*_A1 vs. *S. v.*_A1_A2, **p = 0.0045; *S. v.*_A1_A2 vs. *S. v.*_A1_A2_A3, *p = 0.0454; *S. v.*_A1_A3 vs. *S. v.*_A1_A2_A3, *p = 0.0250; *S. v.*_A2_A3 vs. *S. v.*_A1_A2_A3, *p = 0.0349), *E. coli* (in Δ*efp*) vs. *S. v.*_A1_A2_A3 (in Δ*efp* Δ*epmA*), **p = 0.0096 (a); Δ*efp* Δ*epmA* (*P. g.* vs. *P. g.*_P34Q, *p = 0.0244; *P. g.* vs. *P. g.*_E62K, *p = 0.0132; *W. v.* vs. *W. v.*_P34Q, ***p = 0.0006; *W. v.* vs. *W. v.*_E62K, **p = 0.0094), *E. coli* (in Δ*efp*) vs. *W. v*_P34Q (in Δ*efp* Δ*epmA*), ***p = 0.0005 (B); *E. coli* (in Δ*efp*) vs. *W. v*_S1_S2_S3 (in Δ*efp* Δ*epmA*), ns p = 0.1484 (C). *E. coli* (in Δ*efp*) vs. *E. coli* (in Δ*efp* Δ*epmA*), ***p = 0.0003; *W. v.*_V49R (in Δ*efp*) vs. *W. v.*_V49R (in Δ*efp* Δ*epmA*), ns p = 0.4876; *W. v.*_S1_S2_S3 (in Δ*efp*) vs. *W. v.*_S1_S2_S3 (in Δ*efp* Δ*epmA*), ns p = 0.0506; *R. v.* (in Δ*efp*) vs. *R. v.* (in Δ*efp* Δ*epmA*), ns p = 0.1176; *E. coli* (in Δ*efp*) vs. *W. v.* S1_S2_S3 (in Δ*efp* Δ*epmA*), ns p = 0.2580 (D). See also Figures S1–S3 and S7.

allow contacts with the *E. coli* ribosome (23*S* rRNA) and consequently stabilization and/or orientation of the β-hairpin with the essential lysine at the tip in the absence of a modification (Figure S7B).

Another interesting finding was that substitution of amino acids at position 50 into an arginine (R) resulted in an increase in activity for all tested variants (Figure 4E). This positively charged amino acid is predicted to be in close contact with the ribosome[15,16,42] and therefore important for the corresponding EF-P to establish contacts with the ribosome (Figure S7C). We noted that a positively charged amino acid at position 50 is also found in some EF-P orthologs, the eukaryotic/archaeal initiation factor 5A (e/aIF5A).[31] Finally, the substitution of GAD (positions 64_66 in *E. coli*) against KVR had a major impact on the functionality of the unmodified EF-P variants and might be explained by a stabilization of the interaction between EF-P and the P-site tRNA[15] (Figure S7D).

In conclusion, bacteria, archaea, and eukaryotes have evolved different ways to cope with the synthesis of proteins containing polyproline motifs. There are functional EF-Ps that require PTMs, as well as those that do not necessitate modification. According to previous knowledge, it was suggested that the tip of the modified loop of EF-P reaches the carbonyl group of the P-site tRNA substrate in the peptidyl transferase center and stabilizes it for the transpeptidation reaction.[2,6,43] Our study reveals that naturally occurring unmodified EF-Ps, which have low activity in *E. coli*, can be made functional by altering distinct amino acids. This suggests that the alleviation of the polyproline-dependent ribosome stalling by EF-P may not be exclusively related to the PTM itself but rather to the affinity of EF-P to the ribosome and/or the tRNA-Pro in the P-site.

The modification of EF-P with β-lysine in *E. coli* requires not only a substrate but also energy-intense enzymes to catalyze this process. Here, we describe an EF-P that is functional in *E. coli* without the need for a PTM, making it attractive for future studies aiming to reallocate cellular energy for other energy-consuming processes in bacteria. An optimally functioning EF-P is important for proteome homeostasis and bacterial virulence.[44,45] Thus, decrypting the functional design of unmodified EF-P in *E. coli* not only directs protein production in a resource-efficient manner for scientific and industrial applications but also extends the fundamental understanding of the functional principles of this translation factor.

Limitations of the study

Our study provides insights into the functional design of unmodified translation elongation factor P. There are limitations of the study that we would like to address. All of the EF-P variants we generated were tested for production using the western blot technique. As this is a semi-quantitative technique, we could not calculate the specific activity of the variants. In addition, we did not test their individual stability and conformational heterogeneity. We do not exclude other possible substitution combinations that would lead to active unmodified EF-Ps in *E. coli*. To gain insights into the molecular mechanism for the rescue of ribosome stalling by unmodified EF-P, additional data are required, such as determining the affinity of EF-P variants to bind to the ribosome.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Plasmid and bacterial strain construction
  - β-Galactosidase activity assays
  - Protein purification and structural analysis
  - SDS-PAGE and Western Blot analysis
  - Mass spectrometry for identification of modification status
  - Mass spectrometry for identification of modification status
  - Mass spectrometry for proteome analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.114063.

ACKNOWLEDGMENTS

We thank Prof. Bernhard Schink from University of Konstanz for the support in handling *R. vannielii*. We also thank Haoyu Chen and Thomas Krause for excellent technical assistance. This work was financially supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant JU270/20-1, project number 449926427, RTG2062 (Molecular Principles of Synthetic Biology) to K.J., and SFB 1309, project number 325871075, to K.J., P.K., and A.I.

AUTHOR CONTRIBUTIONS

Conceptualization, U.T. and K.J.; methodology, U.T., P.K., I.F., A.I., and K.J.; investigation, U.T., P.K., R.K., and I.F.; writing – original draft, U.T. and K.J.; funding acquisition and resources, P.K., A.I., and K.J.; supervision, K.J.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 20, 2023
Revised: January 17, 2024
Accepted: March 21, 2024
Published: April 17, 2024

REFERENCES

1. Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. Nat. Rev. Mol. Cell Biol. *19*, 20–30. https://doi.org/10.1038/nrm.2017.91.
2. Doerfel, L.K., Wohlgemuth, I., Kubyshkin, V., Starosta, A.L., Wilson, D.N., Budisa, N., and Rodnina, M.V. (2015). Entropic Contribution of Elongation Factor P to Proline Positioning at the Catalytic Center of the

Ribosome. J. Am. Chem. Soc. *137*, 12997–13006. https://doi.org/10.1021/jacs.5b07427.

3. Muto, H., and Ito, K. (2008). Peptidyl-prolyl-tRNA at the ribosomal P-site reacts poorly with puromycin. Biochem. Biophys. Res. Commun. *366*, 1043–1047. https://doi.org/10.1016/j.bbrc.2007.12.072.
4. Pavlov, M.Y., Watts, R.E., Tan, Z., Cornish, V.W., Ehrenberg, M., and Forster, A.C. (2009). Slow peptide bond formation by proline and other N-alkylamino acids in translation. Proc. Natl. Acad. Sci. USA *106*, 50–54. https://doi.org/10.1073/pnas.0809211106.
5. Wohlgemuth, I., Brenner, S., Beringer, M., and Rodnina, M.V. (2008). Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. J. Biol. Chem. *283*, 32229–32235. https://doi.org/10.1074/jbc.M805316200.
6. Doerfel, L.K., Wohlgemuth, I., Kothe, C., Peske, F., Urlaub, H., and Rodnina, M.V. (2013). EF-P Is Essential for Rapid Synthesis of Proteins Containing Consecutive Proline Residues. Science *339*, 85–88. https://doi.org/10.1126/science.1229017.
7. Ude, S., Lassak, J., Starosta, A.L., Kraxenberger, T., Wilson, D.N., and Jung, K. (2013). Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. Science *339*, 82–85. https://doi.org/10.1126/science.1228985.
8. Tanner, D.R., Cariello, D.A., Woolstenhulme, C.J., Broadbent, M.A., and Buskirk, A.R. (2009). Genetic Identification of Nascent Peptides That Induce Ribosome Stalling. J. Biol. Chem. *284*, 34809–34818. https://doi.org/10.1074/jbc.M109.039040.
9. Peil, L., Starosta, A.L., Lassak, J., Atkinson, G.C., Virumäe, K., Spitzer, M., Tenson, T., Jung, K., Remme, J., and Wilson, D.N. (2013). Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. Proc. Natl. Acad. Sci. USA *110*, 15265–15270. https://doi.org/10.1073/pnas.1310642110.
10. Adzhubei, A.A., Sternberg, M.J.E., and Makarov, A.A. (2013). Polyproline-II helix in proteins: structure and function. J. Mol. Biol. *425*, 2100–2132. https://doi.org/10.1016/j.jmb.2013.03.018.
11. Starosta, A.L., Lassak, J., Peil, L., Atkinson, G.C., Woolstenhulme, C.J., Virumäe, K., Buskirk, A., Tenson, T., Remme, J., Jung, K., and Wilson, D.N. (2014). A conserved proline triplet in Val-tRNA synthetase and the origin of elongation factor P. Cell Rep. *9*, 476–483. https://doi.org/10.1016/j.celrep.2014.09.008.
12. Motz, M., and Jung, K. (2018). The role of polyproline motifs in the histidine kinase EnvZ. PLoS One *13*, e0199782. https://doi.org/10.1371/journal.pone.0199782.
13. Qi, F., Motz, M., Jung, K., Lassak, J., and Frishman, D. (2018). Evolutionary analysis of polyproline motifs in *Escherichia coli* reveals their regulatory role in translation. PLoS Comput. Biol. *14*, e1005987. https://doi.org/10.1371/journal.pcbi.1005987.
14. Gutierrez, E., Shin, B.S., Woolstenhulme, C.J., Kim, J.R., Saini, P., Buskirk, A.R., and Dever, T.E. (2013). eIF5A promotes translation of polyproline motifs. Mol. Cell *51*, 35–45. https://doi.org/10.1016/j.molcel.2013.04.021.
15. Huter, P., Arenz, S., Bock, L.V., Graf, M., Frister, J.O., Heuer, A., Peil, L., Starosta, A.L., Wohlgemuth, I., Peske, F., et al. (2017). Structural Basis for Polyproline-Mediated Ribosome Stalling and Rescue by the Translation Elongation Factor EF-P. Mol. Cell *68*, 515–527.e6. https://doi.org/10.1016/j.molcel.2017.10.014.
16. Blaha, G., Stanley, R.E., and Steitz, T.A. (2009). Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. Science *325*, 966–970. https://doi.org/10.1126/science.1175800.
17. Park, M.H., Cooper, H.L., and Folk, J.E. (1982). The biosynthesis of protein-bound hypusine (N epsilon -(4-amino-2-hydroxybutyl)lysine). Lysine as the amino acid precursor and the intermediate role of deoxyhypusine (N epsilon -(4-aminobutyl)lysine). J. Biol. Chem. *257*, 7217–7222.
18. Behshad, E., Ruzicka, F.J., Mansoorabadi, S.O., Chen, D., Reed, G.H., and Frey, P.A. (2006). Enantiomeric free radicals and enzymatic control of stereochemistry in a radical mechanism: the case of lysine 2,3-aminomutases. Biochemistry *45*, 12639–12646. https://doi.org/10.1021/bi061328t.
19. Yanagisawa, T., Sumida, T., Ishii, R., Takemoto, C., and Yokoyama, S. (2010). A paralog of lysyl-tRNA synthetase aminoacylates a conserved lysine residue in translation elongation factor P. Nat. Struct. Mol. Biol. *17*, 1136–1143. https://doi.org/10.1038/nsmb.1889.
20. Navarre, W.W., Zou, S.B., Roy, H., Xie, J.L., Savchenko, A., Singer, A., Edvokimova, E., Prost, L.R., Kumar, R., Ibba, M., and Fang, F.C. (2010). PoxA, yjeK, and elongation factor P coordinately modulate virulence and drug resistance in *Salmonella enterica*. Mol. Cell *39*, 209–221. https://doi.org/10.1016/j.molcel.2010.06.021.
21. Peil, L., Starosta, A.L., Virumäe, K., Atkinson, G.C., Tenson, T., Remme, J., and Wilson, D.N. (2012). Lys34 of translation elongation factor EF-P is hydroxylated by YfcM. Nat. Chem. Biol. *8*, 695–697. https://doi.org/10.1038/nchembio.1001.
22. Roy, H., Zou, S.B., Bullwinkle, T.J., Wolfe, B.S., Gilreath, M.S., Forsyth, C.J., Navarre, W.W., and Ibba, M. (2011). The tRNA synthetase paralog PoxA modifies elongation factor-P with (R)-beta-lysine. Nat. Chem. Biol. *7*, 667–669. https://doi.org/10.1038/nchembio.632.
23. Lassak, J., Keilhauer, E.C., Fürst, M., Wuichet, K., Gödeke, J., Starosta, A.L., Chen, J.M., Søgaard-Andersen, L., Rohr, J., Wilson, D.N., et al. (2015). Arginine-rhamnosylation as new strategy to activate translation elongation factor P. Nat. Chem. Biol. *11*, 266–270. https://doi.org/10.1038/nchembio.1751.
24. Rajkovic, A., Erickson, S., Witzky, A., Branson, O.E., Seo, J., Gafken, P.R., Frietas, M.A., Whitelegge, J.P., Faull, K.F., Navarre, W., et al. (2015). Cyclic Rhamnosylated Elongation Factor P Establishes Antibiotic Resistance in *Pseudomonas aeruginosa*. mBio *6*, e00823. https://doi.org/10.1128/mBio.00823-15.
25. Rajkovic, A., Hummels, K.R., Witzky, A., Erickson, S., Gafken, P.R., Whitelegge, J.P., Faull, K.F., Kearns, D.B., and Ibba, M. (2016). Translation Control of Swarming Proficiency in *Bacillus subtilis* by 5-Amino-pentanolylated Elongation Factor P. J. Biol. Chem. *291*, 10976–10985. https://doi.org/10.1074/jbc.M115.712091.
26. Hummels, K.R., Witzky, A., Rajkovic, A., Tollerson, R., 2nd, Jones, L.A., Ibba, M., and Kearns, D.B. (2017). Carbonyl reduction by YmfI in *Bacillus subtilis* prevents accumulation of an inhibitory EF-P modification state. Mol. Microbiol. *106*, 236–251. https://doi.org/10.1111/mmi.13760.
27. Pinheiro, B., Scheidler, C.M., Kielkowski, P., Schmid, M., Forné, I., Ye, S., Reiling, N., Takano, E., Imhof, A., Sieber, S.A., et al. (2020). Structure and Function of an Elongation Factor P Subfamily in Actinobacteria. Cell Rep. *30*, 4332–4342.e5. https://doi.org/10.1016/j.celrep.2020.03.009.
28. Tollerson, R., 2nd, Witzky, A., and Ibba, M. (2018). Elongation factor P is required to maintain proteome homeostasis at high growth rate. Proc. Natl. Acad. Sci. USA *115*, 11072–11077. https://doi.org/10.1073/pnas.1812025115.
29. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics *13*, 2513–2526. https://doi.org/10.1074/mcp.M113.031591.
30. Volkwein, W., Krafczyk, R., Jagtap, P.K.A., Parr, M., Mankina, E., Macošek, J., Guo, Z., Fürst, M.J.L.J., Pfab, M., Frishman, D., et al. (2019). Switching the Post-translational Modification of Translation Elongation Factor EF-P. Front. Microbiol. *10*, 1148. https://doi.org/10.3389/fmicb.2019.01148.
31. Hanawa-Suetsugu, K., Sekine, S.i., Sakai, H., Hori-Takemoto, C., Terada, T., Unzai, S., Tame, J.R.H., Kuramitsu, S., Shirouzu, M., and Yokoyama, S. (2004). Crystal structure of elongation factor P from *Thermus thermophilus* HB8. Proc. Natl. Acad. Sci. USA *101*, 9595–9600. https://doi.org/10.1073/pnas.0308667101.
32. Choi, S., and Choe, J. (2011). Crystal structure of elongation factor P from *Pseudomonas aeruginosa* at 1.75 A resolution. Proteins *79*, 1688–1693. https://doi.org/10.1002/prot.22992.

33. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

34. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. *50*, D439–D444. https://doi.org/10.1093/nar/gkab1061.

35. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.

36. Hayter, J.R., Robertson, D.H.L., Gaskell, S.J., and Beynon, R.J. (2003). Proteome analysis of intact proteins in complex mixtures. Mol. Cell. Proteomics *2*, 85–95. https://doi.org/10.1074/mcp.M200078-MCP200.

37. Lee, S.W., Berger, S.J., Martinović, S., Pasa-Tolić, L., Anderson, G.A., Shen, Y., Zhao, R., and Smith, R.D. (2002). Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. Proc. Natl. Acad. Sci. USA *99*, 5942–5947. https://doi.org/10.1073/pnas.082119899.

38. Pfab, M., Kielkowski, P., Krafczyk, R., Volkwein, W., Sieber, S.A., Lassak, J., and Jung, K. (2021). Synthetic post-translational modifications of elongation factor P using the ligase EpmA. FEBS J. *288*, 663–677. https://doi.org/10.1111/febs.15346.

39. Reiling, N., Homolka, S., Walter, K., Brandenburg, J., Niwinski, L., Ernst, M., Herzmann, C., Lange, C., Diel, R., Ehlers, S., and Niemann, S. (2013). Clade-specific virulence patterns of *Mycobacterium tuberculosis* complex strains in human primary macrophages and aerogenically infected mice. mBio *4*, e00250-13. https://doi.org/10.1128/mBio.00250-13.

40. Bradshaw, E., Saalbach, G., and McArthur, M. (2013). Proteomic survey of the *Streptomyces coelicolor* nucleoid. J. Proteonomics *83*, 37–46. https://doi.org/10.1016/j.jprot.2013.02.033.

41. Wards, B.J., and Collins, D.M. (1996). Electroporation at elevated temperatures substantially improves transformation efficiency of slow-growing mycobacteria. FEMS Microbiol. Lett. *145*, 101–105. https://doi.org/10.1111/j.1574-6968.1996.tb08563.x.

42. Katz, A., Solden, L., Zou, S.B., Navarre, W.W., and Ibba, M. (2014). Molecular evolution of protein-RNA mimicry as a mechanism for translational control. Nucleic Acids Res. *42*, 3261–3271. https://doi.org/10.1093/nar/gkt1296.

43. Mudryi, V., Peske, F., and Rodnina, M. (2023). Translation factor accelerating peptide bond formation on the ribosome: EF-P and eIF5A as entropic catalysts and a potential drug targets. BBA Adv. *3*, 100074. https://doi.org/10.1016/j.bbadva.2023.100074.

44. Marman, H.E., Mey, A.R., and Payne, S.M. (2014). Elongation factor P and modifying enzyme PoxA are necessary for virulence of *Shigella flexneri*. Infect. Immun. *82*, 3612–3621. https://doi.org/10.1128/IAI.01532-13.

45. Guo, Q., Cui, B., Wang, M., Li, X., Tan, H., Song, S., Zhou, J., Zhang, L.H., and Deng, Y. (2022). Elongation factor P modulates *Acinetobacter baumannii* physiology and virulence as a cyclic dimeric guanosine monophosphate effector. Proc. Natl. Acad. Sci. USA *119*, e2209838119. https://doi.org/10.1073/pnas.2209838119.

46. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740. https://doi.org/10.1038/nmeth.3901.

47. Tyanova, S., and Cox, J. (2018). Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. Methods Mol. Biol. *1711*, 133–148. https://doi.org/10.1007/978-1-4939-7493-1_7.

48. Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. *30*, 918–920. https://doi.org/10.1038/nbt.2377.

49. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods *17*, 41–44. https://doi.org/10.1038/s41592-019-0638-x.

50. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. *27*, 14–25. https://doi.org/10.1002/pro.3235.

51. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. https://doi.org/10.1002/pro.3943.

52. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

53. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

54. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. *50*, D543–D552. https://doi.org/10.1093/nar/gkab1038.

55. Lassak, J., Henche, A.L., Binnenkade, L., and Thormann, K.M. (2010). ArcS, the cognate sensor kinase in an atypical Arc system of *Shewanella oneidensis* MR-1. Appl. Environ. Microbiol. *76*, 3263–3274. https://doi.org/10.1128/AEM.00512-10.

56. Keseler, I.M., Gama-Castro, S., Mackie, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Kothari, A., Krummenacker, M., Midford, P.E., Muñiz-Rascado, L., et al. (2021). The EcoCyc Database in 2021. Front. Microbiol. *12*, 711077. https://doi.org/10.3389/fmicb.2021.711077.

57. Miller, J.H. (1992). A Short Course in Bacterial Genetics : A Laboratory Manual and Handbook for *Escherichia coli* and Related Bacteria (Cold Spring Harbor Laboratory Press).

58. Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685. https://doi.org/10.1038/227680a0.

59. Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., and Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat. Protoc. *14*, 68–85. https://doi.org/10.1038/s41596-018-0082-x.

60. Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA *98*, 5116–5121. https://doi.org/10.1073/pnas.091062498.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti - *R. v.* EF-P antibodies (α-625)	This study	N/A
6x-His Tag Monoclonal Antibody	Invitrogen	RRID: AB_2536841
Goat anti-Mouse IgG	Abcam	Cat# ab216776; RRID: AB_2933974
Goat anti-Rabbit IgG	Abcam	Cat# ab216773; RRID:AB_2925189
Rabbit polyclonal anti - *E. coli* EF-P antibody	Pfab et al.[38]	N/A
Bacterial and virus strains		
See Data S3 (Sheet 'strain list')	This study	N/A
Chemicals, peptides, and recombinant proteins		
Chymotrypsin	Thermo Fisher Scientific	Cat# 90056
ortho-Nitrophenyl-β-galactoside	Carl Roth GmbH + Co. KG	2492-87-7
Critical commercial assays		
Q5® Site-Directed Mutagenesis Kit	New England BioLabs	Cat# E0554S
iST Kit	Preomics	P.O.00001
Deposited data		
Mass spectrometry proteomics data	ProteomeXchange	ProteomeXchange: PXD044929
Oligonucleotides		
See Data S3 (sheet 'primer list')	This study	N/A
Recombinant DNA		
See Data S3 (sheet 'plasmid list')	This study	N/A
See Data S3 (sheet 'genomic DNA list')	DSMZ	N/A
Software and algorithms		
Freestyle (Xtract Deconvolution algorithm)	Thermo Scientific	https://www.thermofisher.com/order/catalog/product/OPTON-30965?SID=srch-srp-OPTON-30965
Perseus 2.0.9.0	Tyanova et al.[46] Tyanova & Cox[47]	https://maxquant.net/perseus/
The MaxQuant 2.1.0.0	Cox et al.[29]	https://www.maxquant.org/
ProteoWizard	Chambers et al.[48]	http://www.proteowizard.org/download.html
Standalone DIA-NN software 1.8.1	Demichev et al.[49]	https://github.com/vdemichev/DiaNN
UCSF ChimeraX	Goddard et al.,[50] Pettersen et al.[51]	https://www.rbvi.ucsf.edu/chimerax
AlphaFold2 ColabFold 1.5.2	Jumper et al.,[33] Mirdita et al.[52]	N/A
ImageJ 1.54d	Schneider et al.[53]	https://imagej.net/software/imagej/
CLC Main Workbench 8.1.2	QIAGEN, Aarhus, Denmark	https://digitalinsights.qiagen.com/
Other		
Amicon® Ultra-15 Centrifugal Filter Units	Millipore	UFC9003
Carboxylate-coated magnetic beads (hydrophilic)	Cytiva	Cat# 45152105050250
Carboxylate-coated magnetic beads (hydrophobic)	Cytiva	Cat# 65152105050250
Ni-NTA Agarose	Qiagen	Cat# 30230
ZipTip with C4 resin	Millipore	Cat# ZTC04S096
PicoTipTM Emitter, Silica TipTM	New Objectives	FS360-75-8-N-20-C15
PEPMAP100 C18 5UM 0.3 × 5MM	Thermo Fisher Scientific	Cat# 160454
ReproSil-Pur 120 C18-AQ,1.9 μm	Dr. Maisch GmbH	Cat# r119.aq.0001
SERVAPOR® dialysis tubing	SERVA	Cat# 44146.04

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kirsten Jung (jung@lmu.de).

Materials availability

Materials generated in this study are available from the lead contact upon request with a completed Material Transfer Agreement.

Data and code availability

- Mass spectrometry and proteomics data are deposited to the ProteomeXchange Consortium via the PRIDE[54] partner repository (ProteomeXchange: PXD044929).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

All bacterial strains used in this study are listed in Data S3. *E. coli* was cultivated in lysogeny broth (LB) supplemented with antibiotics under agitation (750 rpm) at 37°C. Bacterial growth experiments were conducted in 50 mL glass flasks filled with 15 mL LB. For the spot assay, overnight cultures were resuspended in LB to an optical density (600 nm) (OD_{600}) of 0.01, spotted in dilutions (10^{-3} to 10^{-5}) on LB plates and grown for 18 h at 37°C. For CadC production, cells were grown in buffered LB at pH 5.8.[7] To investigate β-galactosidase activity on solid media, LB agar plates were supplemented with S-Gal® (300 mg/mL) and ferric ammonium citrate (500 mg/L). For EF-P overproduction, growth media were supplemented with L-arabinose [0.2% (w/v)] and chloramphenicol. Antibiotic concentrations used in this study: 34 μg/mL chloramphenicol, 50 μg/mL kanamycin sulfate.

METHOD DETAILS

Plasmid and bacterial strain construction

All primers and plasmids constructed in this study are shown in Data S3. All PCR reactions were conducted using the Q5 polymerase (New England BioLabs) according to manufacturer's instructions. Standard DNA restrictions were performed in rCutSmart buffer and the fragments were ligated into the corresponding vectors using the T4 ligase (New England BioLabs). *efp* genes from various species including a sequence coding for a C-terminal 6xHis tag were cloned in pBAD33. Plasmids were isolated using Hi Yield Plasmid Mini Kit (Sued Laborbedarf), whereas PCR fragments from the agarose gel were purified using the High-Yield PCR Cleanup and Gel Extraction Kit (New England BioLabs). Site-specific mutagenesis of *efp* was carried out with the Q5 Site-Directed Mutagenesis Kit (New England BioLabs) according to manufacturer's instructions.

E. coli mutants expressing *efp* variants were constructed using double homologous recombination with neomycin acetyltransferase and SacB as the selection or counterselection markers, respectively.[55] Overlapping regions of *E. coli efp* with the promoter of its modification enzyme *epmB* and the lipoprotein entericidin A (*ecnA*) were considered by keeping the 5′ and 3′ gene regions of *E. coli efp* present in the chromosome[56] (Figures 2A, S3A, and S3E). All nucleotide and protein sequences were analyzed using CLC Main Workbench 8.1.2 (Qiagen).

β-Galactosidase activity assays

Reporter strains MG1655 Δ*lacZ* P_{cadBA}:*lacZ* Δ*efp* and MG1655 Δ*lacZ* P_{cadBA}:*lacZ* Δ*efp* Δ*epmA* transformed with plasmids expressing *efp* and its variants were inoculated in 1.8 mL buffered LB pH 5.8 (91.5 mM KH_2PO_4; 8.5 mM K_2HPO_4; 0.2% (w/v) arabinose; chloramphenicol) and microaerobically grown under agitation (Thermomixer Comfort; 700 rpm) at 37°C overnight. The overnight culture was split into three tubes to measure optical density, β-galactosidase activity and detect EF-P in a Western Blot. For the β-galactosidase activity measurements, cells were harvested from 500 μL liquid culture by centrifugation and resuspended in 1 mL cooled (4°C) assay buffer (60 mM Na_2HPO_4; 40 mM NaH_2PO_4; 10 mM KCl; 1 mM $MgSO_4$; 50 mM β-mercaptoethanol; pH 7.0). Cells were permeabilized by adding 100 μL chloroform and 50 μL 0.1% sodium dodecyl sulfate (SDS). Prior to the measurements, all samples were incubated for 5 min at 37°C. The reaction was started by adding 0.2 mL of the substrate *ortho*-Nitrophenyl-β-galactoside (*o*-NPG, Carl Roth GmbH + Co. KG; 4 mg/mL in assay buffer). Total incubation time with the substrate was recorded when yellow color was developed until the reaction was stopped by adding 0.5 mL of 1 M Na_2CO_3. Samples were centrifuged and 1 mL of the supernatant was used for absorption measurements at 420 nm.[7,27] The measured β-galactosidase activity is given in Miller Units (MU), calculated according to Miller et al., 1992.[57]

Protein purification and structural analysis

For the recombinant *R. vannielii* EF-P production, *efp* was cloned into the pET expression plasmid and overproduced using the pET expression system in *E. coli* BL21 (DE3) (Invitrogen). Cells were harvested after cultivation in LB supplemented with IPTG (1 mM) at

18°C overnight, and the resulting pellet was frozen at −80°C. Cells with chromosomally incorporated His-tagged *efp* (*R. vannielii* EF-P and *W. virosa* EF-P) were grown in LB until the mid-exponential growth phase, harvested and kept at −80°C until further processing.

For lysis, cells were resuspended in 0.1 M, pH 7.6 sodium phosphate buffer (supplemented with 300 mM NaCl and DNase) and lysed using the high-pressure cell disrupter (Constant Systems), by running the sample twice under 1.9 kbar with subsequent cell fractionation using ultracentrifugation. The His-tagged proteins were purified using Ni-NTA Agarose beads (Qiagen) with 0.1 M, pH 7.6 sodium phosphate buffer supplemented with 300 mM NaCl and 30 mM–200 mM imidazole. Samples were dialyzed to remove imidazole at 4°C for 24 h using SERVAPORE dialysis tubing (MWCO 12000–14000, SERVA) according to manufacturer's instructions. Proteins were concentrated using the Amicon Ultra-15 Centrifugal Filter Units, 3kDa (Millipore).

The *R. v.* EF-P structure was predicted using AlphaFold2 ColabFold (v1.5.2).[33,52] All structural models were generated using UCSF ChimeraX.[50,51]

SDS-PAGE and Western Blot analysis

Proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)[58] using 12.5% Bis-tris acrylamide gels. Gels were stained with 2,2,2-trichloroethanol (TCE) to detect protein bands corresponding to a 72 kDa protein which served as loading controls. All proteins were transferred to a nitrocellulose membrane using the wet-blot apparatus (Mini Trans-Blot Cell, Bio-Rad). *E. c*. EF-P was detected by incubating the membrane in Tris-buffered saline (TBS) with primary polyclonal antibodies against *E. coli* EF-P (rabbit, Eurogentec)[38] with final concentration of 1:5,000, whereas His-tagged EF-P was detected with primary monoclonal antibodies against the His-Tag (1:10,000; RRID: AB_2536841, Invitrogen). Membranes were incubated with the secondary antibodies conjugated with a fluorophore (1:20,000 in TBS) (anti mouse: Cat# ab216776; anti rabbit: Cat# ab216773) and imaged using the Odyssey CLx (LI-COR Biosciences). For detection of *R. vannielii* EF-P, recombinantly produced EF-P was purified and sent to Eurogentec for polyclonal antibody generation (Speedy 28-day program in rabbits, Eurogentec). The blood serum containing antibodies against the *R. vannielii* EF-P (α 652) was diluted to final concentration of 1:100. Relative protein band intensities in Western blot analysis were calculated using ImageJ1.54d.[53]

Mass spectrometry for identification of modification status

For top-down EF-P measurements the purified proteins were desalted on the ZipTip with C4 resin (Millipore, ZTC04S096) and eluted with 50% (v/v) acetonitrile 0.1% (v/v) formic acid (FA) buffer resulting in ∼10 μM final protein concentration in 200–400 μL total volume. MS measurements were performed on an Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Fisher Scientific) via direct injection, a HESI-Spray source (Thermo Fisher Scientific) and FAIMS interface (Thermo Fisher Scientific) in a positive, peptide mode. Typically, the FAIMS compensation voltage (CV) was optimized by a continuous scan. The most intense signal was usually obtained at −5 CV. The MS spectra were acquired with at least 120,000 FWHM, AGC target 100 and 2–5 microscans covering the 800–1100 m/z range. The spectra were deconvoluted in Freestyle (Thermo Scientific) using the Xtract Deconvolution algorithm.

For bottom-up proteomics, samples were prepared in 96-well plate using the optimized SP3 protocol.[59] The purified EF-P sample (1 μL, 10 μM) was diluted to total volume of 50 μL with 1% NP40, 0.2% (w/v) SDS in 25 mM HEPES, pH 7.5. The protein was loaded onto a mixture of hydrophilic and hydrophobic carboxylate-coated magnetic beads (10 μL each) pre-washed three times with 100 μL of MS-grade H_2O. The magnetic beads with protein sample were mixed at 850 rpm, 1 min at room temperature (RT). To initiate the binding, 60 μL of absolute EtOH was added, and the mixture was incubated at RT for 5 min at 850 rpm. Subsequently, the beads were washed three times with 80% (v/v) EtOH, with incubation at RT for 1 min and 850 rpm between each wash. After the last wash the beads were resuspended in 50 μL of 100 mM ammonium acetate buffer (ABC). The proteins were reduced and alkylated by addition of 5 μL of 100 mM tris(2-carboxyethyl) phosphine (TCEP) and 5 μL of 400 mM chloroacetamide (CAA) and incubation at 95°C for 5 min, 850 rpm. Samples were cooled to RT. The on-beads digestion was performed with chymotrypsin. Chymotrypsin digestion: ABC buffer was supplemented with 10 mM $CaCl_2$, 1 μg of chymotrypsin (Thermo Scientific, 90056) and incubated at 25°C overnight. The resulting peptide mixture was eluted from the magnetic beads into a new 1.5 mL tube. The magnetic beads were washed with 50 and 30 μL of 1% (v/v) formic acid and incubated at 40°C, 850 rpm for 5 min. The fractions were added to the first elution fraction. The combined fractions were further purified from remaining magnetic beads. Alternatively, for *R. v.* EF-P overproduced in *E. coli*, the EF-P was directly digested in ABC buffer (50 μL) supplemented with 10 mM $CaCl_2$ and 1 μg of chymotrypsin (Thermo Scientific, 90056), without previous clean-up on magnetic beads.

MS measurements were performed on an Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Fisher Scientific) coupled to an UltiMate 3000 Nano-HPLC (Thermo Fisher Scientific) via a nanospray Flex ion source (Thermo Fisher Scientific) equipped with column oven (Sonation) and FAIMS interface (Thermo Fisher Scientific). Peptides were loaded on an Acclaim PepMap 100 μm-precolumn cartridge (5 μm, 100 Å, 300 μm ID x 5 mm, Thermo Fisher Scientific) and separated at 40°C on a PicoTip emitter (noncoated, 15 cm, 75 μm ID, 8 μm tip, New Objective) that was *in-house* packed with Reprosil-Pur 120 C18-AQ material (1.9 μm, 150 Å, Dr. A. Maisch GmbH). Buffer composition. Buffer A consists of MS-grade H_2O supplemented with 0.1% FA. Buffer B consists of acetonitrile supplemented with 0.1% FA. The 41- or 126-min LC gradient from 4 to 35.2% buffer B was used. The flow rate was 0.3 μL/min.

Data-independent acquisition

The DIA duty cycle consisted of one MS1 scan followed by 30 MS2 scans with an isolation window of the 4 m/z range, overlapping with an adjacent window at the 2 m/z range. MS1 scan was conducted with Orbitrap at 60000 resolution power and a scan range of 200–1800 m/z with an adjusted RF lens at 30%. MS2 scans were conducted with Orbitrap at 30000 resolution power, RF lens was set

to 30%. The precursor mass window was restricted to a 500–740 m/z range. HCD fragmentation was enabled as an activation method with a fixed collision energy of 35%. FAIMS was performed with one CV at −45V for both MS1 and MS2 scans during the duty cycle.

Data-dependent acquisition

(*R. v.* EF-P overproduced in *E. coli*) For measurements of DDA, the Orbitrap Eclipse Tribrid Mass Spectrometer was operated with the following settings: Polarity: positive; MS1 resolution: 240k; MS1 AGC target: standard; MS1 maximum injection time: 50 ms; MS1 scan range: m/z 375–1500; MS2 ion trap scan rate: rapid; MS2 AGC target: standard; MS2 maximum injection time: 35 ms; MS2 cycle time: 1.7 s; MS2 isolation window: m/z 1.2; HCD stepped normalised collision energy: 30%; intensity threshold: 1.0e4 counts; included charge states: 2–6; dynamic exclusion: 60 s. FAIMS was performed with two alternating CVs, including −50 V and −70 V.

Mass spectrometry for identification of modification status

Computational evaluation of DIA raw files

Raw files were converted in the first step with "MSConvertGUI" as a part of the "ProteoWizard" software[48] package (http://www.proteowizard.org/download.html) to an output mzML format applying the "peakPicking" filter with "vendor msLevel = 1", and the "Demultiplex" filter with parameters "Overlap Only" and "mass error" set to 10 ppm.

Standalone DIA-NN software under version 1.8.1 was used for protein identification and quantification[49]

First, a spectral library was predicted *in silico* by the software's deep learning-based spectra, RTs and IMs prediction using Uniprot *E. coli* decoyed FASTA (canonical and isoforms with added *R. vannielii* EF-P sequence). DIA-NN search settings: FASTA digest for library-free search/library generation option was enabled, together with a match between runs (MBR) option and precursor FDR level set at 1%. Library generation was set to smart profiling, Quantification strategy - Robust LC. The mass accuracy and the scan window were set to 0 to allow the software to identify optimal conditions. The precursor m/z range was changed to 500–740 m/z to fit the measuring parameters. Carbamidomethylation was set as a fixed modification, oxidation of methionine and N-term acetylation were set as variable modifications. On the contrary, the small-scale samples of the 96-well plate were calculated without carbamidomethylation as a fixed modification.

Computational evaluation of DDA raw files

MS Raw files were analyzed using MaxQuant software. Searches were performed against the Uniprot database for *E. coli* (*R. vannielii* EF-P sequence). At least two unique peptides were required for protein identification. False discovery rate determination was carried out using a decoy database and thresholds were set to 1% FDR both at peptide-spectrum match and at protein levels.

Mass spectrometry for proteome analysis

Cells were cultivated in LB under constant shaking at 37°C until reaching the exponential growth phase. Cells were harvested (OD_{600} = 0.5) and were processed with the iST kit (Preomics) as recommended by the manufacturer. Samples were evaporated to dryness, resuspended in LC-LOAD buffer to 0.2μg/μL and injected in an Ultimate 3000 RSLCnano system (Thermo) separated in a 25-cm Aurora column (Ionopticks) with a 100-min gradient from 4 to 40% acetonitrile in 0.1% formic acid. The effluent from the HPLC was directly electrosprayed into an Orbitrap Exploris 480 (Thermo) operated in data dependent mode to automatically switch between full scan MS and MS/MS acquisition. Survey full scan MS spectra (from m/z 350–1200) were acquired with a resolution of R = 60,000 at m/z 400 (AGC target of $3x10^6$). The 20 most intense peptide ions with charge states between 2 and 6 were sequentially isolated to a target value of $1x10^5$ and fragmented at 30% normalized collision energy. Typical mass spectrometric conditions were: spray voltage, 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 275°C; intensity selection threshold, $3x10^5$.

The MaxQuant 2.1.0.0 software was used for protein identification and quantification by label-free quantification (LFQ)[29] with the following parameters: Database Uniprot_UP000000625_Ecoli_20220309.fasta including the *Rhodomicrobium vannielii* EFP sequence; MS tol, 10 ppm; MS/MS tol, 20 ppm Da; Peptide FDR, 0.1; Protein FDR, 0.01 min; Peptide Length, 7; Variable modifications, Oxidation (M); Fixed modifications, Carbamidomethyl (C); Peptides for protein quantitation, razor and unique; Min. peptides, 1; Min. ratio count, 2. For display and analysis, the Perseus software[46,47] was used. A list as a reference for proteins with polyproline-motifs in *E. coli* was taken from Qi *et* al..[13] Data have been uploaded to the PRIDE repository[54] (ProteomeXchange: PXD044929).

QUANTIFICATION AND STATISTICAL ANALYSIS

All measurements, except intact protein mass measurements and chymotrypsin digestions, are from at least three biological replicates. Statistical analysis was done using Microsoft Office *Excel* 2019 (student's unpaired two-sided t test). Error bars in all growth curves and bar graphs represent the standard deviation (SD). Values were considered as significantly different when the calculated p *value* was below 0.05.

Perseus (2.0.9.0) was used to $\log_2$ transform LFQ intensities, replace missing values from normal distribution and construct the volcano plots. To determine which proteins were differentially expressed between experimental conditions, we applied a t test with a permutation-based FDR calculation with n = 3, FDR = 0.05 and s0 = 0.1 to the log 2 LFQ protein values, where s0 controls the relative importance of t test p value and difference between means. At s0 = 0 only the p value matters, while at nonzero s0 also the difference of means plays a role.[60]

Cell Reports, Volume *43*

Supplemental information

Decrypting the functional design of unmodified translation elongation factor P

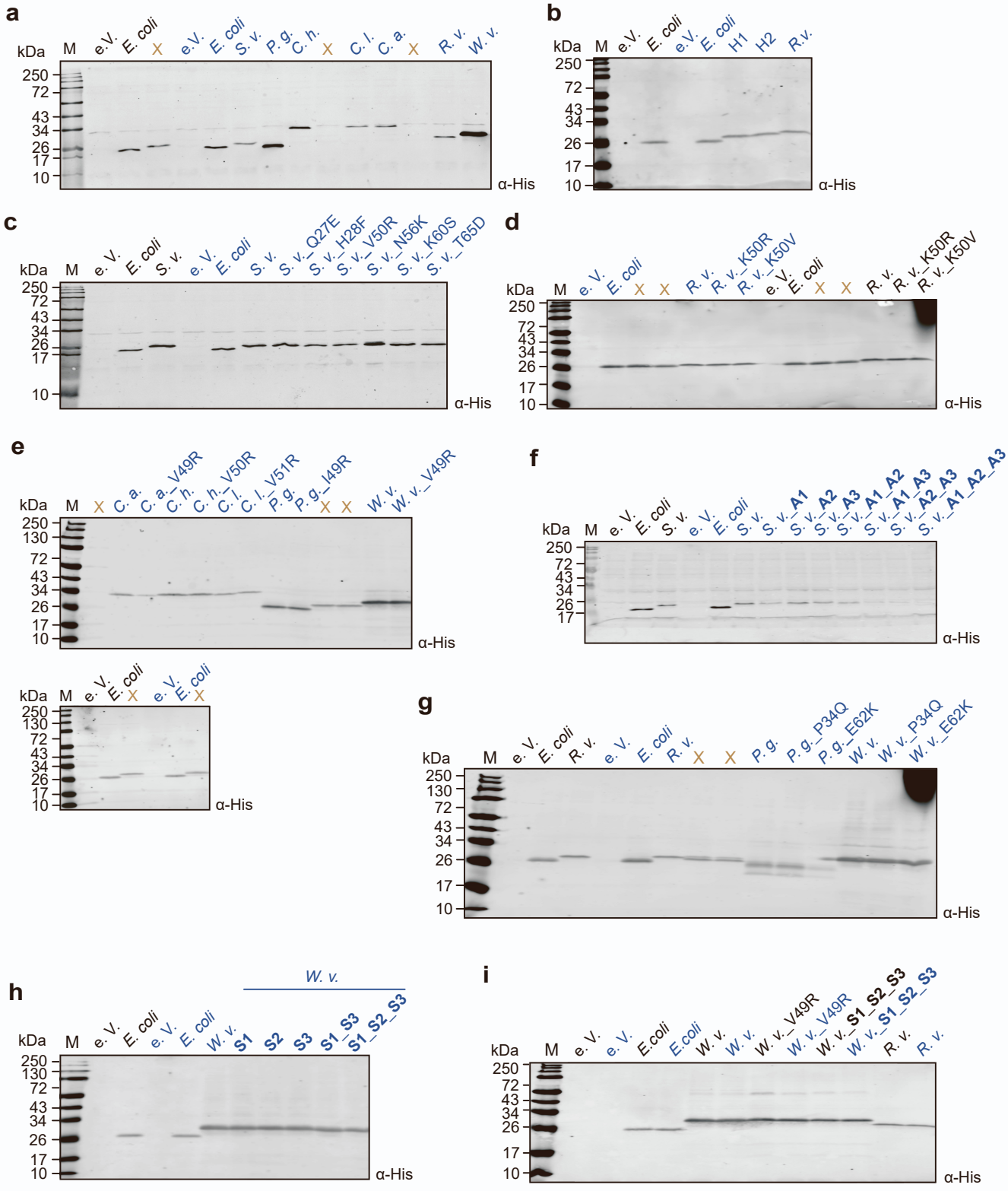
Urte Tomasiunaite, Pavel Kielkowski, Ralph Krafczyk, Ignasi Forné, Axel Imhof, and Kirsten Jung

Supplementary Table S1. Protein sequence identity of *E. coli* EF-P and EF-Ps of the PGKGP-subfamily. Related to Figure 1.

Multiple sequence alignment and percent identity matrix were calculated using the Multiple Sequence Alignment Tool from Clustal Omega (Clustal 2.1) and is given in percentage (%).

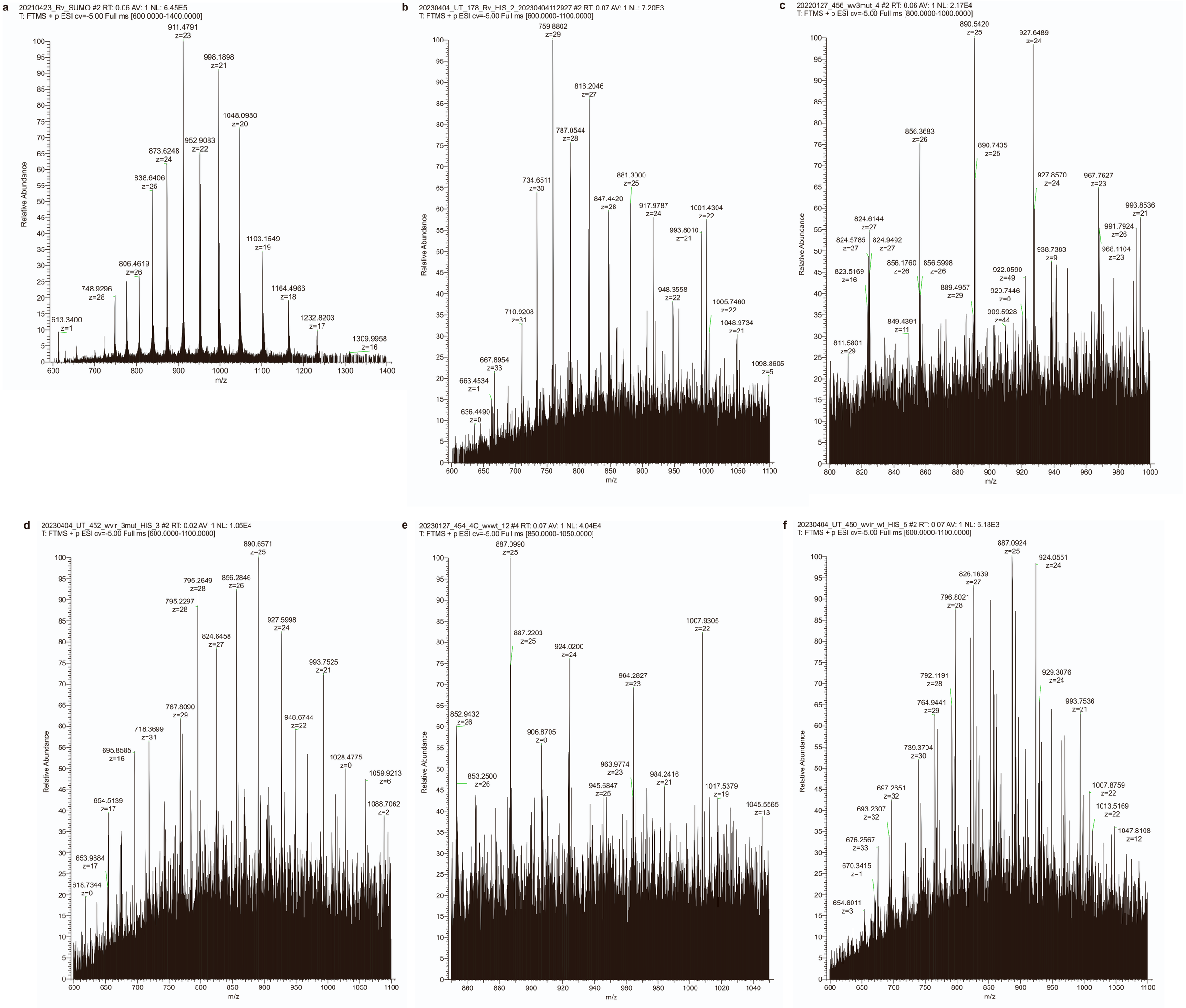
	E. c.	***C. h.***	***C. l.***	***C. a.***	***C. w.***	***P. g.***	***R. v.***	***S. v.***	***W. v.***
E. c.		44.92	44.15	39.13	40.54	41.08	32.09	41.94	37.84
C. h.	44.92		88.83	40.86	44.86	43.01	40.64	43.32	40.86
C. l.	44.15	88.83		40.86	45.95	40.86	37.97	42.25	40.32
C. a.	39.13	40.86	40.86		44.86	53.19	32.97	43.01	66.49
C. w.	40.54	44.86	45.95	44.86		45.65	35.14	46.49	50.54
P. g.	41.08	43.01	40.86	53.19	45.65		35.68	49.46	57.98
R. v.	32.09	40.64	37.97	32.97	35.14	35.68		36.56	31.89
S. v.	41.94	43.32	42.25	43.01	46.49	49.46	36.56		41.94
W. v.	37.84	40.86	40.32	66.49	50.54	57.98	31.89	41.94	

E. c. - *Escherichia coli*
C. h. - *Campylobacter hominis*
C. l. - *Campylobacter lari*
C. a. - *Cellulophaga algicola*
C. w. - *Conexibacter woesei*
P. g. - *Porphyromonas gingivalis*
R. v. - *Rhodomicrobium vannielii*
S. v. - *Streptomyces venezuelae*
W. v. - *Weeksella virosa*



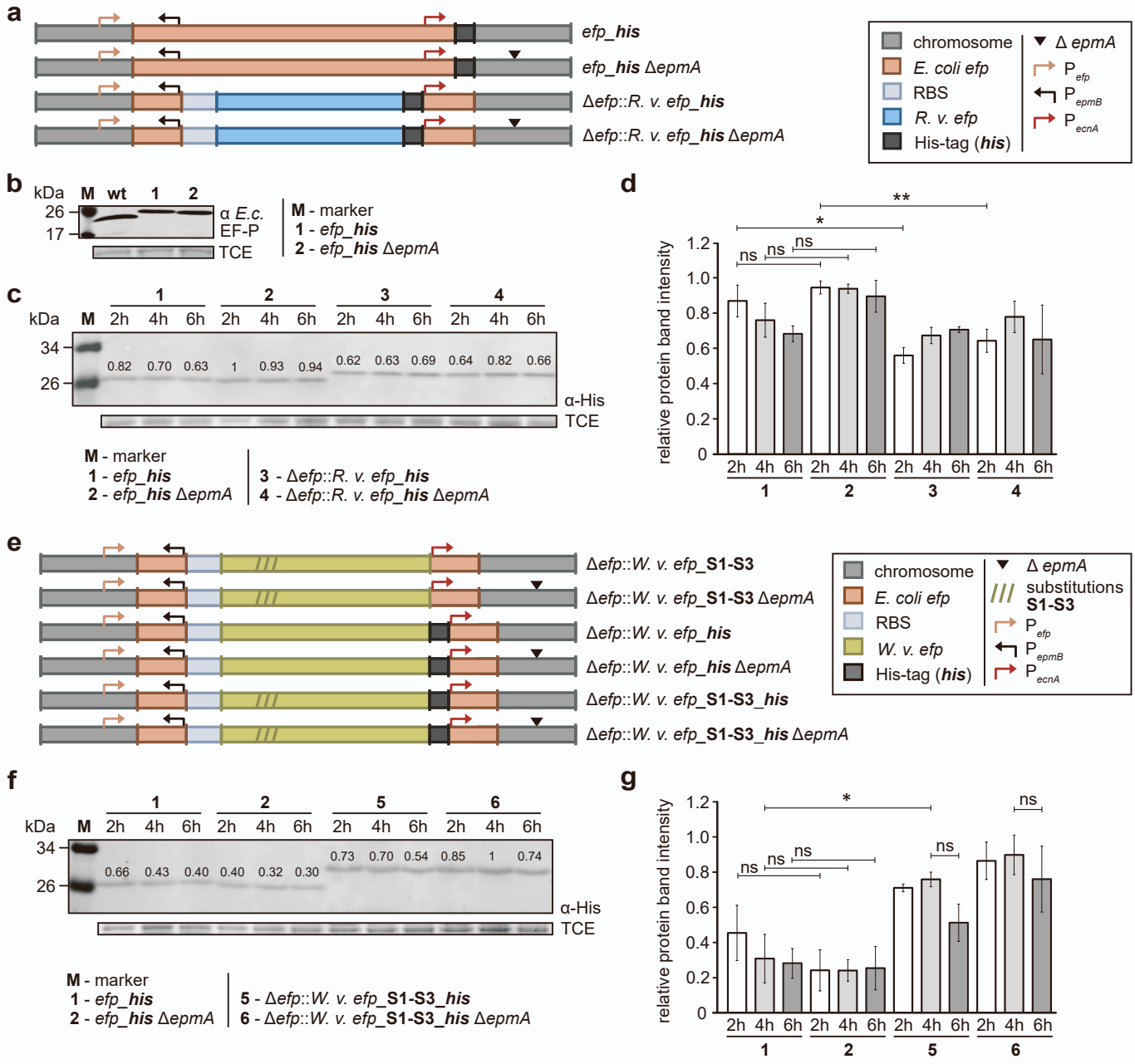
Supplementary Figure S1. Western blot analysis of His-tagged EF-P variants in *E. coli*. Related to Figures 1, 3, 4 and 5.

Uncropped Western blot membranes related to Fig. 1c (**a**), Fig. 3b (**b**), Fig. 4c (**c**), Fig. 4d (**d**), Fig. 4e (**e**), Fig. 5a (**f**), Fig. 5b (**g**), Fig. 5c (**h**) and Fig. 5d (**i**). EF-P production was confirmed by Western Blot analysis using antibodies against the His-tag. Lanes labelled with "X" correspond to samples not shown in the cropped images.



Supplementary Figure S2. Mass spectrometry analysis of EF-Ps produced in *E. coli*. Related to Figures 1 and 5.

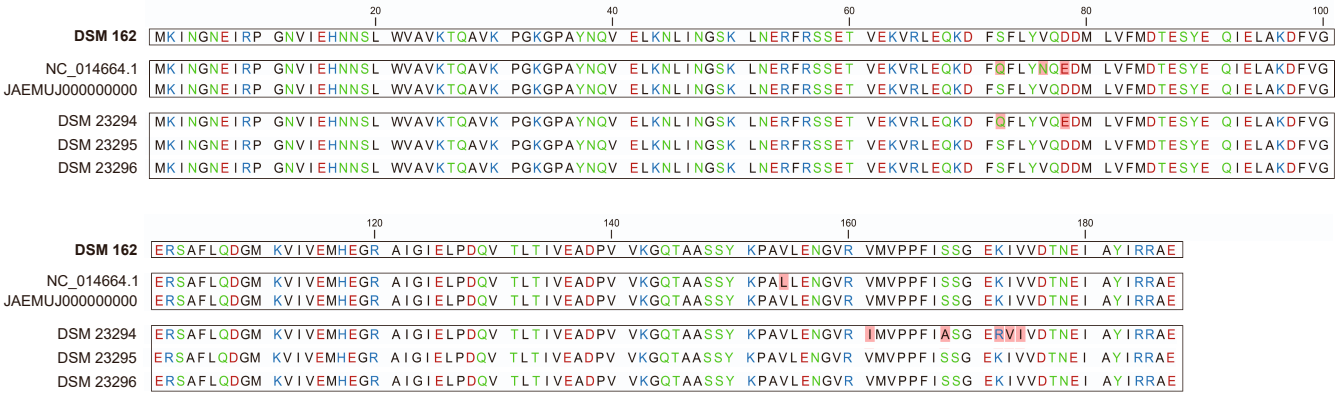
a Intact mass spectra of recombinantly overproduced *R. vannielii* EF-P, purified from *E. coli* wt. **b** Intact mass spectra of His-tagged *R. vannielii* EF-P (chromosomally encoded), purified from *E. coli* wt. **c** Intact mass spectra of His-tagged *W. virosa* EF-P variant with substitutions S1-S3 (chromosomally encoded), purified from *E. coli* Δ*epmA*. **d** Intact mass spectra of His-tagged *W. virosa* EF-P variant with substitutions S1-S3 (chromosomally encoded), purified from *E. coli* wt. **e** Intact mass spectra of His-tagged *W. virosa* EF-P wt (chromosomally encoded), purified from *E. coli* Δ*epmA*. **f** Intact mass spectra of His-tagged *W. virosa* EF-P wt (chromosomally encoded), purified from *E. coli* wt. S – substitution (S1 – P34Q; S2 – V49R; S3 – E62K).



Supplementary Figure S3. Protein production of His-tagged EF-P variants in *E. coli*. Related to Figures 1, 2 and 5.

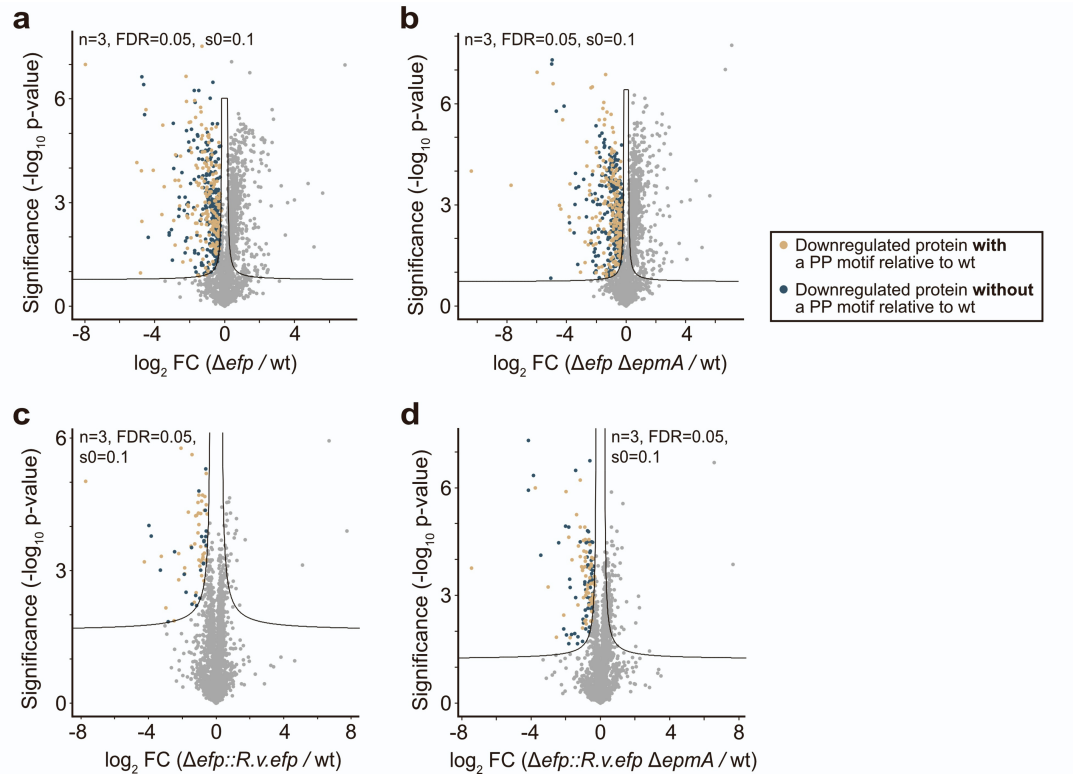
a Schematic overview of the constructed *E. coli* mutants expressing *his*-tagged *efp* or *R. v. efp*. **b** Western blot analysis of EF-P-His production in *E. coli* mutants after cells were grown in LB medium to mid-exponential phase using antibodies against *E. coli* EF-P. **c** Western blot analysis of *R. v.* EF-P production in *E. coli* mutants after cells were grown in LB medium to early-exponential (2h), mid-exponential (4h) and early-stationary (6h) phase using antibodies against the His-tag. **d** Quantification of His-tagged *E. coli* and *R.vannielii* EF-P production in cells of three independent biological replicates. The numbers in bold correspond to the strains used in **c**. **e** Schematic overview of the constructed *E. coli* mutants expressing *his*-tagged *efp* variants of *W. virosa*. **f** Western blot analysis of *W. virosa* EF-P variant production in *E. coli* mutants after cells were grown in LB medium to early-exponential (2h), mid-exponential (4h) and early-stationary (6h) phase using antibodies against the His-tag. **g** Quantification of His-tagged *E. coli* and *W. virosa* EF-P variant production in cells of three independent biological replicates. The numbers in bold correspond to the strains used in **f**. Relative band intensities calculated using ImageJ are displayed above the corresponding protein bands. Protein bands corresponding to a 72 kDa protein after staining with 2,2,2-trichloroethanol (TCE) were used as loading controls. Gene insertions in **a** and **e** are depicted in coloured

boxes, and promoter locations as coloured arrows. S1-S3 - amino acid substitutions (S1 – P34Q; S2 – V49R; S3 – E62K). Statistics: error bars represent the standard deviation (SD) of at least three independent biological replicates; student's unpaired two-sided t-test (****p < 0,0001; ***p < 0,001; **p <0,01; *p < 0,05; ns p > 0,05). 2h (Δ*efp_his* vs. Δ*efp_his* Δ*epmA*, ns p = 0.3557; Δ*efp_his* vs. Δ*efp*::*R. v.efp_his*, *p = 0.0232; Δ*efp_his* Δ*epmA* vs. Δ*efp*::*R. v. efp_his* Δ*epmA*, **p = 0.0092), 4h (Δ*efp_his* vs.Δ*efp_his* Δ*epmA*, ns p = 0.1104), 6h (Δ*efp_his* vs. Δ*efp_his* Δ*epmA*, ns p = 0.0608) (**d**). 2h (Δ*efp_his* vs.Δ*efp_his* Δ*epmA*, ns p = 0.2041), 4h (Δ*efp_his* vs. Δ*efp_his* Δ*epmA*, ns p = 0.5752; Δ*efp_his* vs. Δ*efp*::*W. virosa_efp*_S1-S3_*his*, *p = 0.0350, 6h (Δ*efp_his* vs. Δ*efp_his* Δ*epmA*, ns p = 0.7399), 4h vs. 6h (Δ*efp*:: *W. virosa_efp*_S1-S3_*his,* ns p = 0.0657; Δ*efp*:: *W. virosa_efp*_S1-S3_*his* Δ*epmA,* ns p = 0.4331) (**g**).



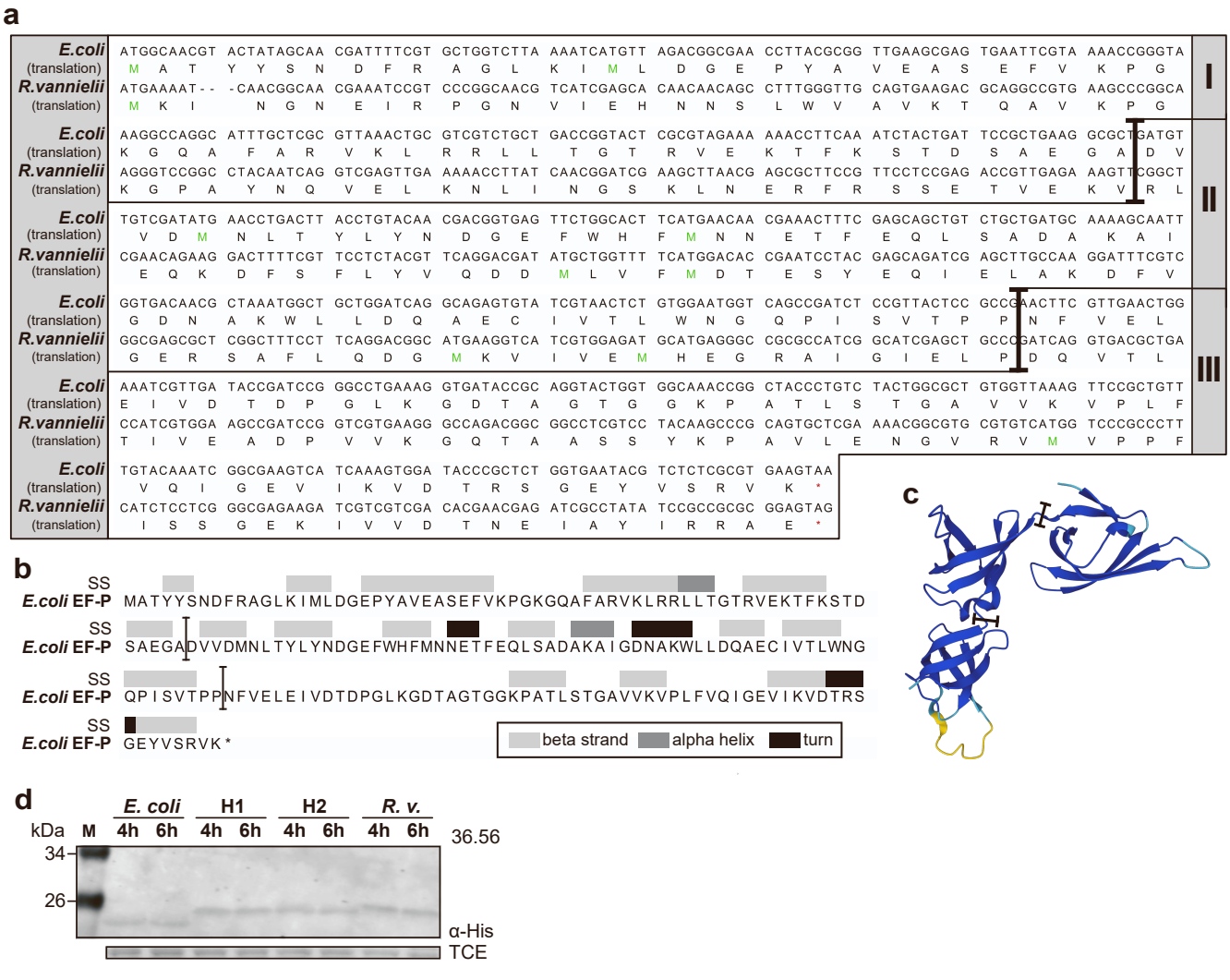
Supplementary Figure S4. Comparison of the sequence of EF-Ps from different *R. vannielii* strains and the publicly available sequences. Related to Figure 1.

Different colors represent the polarity of the amino acids (black- hydrophobic, light green-hydrophilic, red-acidic, blue-basic). Differences in amino acids between strains are highlighted in pink. Discrepancies between the EF-P sequences of the type strain DSM 162 used in this study and the publicly available sequence of *R. vannielii* ATCC171000 (GenBank accession number NC_014664.1) were ruled out by comparison with the revised EF-P sequence of *R. vannielii* ATCC171000 (GenBank accession number JAEMUJ000000000.1) [S1], and the sequencing results of EF-P from evolutionarily related strains (DSM 23294, DSM 23295, and DSM 23296).



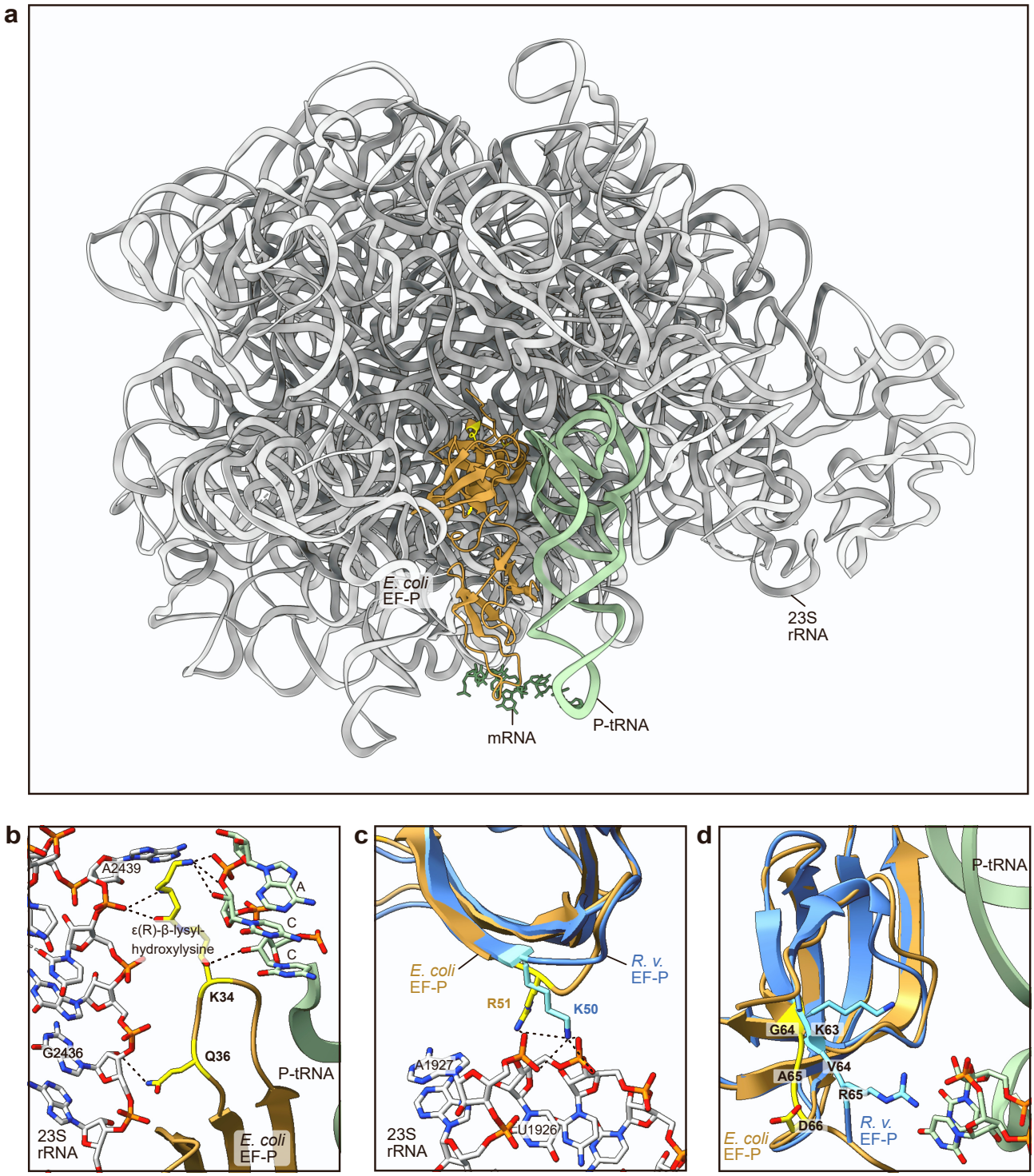
Supplementary Figure S5. Proteome-wide analysis of polyproline-containing proteins in the *E. coli* mutants Δ*efp,* Δ*efp* Δ*epmA* after complementation with *R. v. efp*. Related to Figure 2.

a-d Volcano plot analysis highlight the downregulation of proteins with/without polyproline motifs (PP-motifs) in *E. coli* mutants Δ*efp* (**a**), Δ*efp* Δ*epmA* (**b**), Δ*efp*::*R. v. efp* (**c**) and Δ*efp*::*R. v. efp* Δ*epmA* (**d**) compared to wild type (wt). The x-axes show for each protein the fold change (FC) of the mean value of the $\log_2$ protein intensity (LFQ) between two strains. The y-axes show the significance level of the observed difference between the two strains ($-\log_{10}$ p-value of the t-test). The test was adjusted for multiple comparisons (permutation-based FDR with n=3, FDR=0.05 and s0=0.1).



Supplementary Figure S6. Construction principles of *R. vannielii* EF-P and *E. coli* EF-P hybrids. Related to Figure 3.

a Nucleotide and protein sequence alignment of *E. coli* EF-P and *R. vannielii* EF-P with selected boundaries (black bulk arrows) for each EF-P domain (I-III). **b** Secondary structure and **c** 3D structural model of *E. coli* EF-P with selected boundaries between all domains (I-III) for hybrid construction (adapted from UniProt P0A6N4, and AlphaFold AF-P0A6N4-F1). **d** Western blot analysis of EF-P hybrid production in cells grown in LB medium to mid-exponential (4h) and early-stationary (6h) phase using antibodies against the His-tag. Protein bands corresponding to a 72 kDa protein after staining with 2,2,2-trichloroethanol (TCE) were used as loading controls. Structure prediction and the confidence score (pLDDT) were calculated by AlphaFold. 3D model confidence: very high – in dark blue (pLDDT > 90); confident – in light blue (90 >pLDDT> 70); low – yellow (70 > pLDDT >50); very low – red (pLDDT < 50).


a
E. coli
EF-P
23S
rRNA
mRNA
P-tRNA
b
A2439
A
C
C
ε(R)-β-lysyl-
hydroxylysine
K34
G2436
Q36
P-tRNA
23S
rRNA
E. coli
EF-P
c
E. coli
EF-P
R. v.
EF-P
R51
K50
A1927
U1926
23S
rRNA
d
P-tRNA
G64
K63
A65
V64
R65
D66
R. v.
EF-P
E. coli
EF-P

Supplementary Figure S7. Model of potential interactions between *E. coli* / *R. vannielii* EF-P and 23S rRNA / P-site-tRNA in context of amino acid substitutions P35Q, V50R, and 63-65_KVR in *S. venezuelae* EF-P. Related to Figure 5.

The Cryo-EM structure of polyproline-stalled ribosome in the presence of *E. coli* EF-P was adopted from Huter *et al.*, 2017 [S2] (PDB accession number: 6ENU). Structural models were generated using UCSF ChimeraX [S3,S4]. The *R. v.* EF-P structure was predicted using AlphaFold2 ColabFold (v1.5.2) [S5,S6]. **a** Schematic presentation of *E. coli* EF-P (brown), localized relative to the ribosome (23S rRNA) (grey), mRNA (dark green) and P-site-bound $tRNA^{Pro}$ (light green). **b** Potential interaction between Q36 of *E. coli* EF-P and 23S rRNA (G2436) is indicated with a dashed line. As described before [S2] the interactions of ε(R)-ß-lysyl-hydroxylysine (after post translational modification of K34) with the CCA end of the P-site tRNA (P-tRNA) and the 23S rRNA (A2439) are shown in dashed lines. **c** Potential interactions between R51 of *E. coli* EF-P or K50 of *R. v.* EF-P with the phosphate backbone of the 23S rRNA (U1926, A1927), indicated in dashed lines. Domain I of the predicted *R. v.* EF-P was superimposed with domain I of the cryo-EM structure of *E. coli* EF-P. **d** Potential interactions between the motif KVR (positions 63-65 in *R. v.* EF-P) and the P-tRNA. For comparison, the position of the corresponding GAD motif (amino acids 64-66) in *E. coli* EF-P is marked. Amino acids K34, Q36, R51, G64, A65 and D66 in *E. coli* EF-P correspond to K33, P35, V50, T63, A64 and T65, respectively, in *S. venezuelae* EF-P (**Fig. 4a**).

References

S1. Conners, E.M., Davenport, E.J., and Bose, A. (2021). Revised Draft Genome Sequences of *Rhodomicrobium vannielii* ATCC 17100 and *Rhodomicrobium udaipurense* JA643. Microbiol Resour Announc 10. 10.1128/MRA.00022-21.

S2. Huter, P., Arenz, S., Bock, L.V., Graf, M., Frister, J.O., Heuer, A., Peil, L., Starosta, A.L., Wohlgemuth, I., Peske, F., et al. (2017). Structural Basis for Polyproline-Mediated Ribosome Stalling and Rescue by the Translation Elongation Factor EF-P. Molecular Cell 68, 515-+. 10.1016/j.molcel.2017.10.014.

S3. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci 27, 14-25. 10.1002/pro.3235.

S4. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci 30, 70-82. 10.1002/pro.3943.

S5. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-+. 10.1038/s41586-021-03819-2.

S6. Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat Methods 19, 679-682. 10.1038/s41592-022-01488-1.