# The problem with proline – Regulation of ribosome stalling by translation factors and codon choice

## DISSERTATION

Zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

der Fakultät für Biologie
der Ludwig-Maximilians-Universität München
vorgelegt von

Alina Maria Sieber
München, Januar 2025

Gutachter 1: PD Dr. Jürgen Lassak

Gutachter 2: Prof. Dr. Dmitrij Frishman

Datum der Abgabe: 29.01.2025

Datum der mündlichen Prüfung: 28.05.2025

## Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass die vorliegende Dissertation von mir selbstständig verfasst wurde und dass keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden. Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinne nach entnommen sind, wurden in jedem Fall unter Angabe der Quellen (einschließlich des World Wide Web und anderer elektronischer Text- und Datensammlungen) kenntlich gemacht. Weiterhin wurden alle Teile der Arbeit, die mit Hilfe von Werkzeugen der künstlichen Intelligenz *de novo* generiert wurden, durch Fußnote/Anmerkung an den entsprechenden Stellen kenntlich gemacht und die verwendeten Werkzeuge der künstlichen Intelligenz gelistet. Diese Erklärung gilt für alle in der Arbeit enthaltenen Texte, Graphiken, Zeichnungen, Kartenskizzen und bildliche Darstellungen. Des Weiteren erkläre ich, dass ich nicht anderweitig ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen. Die folgende Dissertation liegt weder ganz, noch in wesentlichen Teilen einer anderen Prüfungskommission vor.

München, 29.01.2025 Alina Sieber

## Statutory Declaration

Herewith I certify under oath that I wrote the accompanying Dissertation myself. In the thesis no other sources and aids have been used than those indicated. The passages of the thesis that are taken in wording or meaning from other sources have been marked with an indication of the sources (including the World Wide Web and other electronic text and data collections). Furthermore, all parts of the thesis that were *de novo* generated with the help of artificial intelligence tools were identified by footnotes/annotations at the appropriate places and the artificial intelligence tools used were listed. This statement applies to all text, graphics, drawings, sketch maps, and pictorial representations contained in the Work. As well I declare, that I have not submitted a dissertation without success and not passed the oral exam. The present dissertation (neither the entire dissertation nor parts) has not been presented to another examination board.

Munich, 29.01.2025 Alina Sieber

## Contents

## Nomenclature

All genes are written in italics, all proteins are written with a first capital letter. Gene deletions are indicated by the symbol Δ, gene fusions are indicated by the symbol : and gene replacements are indicated by the symbol :: . Promotor regions of an indicated gene are represented by $P_{gene}$. Gene products are numbered such that the first methionine of the wild-type protein, when present, is assigned position "1" in the amino acid sequence, independent of any N-terminal affinity tag. Structural elements in proteins are numbered according to their order (e.g. helix h29) Amino acids are designated using their single-letter code, followed by their specific position within the primary sequence (e.g. Y180), and if not mention differently, X is a placeholder for any amino acid (e.g. PPX for proline-proline-amino acid). Codons are designated using their single-letter abbreviation, with N being a placeholder for any nucleotide (e.g. GGN for guanosine-guanosine-nucleotide).

## Publications originating from this Thesis

### Chapter 2:

Krafczyk R*, Qi F*, Sieber A, Mehler J, Jung K, Frishman D & Lassak J. 2021. **Proline codon pair selection determines ribosome pausing strength and translation efficiency in bacteria**. *Commun. Biol.* **4**, 589

### Chapter 3:

Sieber A*, Parr M*, von Ehr J, Dhamotharan K, Kielkowski P, Brewer T, Schäpers A, Krafczyk R, Qi F, Schlundt A, Frishman D & Lassak J. 2024. **EF-P and its paralog EfpL (YeiP) differentially control translation of proline-containing sequences.** *Nat. Commun.* **15**, 10465

*Authors contributed equally

### Further publications not included in the dissertation

### Review article

Lassak J, Sieber A, Hellwig M. 2022. **Exceptionally versatile take II: post-translational modifications of lysine and their impact on bacterial physiology**. *Biol Chem*. **403(**8-9):819-858.

## Contributions to Publications presented in this Thesis

### Chapter 2:

The HisL*-Lux reporter concept was developed by Jürgen Lassak. Ralph Krafczyk, Alina Sieber, Judith Mehler, and Jürgen Lassak constructed HisL*-lux reporter strains and plasmids. Corresponding *in vivo* measurements were performed by Ralph Krafczyk, Alina Sieber and Judith Mehler. tRNA gene deletions were made by Ralph Krafczyk and Judith Mehler. Reporters for quantification of +1 frameshifting were constructed and corresponding assay performed by Ralph Krafczyk and Judith Mehler. Ralph Krafczyk constructed *cadC* mutations and recorded their effect on pH-regulation. All bioinformatic analyses were performed by Fei Qi and Dmitrij Frishman. The study was designed by Fei Qi, Ralph Krafczyk, Dmitrij Frishman and Jürgen Lassak with contributions of Kirsten Jung. The manuscript was written by Fei Qi, Ralph Krafczyk, and Jürgen Lassak with contributions of Dmitrij Frishman and Kirsten Jung.

### Chapter 3:

Alina Sieber, Ralph Krafczyk, Anna Schäpers and Jürgen Lassak constructed strains and plasmids. Alina Sieber and Ralph Krafczyk performed the biochemical *in vivo* / *in vitro* characterization of EfpL. Anna Schäpers performed qPCR experiments. Alina Sieber purified all proteins used for *in vitro* assays, MS analysis and X-ray crystallography. MS experiments and analysis was done by Pavel Kielkowski. The crystallization screen was set up by Julian von Ehr, Karthikeyan Dhamotharan and Andreas Schlundt, and Julian von Ehr, Karthikeyan Dhamotharan and Andreas Schlundt solved the crystal structure of $EfpL_{E.\ coli}$. Andreas Schlundt performed *in silico* interaction analyses. All bioinformatic analyses were performed by Marina Parr, Tess Brewer, Fei Qi and Dmitrij Frishman with contributions of Alina Sieber and Jürgen Lassak. Ribosome profiling analyses were done by Marina Parr and Dmitrij Frishman. Marina Parr and Dmitrij Frishman performed phylogenetic analyses of the EF-P subgroups and Tess Brewer performed phylogenetic analyses of bacterial growth rates. The study was designed by Jürgen Lassak with contributions from Ralph Krafczyk and Dmitrij Frishman. The manuscript was written by Alina Sieber, Marina Parr and Jürgen Lassak with contributions from Andreas Schlundt, Tess Brewer and Dmitrij Frishman.

We hereby confirm the above statements:

Alina Sieber Jürgen Lassak

**Chapter 3:**

To achieve a comprehensive analysis of the EF-P paralog EfpL, Alina Sieber and Marina Parr combined biochemical and phylogenetic approaches to characterize the protein. Alina Sieber performed the biochemical *in vivo* and *in vitro* analysis, demonstrating that EfpL and EF-P have overlapping functions. The extensive phylogenetic analyses were conducted by Marina Parr, demonstrating unique features in the structure of EfpL. Alina Sieber purified the proteins for structural and modification analysis. She also performed the subsequent confirmation of EfpL's deactivation through acylation. Alina Sieber prepared the samples for ribosome profiling, and Marina Parr analyzed the sequencing data. With this analysis the EfpL arrest motif spectrum was detected and additional sequences beyond the already described motifs, that both EF-P and EfpL can resolve, were found. Together, Alina Sieber and Marina Parr elucidated the role of EfpL in ribosome rescue and translational stress response.

We hereby confirm the above statements:

Alina Sieber

Marina Parr

Jürgen Lassak

Dmitrij Frishman

## Zusammenfassung

Die bakterielle Translation ist ein essenzieller und hochkomplexer Prozess, der von zahlreichen Faktoren beeinflusst wird. Prolinreiche Sequenzen stellen hierbei eine besondere Herausforderung dar, da Proline aufgrund ihrer Ringstruktur nur langsam in Polypeptidketten eingebaut werden, und somit einen Ribosomenarrest auslösen können. Dennoch sind Polyprolinmotive aufgrund ihrer strukturellen und funktionellen Bedeutung zentrale Elemente für die Proteinarchitektur. Um die Translation von Polyprolinsequenzen zu ermöglichen, haben Bakterien spezialisierte Strategien entwickelt, die die Translationseffizienz regulieren und die Anpassungsfähigkeit an zelluläre Anforderungen verbessern. Nicht alle dieser Mechanismen sind bisher vollständig bekannt und um die Prozesse der Translation gezielt optimieren zu können, ist ein tieferes Verständnis dieser Systeme von entscheidender Bedeutung.

Zunächst konnten mithilfe von bioinformatischen Analysen und eines Lumineszenz-Reportersystem, das Ribosomenpausen in lebenden Zellen misst, signifikante Unterschiede in der Codonnutzung bei Prolinen in *Escherichia coli* aufgedeckt werden. Dabei beeinflusst die Wahl der Prolincodons und die Verfügbarkeit spezifischer tRNAs maßgeblich die Effizienz des Einbaus von Prolinen. Die selektive Verwendung verschiedener Codons optimiert somit nicht nur die Translationsseffizienz, sondern dient auch als Mechanismus zur flexiblen Anpassung der Proteinkopienzahl an die Bedürfnisse der Zelle.

Darüber hinaus konnte EfpL als Paralog des Elongationsfaktors EF-P, der die Translation von Polyprolinsequenzen erleichtert, charakterisiert werden. Biochemische und strukturelle Analysen zeigten, dass EfpL eine Schlüsselfunktion bei der Rettung von Ribosomen hat, die an Prolinreichen Motiven feststecken. Die Koexistenz von EF-P und EfpL kann als evolutionärer Mechanismus gesehen werden, der das Wachstum beschleunigt und den Umgang mit Translationsstress optimiert. Ribosomenprofilanalysen enthüllten, dass sowohl EF-P als auch EfpL neben Polyprolinmotiven auch andere Sequenzen erkennen können, die zu einem Ribosomenarrest beitragen. Interessanterweise führte die Überexpression von *efp* und *efpL* zu Translationspausen an bestimmten Motiven, was auf eine komplexe Regulierung dieser Faktoren hindeutet. Hervorzuheben ist, dass EfpL den metabolischen Zustand der Zelle durch Lysinacylierungen erfassen kann, wodurch eine präzise Anpassung der Translation an die zellulären Bedingungen ermöglicht wird.

Diese Erkenntnisse tragen zu einem besseren Verständnis der Regulation von Ribosomenarrestsituationen durch spezifische Elongationsfaktoren bei und liefern wichtige Einblicke in die Komplexität der bakteriellen Translation und die evolutionären Mechanismen.

## Summary

Bacterial translation is an essential and highly complex process influenced by numerous factors. Proline-rich sequences pose a particular challenge, as the ring structure of proline slows its incorporation into polypeptide chains, often resulting in ribosome stalling. Despite these challenges, polyproline motifs are crucial for protein architecture due to their structural and functional significance. To enable the translation of polyproline sequences, bacteria have evolved specialized strategies that regulate translation efficiency and enhance adaptability to cellular demands. However, not all of these mechanisms are fully understood, and to optimize translation processes effectively, a deeper understanding of these systems is critical.

First, significant differences in proline codon usage in *Escherichia coli* were identified using bioinformatic analyses and a luminescence reporter system to measure ribosome pausing in living cells. The choice of proline codons and the availability of specific tRNAs significantly influence the efficiency of proline incorporation. The selective use of different codons not only optimizes translation efficiency but also functions as a mechanism to flexibly adjust protein copy numbers to the needs of the cell.

Additionally, EfpL was characterized as a paralog of the elongation factor EF-P, which facilitates the translation of polyproline sequences. Biochemical and structural analyses revealed that EfpL plays a key role in rescuing ribosomes stalled at proline-rich motifs. The coexistence of EF-P and EfpL represents an evolutionary mechanism that accelerates growth and improves the response to translational stress. Ribosome profiling analyses revealed that EF-P and EfpL can recognize other sequences in addition to polyproline motifs that induce ribosome stalling. Interestingly, overexpression of *efp* and *efpL* was found to cause translation pauses at specific motifs, indicating a complex regulation of these factors. Notably, EfpL can sense the metabolic state of the cell through lysine acylations, enabling precise adjustment of translation to cellular conditions.

These findings contribute to a better understanding of ribosome stalling and its resolution by specific factors, providing important insights into the complexity of bacterial translation and the evolutionary mechanisms that have led to the development of efficient ribosome rescue processes.

# 1 Introduction

## 1.1 Bacterial translation

The translation of genetic information into functional proteins is a critical process in all living organisms. In bacteria, the ribosome is the central machinery to decode messenger RNA (mRNA) into polypeptide chains, ensuring cellular function and survival[1]. Translation involves three main stages – initiation, elongation, and termination – all being checkpoints for protein synthesis to secure protein quality. In most bacteria, transcription and translation are coupled and occur in a coordinated manner[2] **(Fig. 1)**. The mRNA, transcribed by the RNA polymerase, is subsequently bound by the small ribosomal subunit to find the Shine Dalgarno sequence at the ribosome binding site (RBS). This ensures the correct positioning of the start codon[3]. The large ribosomal subunit then assembles with the small subunit to form a complete ribosome[4]. During translation elongation, the ribosome moves along the mRNA, decoding each codon. Therefore, the ribosome has three active sites. The aminoacyl (A)-site is the point of entry for the corresponding aminoacyl-transfer RNAs (tRNAs). Peptide bonds form between sequential amino acids, extending the polypeptide chain at the peptidyl (P)-site. At the exit (E)-site, the uncharged tRNA is released[5]. After the completed translation of the mRNA into a polypeptide chain, the ribosome encounters a stop codon, leading to the release of the newly synthesized protein and the disassembly of the translation machinery[6]. Meanwhile, new ribosomes can bind to the same mRNA strand and initiate translation even before the previous ones completed, leading to the formation of polysomes[7]. With higher growth rates, the spacing between ribosomes on the mRNA decreases from 120 to 60 nucleotides[8]. Therefore, a coordinated interplay of the ribosome with several cellular compounds is important to ensure correct and efficient protein synthesis[9].

### 1.1.1 Translational pausing and ribosome rescue

Translation is influenced by various factors such as ribosomal integrity, mRNA structure, codon usage, amino acid nature, and cellular conditions, making the translation rate not uniform[2,5,10-13]. Damage to mRNA halts translation, but certain mRNA sequences can also intrinsically cause the ribosome to pause **(Fig. 1)**. Ribosome stalling, however, is not necessarily harmful, as pausing can play a regulatory role, helping to coordinate co-translational processes and protein maturation[14,15]. Specific mRNA sequences serve as regulatory elements by forming stable secondary structures, such as hairpins, that physically obstruct ribosome progression[16]. Additionally, the presence of rare codons, where corresponding tRNAs are in low abundance, can slow down translation for co-translational folding[14,17]. Furthermore, specific nascent peptide sequences can interact with the ribosomal exit tunnel, causing the ribosome to stall. These factors form part of regulatory mechanisms that adjust protein synthesis in response to various cellular conditions[18,19].

However, ribosome stalling halts ribosomal progression, potentially leading to the accumulation of incomplete or misfolded proteins. Stalled ribosomes pose an even more challenging situation during polysome formation, as a single stalled ribosome can obstruct the progression of others in the queue[20]. These collisions not only block translation but can also recruit quality control mechanisms, which may degrade functional mRNA or prematurely terminate translation, reducing protein output[21-24]. In bacteria, ribosome stalling is common, often worsened by ribosome-targeting antibiotics or environmental factors that damage mRNA[25,26]. Often, stalling events are temporary and can be resolved by the ribosome alone or with the assistance of specific factors[27-32]. In certain cases, stalling persists, leading to prolonged pauses, ultimately requiring targeted resolution mechanisms.

One of the primary bacterial mechanisms to address ribosome stalling is the *trans*-translation system, which rescues ribosomes stalled on non-stop mRNAs[21,23]. In this process, the incomplete A-site of the ribosome is recognized by a complex of the small transfer-messenger-RNA (tmRNA), the SsrA-binding protein SmpB, the elongation factor Tu (EF-Tu) and guanosine-5'-triphosphate (GTP)[33]. The tmRNA, along with its protein partner SmpB, acts to release the ribosome from defective mRNA that lacks a stop codon, allowing the ribosome to be recycled. This system not only clears stalled ribosomes but also tags incomplete proteins for degradation, ensuring that only fully synthesized and functional proteins persist in the cell[21].

In addition to *trans*-translation, alternative ribosome rescue pathways have been identified, including the actions of proteins such as ArfA[24,34] and ArfB[35]. These factors act as backup systems to rescue stalled ribosomes when the *trans*-translation system is compromised or insufficient[25]. ArfA recognizes stalled ribosomes and recruits release factors to trigger peptide release[34], while ArfB directly promotes ribosome disassembly by cleaving the peptidyl-tRNA bond[35]. These mechanisms ensure that ribosomes are freed from stalled mRNAs and made available for productive rounds of translation, but the mRNA and the incomplete polypeptide are not targeted for degradation[34-36].

Some bacteria additionally encode for RqcH, a homolog of the eukaryotic ribosome-associated quality control factor Rqc2. The bacterial ribosome-associated quality control targets the stalled ribosome and tags aberrant peptides for degradation, recycling the large ribosomal subunit for future translation[37-39]. Moreover, peptidyl-tRNAs can drop-off without normal termination, triggered by very short nascent peptides, specific leader sequences, codon patterns, or antibiotics[40-43]. After drop-off, peptidyl-tRNA hydrolase (Pth) cleaves the peptide-tRNA bond, preventing toxic buildup and maintaining a functional tRNA pool[44-49]. The efficiency of ribosome rescue systems is vital for bacterial survival, managing translational stress efficiently, and preventing the accumulation of incomplete or faulty proteins[22].

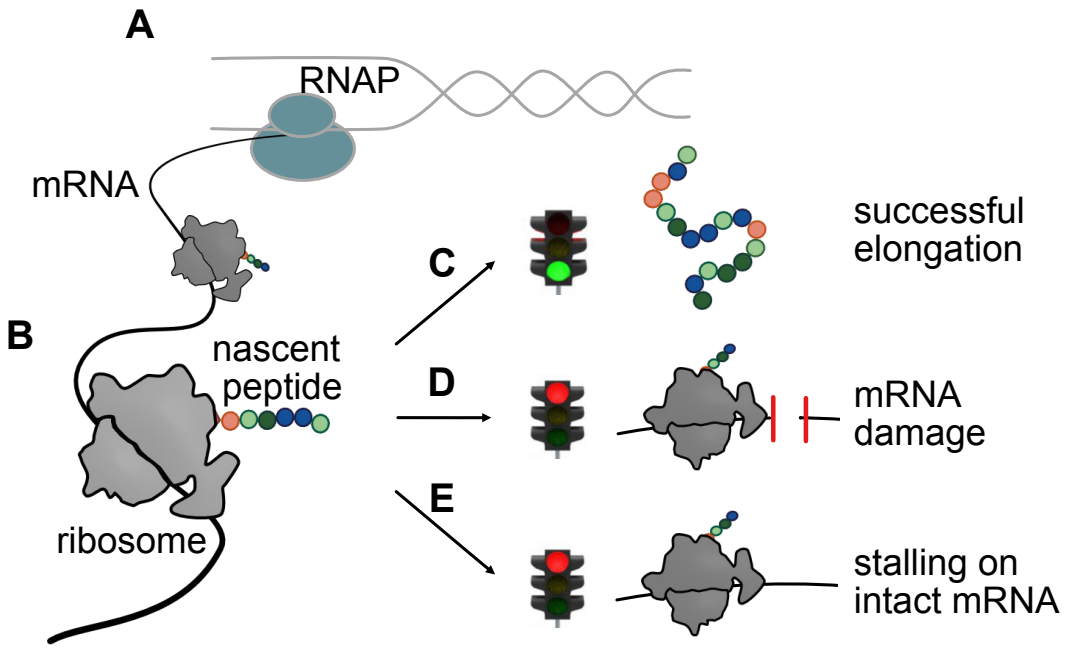



**Figure 1: Illustration of protein synthesis and ribosome stalling**

In most bacteria, transcription (**A**) and translation (**B**) are tightly coupled processes, and predominantly ribosomes successfully elongate the polypeptide chain, producing a full-length protein (**C**). Damage to the mRNA can lead to translational stops (**D**). Ribosome stalling can also occur on intact mRNA sequences (**E**). Figure is adapted from Keiler (2015)[22].

### 1.1.2 Proline-mediated ribosome stalling

One specific cause of ribosome stalling is the presence of polyproline motifs[50-52]. Proline is unique among the amino acids due to its rigid cyclic structure. In proteins, both *cis* and *trans* conformations of proline occur naturally, with nearly equivalent energetic stability[53,54]. When proline appears in consecutive sequences, they can form either a right-handed polyproline helix I (PPI, *cis*) or a left-handed polyproline helix II (PPII, *trans*). In addition to the common α-helix and β-sheet structures, the PPII helix is considered a major secondary structure element, playing an essential role in mediating protein-protein and protein-nucleic acid interactions[55,56]. Unlike other amino acids, the side chain of proline is covalently bonded to the nitrogen of its amino group, resulting in a pyrrolidine ring that restricts the flexibility of the peptide backbone **(Fig. 2)**. This rigidity hinders the ability of the ribosome to accommodate proline residues efficiently, particularly when they occur consecutively[31,50,57]. When ribosomes encounter sequences with multiple proline residues, these constraints can lead to a significant slowdown or stalling of the translation process[50-52].

Despite this, proline-mediated pauses are more than just biochemical bottlenecks — they can have various functional consequences in the cell. These pauses can directly influence the folding and processing of nascent proteins by providing additional time for proper co-translational folding, which is critical for proteins that need to acquire specific conformations

as they exit the ribosome[14,15]. Furthermore, proline-induced stalling can serve as a form of translational regulation. For instance, certain proline-rich sequences act as molecular sensors, responding to environmental conditions such as nutrient levels or cellular stress. These sequences allow cells to modulate the rate of translation in response to such signals, adapting protein synthesis rates to optimize cellular function under changing conditions[17,58]. Bioinformatic analyses in *E. coli* have revealed that proline-rich sequences are not randomly distributed within proteins but are often positioned in key regulatory regions or domains that benefit from translational pausing[17]. For example, proline-rich motifs are frequently found in proteins involved in signal transduction, transcriptional regulation, and stress responses, suggesting that proline-mediated stalling is an evolutionary feature that enables fine-tuning of protein synthesis and function in response to internal and external stimuli[17]. This underscores the importance of proline not just as an amino acid with structural peculiarities but as a critical regulatory element in translation elongation and cellular homeostasis.

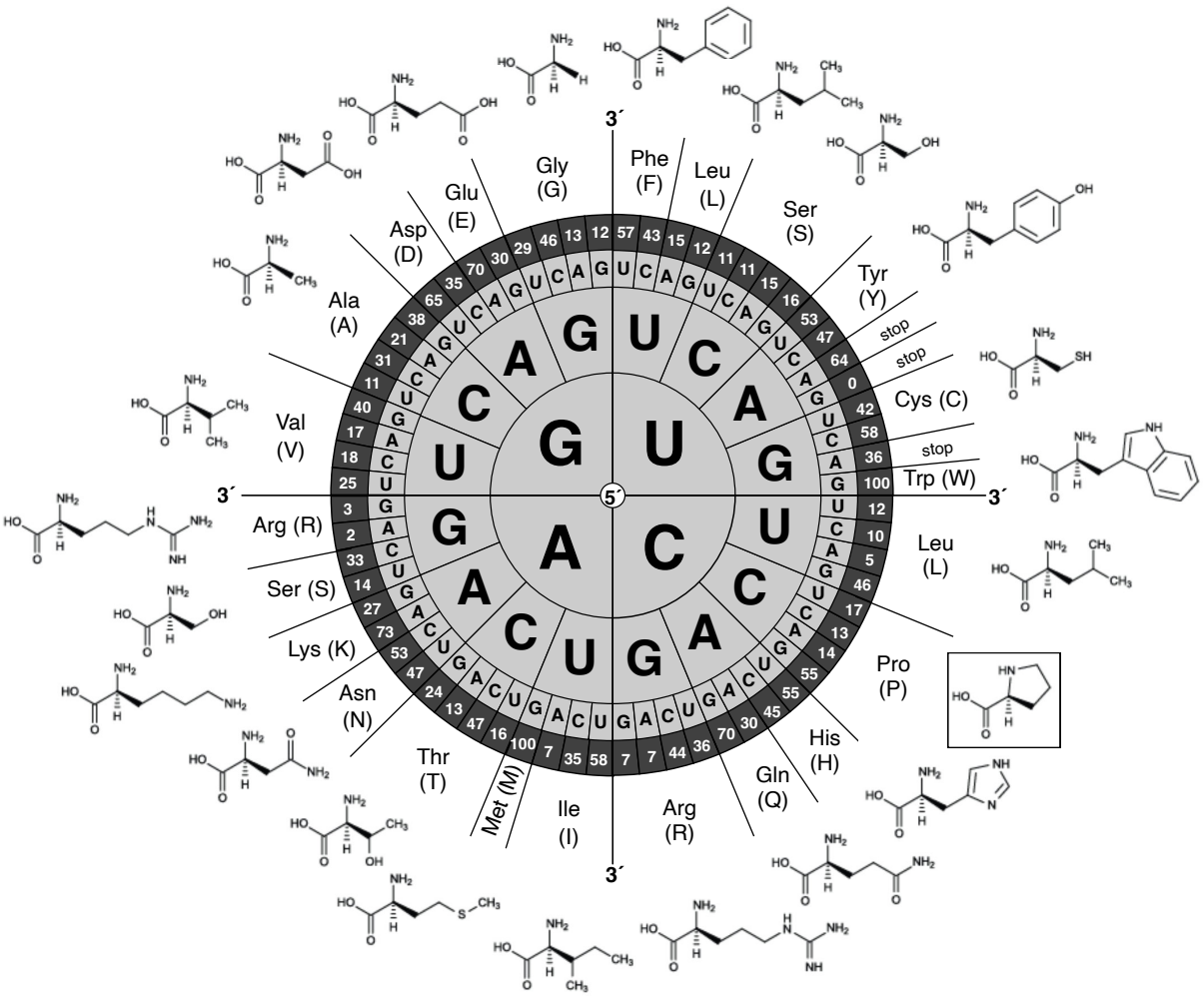



**Figure 2: Proteogenic amino acids and codon usage in *E. coli***

Circular representation of the genetic code. Codons encoding amino acids are read from the center letter toward the outer layer with their associated codon usage in percentage in *E. coli* K12 from the codon usage database[59]. Amino acid structures are represented next to their respective boxes. Proline with its pyrrolidine ring is marked.

## 1.2 The Elongation factor P

Despite their inherent complexity, polyproline sequences are not uncommon in nature and play important roles in protein structure[53,55,56,58,60]. In *E. coli*, approximately one-third of all proteins contain at least one diproline motif[17]. To facilitate the translation of these challenging sequences, almost all living organisms possess specialized factors that alleviate polyproline-induced ribosome stalling, thereby resuming protein synthesis[28,29,61]. In eukaryotes and archaea, this role is fulfilled by Initiation Factor 5A (e/aIF5A)[62-65]. The orthologous factor in bacteria is known as Elongation factor P (EF-P)[66]. The significance of IF5A and EF-P in translation efficiency is underscored by its ubiquitous presence. An EF-P encoding gene is present in nearly all bacteria, including those with highly reduced genomes, and seems to be absent only in a few rare cases of obligate symbionts[30]. In contrast, aIF5A is found in all known archaeal proteomes. Similarly, eIF5A is strictly conserved across all eukaryotes, including the single-celled green alga *Ostreococcus tauri*, one of the smallest free-living eukaryotes or the obligate fungal parasites of the genus *Microsporidia*[67,68]. Together, these factors exemplify the evolutionary pressures that have shaped the translation machinery to accommodate the requirements of proline-containing protein synthesis.

### 1.2.1 Discovery of EF-P

In 1975, Bernard Glick and Clelia Ganoza first isolated EF-P as an extraribosomal factor that stimulates the formation of N-formyl-[35S]Met-Pmn, suggesting a role in translation elongation[66]. Subsequent research demonstrated that this stimulation of peptidyl transferase activity was not universal but rather specific to certain aminoacyl-tRNAs[66]. From systematic *in vitro* analyses, Ganoza and colleagues later hypothesized an inverse correlation between the relative size of the amino acid side chain and the necessity of EF-P for translation[69]. In 2009, a co-crystal structure of the 70S ribosome from *Thermus thermophilus* bound to EF-P, shedding light on how EF-P might facilitate peptide bond formation[70]. Despite EF-P closely resembling the size and shape of a tRNA[71] **(Fig. 3)**, it does not occupy the traditional ribosomal binding sites. Instead, EF-P binds between the P-site and the E-site, where the deacylated tRNA exits the ribosome. The bioinformatic identification and experimental validation of the post-translational modification system for activating EF-P in *E. coli* and *Salmonella enterica* in 2010[72] laid the groundwork for elucidating the exact role of this elongation factor in protein biosynthesis some years later. Parallel *in vitro* translation experiments and additional *in vivo* analyses in Δ*efp* cells confirmed the initial hypothesis that EF-P is required to overcome ribosome arrest occurring at three or more consecutive prolines[28,29]. Shortly after deciphering the function of EF-P in bacteria, it was demonstrated that the eukaryotic ortholog, eIF5A, is similarly essential for the effective translation of proline sequences[65]. This finding not only confirmed the universal importance of resolving a ribosome arrest at polyproline sequences, but also indicated that EF-P and IF5A perform analogous molecular functions in elongation.

While the role of bacterial EF-P appears to be specifically linked to proline translation, the function of the eukaryotic ortholog eIF5A was further extended. Experimental evidence from studies on *Saccharomyces cerevisiae* eIF5A showed a global effect on translation elongation, not limited to polyproline sequences[65,73]. Ribosome profiling data from *eIF5A*-deficient cells reveal widespread stalling across diverse coding sequences and a role for eIF5A in translation termination[74]. Additionally, eIF5A was discovered as a crucial ribosome-associated quality control factor, enabling efficient peptidyl transfer[75].

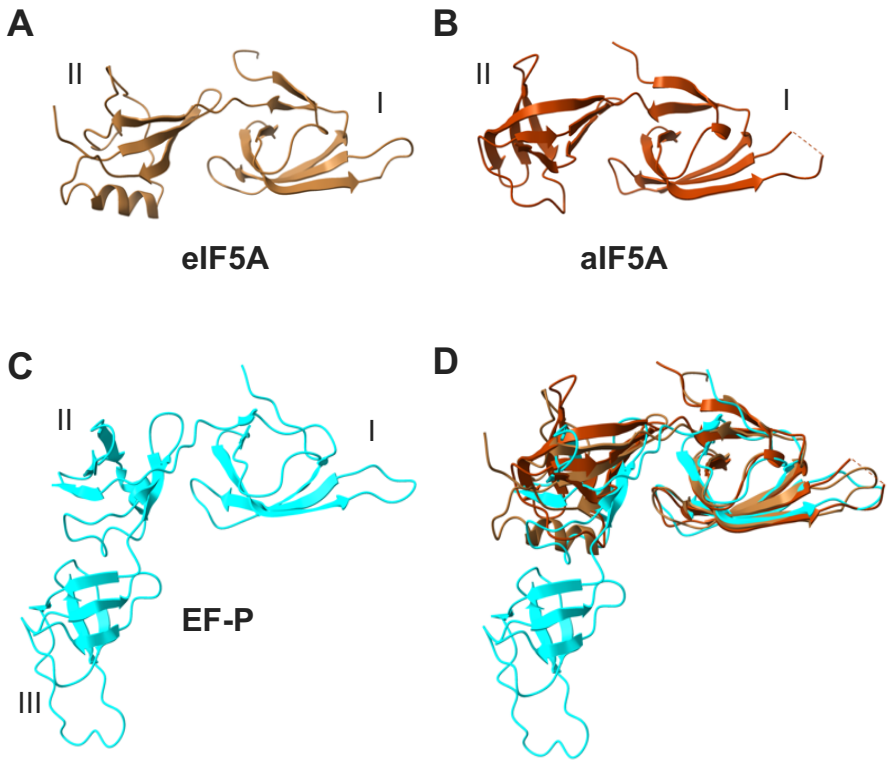



**Figure 3: Structural comparison of IF5A and EF-P**

Crystal structures of eIF5A (**A**; *S. cerevisiae*; pdb: 3ER0) and aIF5A (**B**; *Methanococcus jannaschii;* pdb: 1EIF) depict the two domain structure (I: SH3-like N-domain; II: Oligonucleotide binding C-domain)[76]. The crystal structure of EF-P (**C**; *E. coli*; pdb: 6ENU) shows an additional third domain (I: KOW-like N-domain; II and III: Oligonucleotide binding domains)[31]. Overlay with the Matchmaker tool from UCSF Chimera[77] (**D**) highlights the structural similarity.

### 1.2.2 EF-P structure and function

Structure analysis of aIF5A, eIF5A, and EF-P from various organisms revealed a significant level of overall structure similarity[30] **(Fig. 3)**. The initial published structures originated from thermophilic archaea[76,78,79]. The archaeal factor aIF5A comprises two distinct domains: the N-terminal domain (I) forms a SH3-like β-barrel fold, while the C-terminal domain (II) resembles the oligonucleotide binding (OB) fold of RNA-binding proteins **(Fig. 3A)**. Compared to its eukaryotic ortholog eIF5A, aIF5A is slightly shorter, but with a high sequence similarity, especially in the N-terminal domain **(Fig. 3B)**[30]. Unlike the two-domain structure of IF5A,

bacterial EF-P is composed of three domains, with an additional C-terminal OB domain (III) that completes the L-form of EF-P[71] **(Fig. 3C)**. The structural similarity of the two EF-P OB domains suggests that one may have originated from the duplication of the other[71]. As with IF5A, the most conserved region is found in the β3Ωβ4 loop within domain I[71]. Like a tRNA, EF-P spans both ribosomal subunits, while domain I interacts with the large subunit and domain III reaches toward the small subunit **(Fig. 4)**. Residues Y180 and R183 in domain III of EF-P interact with both the anticodon stem of the P-site tRNA and specific nucleotides of the 16S ribosomal RNA (rRNA), likely preventing the P-site tRNA from translocating to the E-site[70,71]. Domain III also contains conserved residues near the E-site codon of the mRNA[70]. At the other end of EF-P, multiple positively charged residues interact with the P-site tRNA and the 23S rRNA near the peptidyl transferase center[70]. A key feature of EF-P and IF5A is the post-translational modification at the tip of the highly conserved β3Ωβ4 loop within domain I. This modification enhances interactions with the CCA end of the P-site tRNA, allowing EF-P and IF5A to stabilize and correctly position the tRNA. Through these interactions, EF-P indirectly stimulates peptide bond formation, ensuring efficient protein biosynthesis[28,61,80] **(Fig. 4)**.

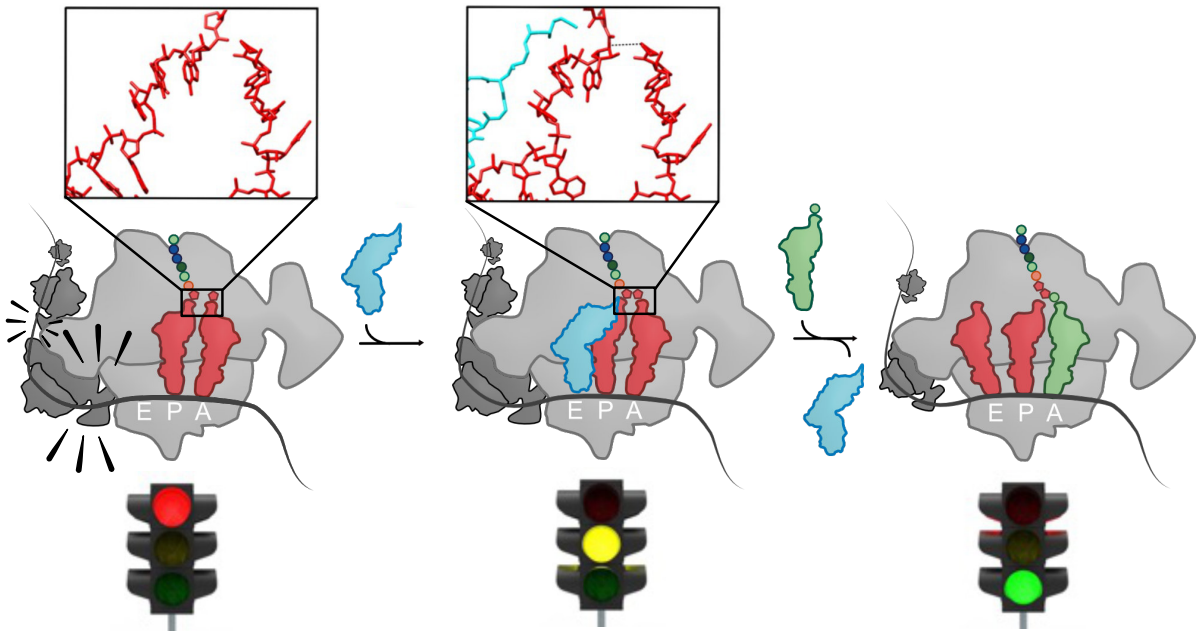



**Figure 4: Molecular function of EF-P**

(Left) Translation of consecutive prolines leads to ribosomal stalling. (Middle) EF-P is recruited to the stalled ribosome and binds close to the site of tRNA-exiting (E-site). By interacting with the P-site prolyl-tRNA, EF-P stimulates proline-proline peptide bond formation. (Right) Translation proceeds upon detachment of EF-P and with the binding of a new aminoacyl-tRNA at the ribosomal A-site. Depicted are the structural images of P- and A-site tRNAs (red) in a polyproline stalled ribosome. The distance between the ester carbonyl carbon of the peptidyl-tRNA and the α-amino group of the aminoacyl-tRNA with (pdb: 6ENJ) and without (pdb: 6ENF) modified EF-P (turquoise) is shown[31]. Figure is adapted from Lassak *et al.* (2021)[81].

### 1.2.3 Post-translational modifications of EF-P

The activity of both IF5A and EF-P relies on the post-translational extension of a positively charged residue[80,82,83]. While the modification systems for IF5A are highly conserved in eukaryotes and archaea, bacteria employ various analogous strategies. The (deoxy-)hypusination of IF5A involves the unusual amino acid hypusine (N$\varepsilon$-(4-amino-2-hydroxybutyl)lysine). In the cell, hypusine is rapidly and likely irreversibly formed following eIF5A synthesis[65,84-87]. Two enzymatic steps are required for complete modification: the first is catalyzed by deoxyhypusine synthase (DHS), which transfers the 4-aminobutyl moiety from spermidine to a conserved lysine in eIF5A[88,89]. DHS is nicotinamide adenine dinucleotide ($NAD^+$)-dependent and forms the intermediate deoxyhypusine in a four-step reaction[89-91]. The second modification enzyme is the eIF5A-specific deoxyhypusine hydroxylase (DOHH)[92]. While DHS is essential in archaeal and eukaryotic domains of life, DOHH seems to be essential only in higher eukaryotes[30].

The first modification type of EF-P was discovered in 2010 for *E. coli* and other γ-proteobacteria[72,93-95]. This modification is characterized by a two-step β-lysine addition process that involves the catalytic action of specialized enzymes. The first enzyme, 2,3-aminomutase EpmB, isomerizes (*S*)-α-lysine to (*R*)-β-lysine, which serves as a donor substrate for the second enzyme. The β-lysine is ligated to the *ε*-amino group of the lysine residue K34 of EF-P by the EF-P-specific ligase EpmA[96]. Notably, EpmA, shares homology with a type II aminoacyl-tRNA synthetase, but evolved substrate specificity for EF-P instead of tRNA[72,94]. Additionally, some bacteria encode for the EF-P hydroxylase EpmC[97], but hydroxylation of EF-P is not essential for the function[28,29,98]. Comparative proteomic analyses consistently indicate that EpmA and EpmB are present in a quarter of all bacterial proteomes sequenced to date[72,80] **(Fig. 5)**.

In contrast, approximately 70% of bacteria encode EF-P, with a lysine positioned at the equivalent site to *E. coli* K34[30]. In *Bacillus subtilis*, EF-P undergoes a distinct type of modification at this lysine side chain, involving the β-lysine structural analog 5-aminopentanol[99]. YmfI was identified as the enzyme responsible for reducing 5-aminopentanone to 5-aminopentanol, which constitutes the final step in the modification pathway[100]. Additional modification states were identified through tandem mass spectrometry analysis, leading to the proposal of a putative pathway. According to this data, hydroxypentanone is the first intermediate, followed by pentenone, and subsequently 5-aminopentanone[101]. Orthologs of YfmI have been identified in bacteria of the Firmicutes order, and it appears that EF-P 5-aminopentanolylation extends to genera such as *Listeria* and *Staphylococcus*[102] **(Fig. 5)**.

The discovery of 5-aminopentanolylation in *B. subtilis* expands the spectrum of EF-P lysine modifications. However, approximately 30% of EF-P homologs have a different amino acid at the corresponding site, with arginine being the predominant alternative. β-proteobacteria and some γ-proteobacteria, including *Pseudomonas* species, mono-rhamnosylate EF-P at the conserved arginine residue R32[80] **(Fig. 5)**. This modification is catalyzed by the glycosyltransferase EarP, which utilizes the nucleotide sugar donor dTDP-β-L-rhamnose. dTDP-β-L-rhamnose itself is synthesized through the RmlABCD pathway[103-106]. The addition of this sugar moiety is an unexpected and unique form of EF-P modification, as it stands in stark contrast to all previously characterized lysine-based modifications of EF-P and its eukaryotic analog, eIF5A. Nevertheless, despite the structural differences, all EF-P modifications alleviate polyproline-induced ribosome stalling[80,97,101,103].

Enzymes involved in β-lysinylation, rhamnosylation, or 5-aminopentanolylation are found in approximately 40 % of sequenced genomes **(Fig. 5)**. Thus, it remains unclear whether additional, yet undiscovered modification strategies exist. Instead of post-translational modification, EF-Ps of Actinobacteria stabilizes the loop due to the rigidity of a palindromic consensus sequence PGKGP[102]. This exemplifies a modification-independent strategy for EF-P functionality. The remarkable diversity in EF-P modification strategies underscores the evolutionary adaptability of bacterial systems in addressing translational challenges.

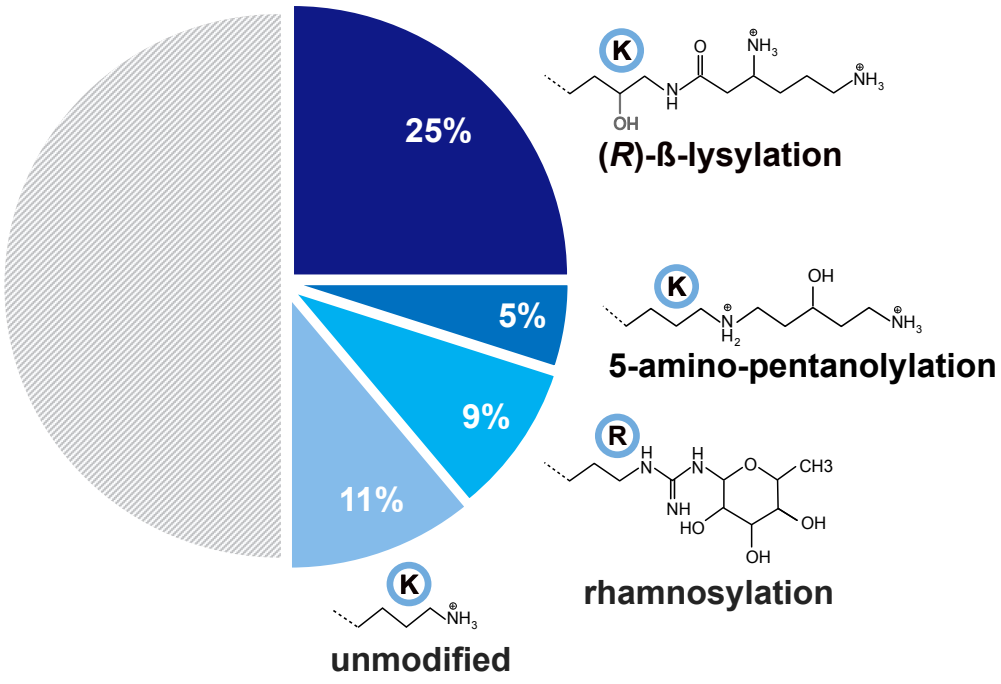



**Figure 5: Diversity of EF-P modifications**

The distribution of EF-P modifications in bacteria of yet analysed EF-Ps is shown. Approximately 25% of bacteria modify lysine (K) with (*R*)-β-lysine, while 5% modify lysine with 5-amino-pentanol. Another 9% of bacteria rhamnosylate an arginine (R). Up to 11% use an unmodified EF-P with a lysine. Around 50% have unknown and uncharacterized EF-Ps.

## 1.3 Alternative ways to alleviate stalling at proline-rich sequences

While a/eIF5A is essential in archaea and eukaryotes[65,107], EF-P can be deleted in many bacteria[80,95,108-113]. Additional to EF-P, alternative enzymes to alleviate ribosome stalling at proline-containing sequences in bacteria were described[114-118]. Unlike the translational stress response pathways of *trans*-translation or ArfA/ArfB, these enzymes most likely enter the ribosome from the E-site and enable translation to continue. They play a crucial role in maintaining efficient translation, especially when EF-P levels are insufficient or absent, ensuring the cellular machinery can cope with translational challenges effectively. Their existence underscores the adaptability of bacterial systems to diverse translational challenges, enabling bacteria to preserve growth and viability across various environmental conditions.

### 1.3.1 ABCF ATPases

ABCF ATPases are an important family of ATP-binding cassette (ABC) proteins of both antibiotic resistance factors and essential housekeeping proteins involved in assisting ribosome assembly and protein synthesis, including alleviating ribosome stalling[116,117,119-121]. Unlike other members of the ABC family, typically associated with transport across membranes, ABCF ATPases lack membrane-spanning domains[122]. ABCF ATPases interact directly with ribosomes and affect translation by binding to the ribosomal subunits with their P-site tRNA interaction motif (PtIM) domain[119,120,123]. ABCFs are widely distributed across bacteria and eukaryotes, with an average of four ABCF proteins encoded per bacterial genome[116]. *E. coli* has four ABCF ATPases: EttA, YheS, YbiT, and Uup[119]. *E. coli* EttA seems to regulate the early stages of translation elongation[114,119,120]. YheS responds to translational arrest caused by SecM and YbiT manages stalling caused by poly-basic sequences as well as poly-acidic sequence-induced intrinsic ribosome destabilization[114]. Uup, which was initially suggested to play a role in replication fork progression[124], was found to help in the translation of polyproline motifs[114]. *B. subtilis* YfmR, a member of the Uup subfamily[116], cooperates with EF-P to alleviate ribosome stalling at aspartate-proline[117] and polyproline motifs[125]. In general, ABCF ATPases can help to manage hard-to-translate sequences.

### 1.3.2 RNA binding protein YebC

Another potential player in the bacterial ribosome rescue system is YebC. YebC is a conserved bacterial protein belonging to the YebC/PmpR family of transcriptional regulators, but its role extends beyond transcriptional regulation[118,125]. A phylogenetic analysis revealed that the YebC-family proteins have evolved distinct functions in transcription or translation separately[125]. Recent studies showed function in rescuing stalled ribosomes by directly interacting with the 23S rRNA near the peptidyl-transferase center. Interestingly, many bacterial species encode two YebC paralogs, YebC and YebC2[126]. YebC in *Streptococcus pyogenes*[118] and YebC2 in *B. subtilis*[115] were shown to enhance the translation of polyproline stretches in the absence of EF-P and the corresponding ABCF ATPase YfmR.

This indicates that YebC and/or YebC2 can serve as an alternative ribosome rescue mechanism, compensating for the absence of other factors. Notably, YebC is also homologous to the mitochondrial TACO1, which alleviates polyproline-mediated stalling of mitoribosomes, while a mitochondrial counterpart of EF-P or eIF5A remains unidentified[127]. In bacterial systems, YebC/YebC2, EF-P, and YfmR/Uup act independently, forming distinct mechanisms to ensure efficient translation[115].

### 1.3.3 Elongation factor P-like proteins

Numerous bacteria additionally possess an EF-P paralog, called EF-P like (EfpL; also known as YeiP) of unknown function. Bioinformatic analyses using structural predictions from AlphaFold suggest that EfpL proteins share a three-domain structure similar to EF-P. The C-terminal OB-domain III shows the greatest similarity, primarily functioning in interactions with the small ribosomal subunit and the anticodon stem loop of the P-site tRNA. Notably, conserved residues in both EF-P and EfpL are positioned to form hydrogen bonds with the P-site tRNA and helix h29 of the 16S rRNA[31]. However, key residues within the EF-P KOW domain, critical for recognizing prolyl-tRNA in stalled ribosomes, are less conserved in EfpL. This divergence suggests that EfpL might operate through a mechanism distinct from the canonical EF-P, possibly targeting different translational challenges or stalling motifs.

The translation of proline-rich sequences represents a complex system involving numerous different key factors that either support the ribosome or further complicate the process. These factors include specialized elongation proteins, post-translational modification, and specific mRNA codon contexts that influence ribosomal dynamics. The interplay between these components adds multiple layers of complexity to the regulation of ribosome stalling and ensures the accurate synthesis of proline-rich proteins. However, the elements that assist the ribosome in navigating these stalls can also amplify translational challenges under certain conditions. Despite recent advancements in understanding translation regulation, many questions remain unanswered regarding ribosome stalling and the specific roles of translation factors. To address these gaps, this thesis aims to investigate the molecular mechanisms underlying stalling motifs and to conduct a comprehensive biochemical characterization of EfpL. Understanding how these systems function and how they are regulated is essential for deciphering the mechanisms cells use to maintain translation efficiency and adapt to stress.

## 2 Proline codon pair selection determines ribosome pausing strength and translation efficiency in bacteria

Krafczyk R*, Qi F*, Sieber A, Mehler J, Jung K, Frishman D & Lassak J. 2021. Proline codon pair selection determines ribosome pausing strength and translation efficiency in bacteria. *Commun. Biol.* **4**, 589

*Authors contributed equally

communications biology

ARTICLE

https://doi.org/10.1038/s42003-021-02115-z OPEN

# Proline codon pair selection determines ribosome pausing strength and translation efficiency in bacteria

Ralph Krafczyk [1,4], Fei Qi [2,3,4], Alina Sieber [1], Judith Mehler [1], Kirsten Jung[1], Dmitrij Frishman [3✉] & Jürgen Lassak [1✉]

The speed of mRNA translation depends in part on the amino acid to be incorporated into the nascent chain. Peptide bond formation is especially slow with proline and two adjacent prolines can even cause ribosome stalling. While previous studies focused on how the amino acid context of a Pro-Pro motif determines the stalling strength, we extend this question to the mRNA level. Bioinformatics analysis of the *Escherichia coli* genome revealed significantly differing codon usage between single and consecutive prolines. We therefore developed a luminescence reporter to detect ribosome pausing in living cells, enabling us to dissect the roles of codon choice and tRNA selection as well as to explain the genome scale observations. Specifically, we found a strong selective pressure against CCC/U-C, a sequon causing ribosomal frameshifting even under wild-type conditions. On the other hand, translation efficiency as positive evolutionary driving force led to an overrepresentation of CCG. This codon is not only translated the fastest, but the corresponding prolyl-tRNA reaches almost saturating levels. By contrast, CCA, for which the cognate prolyl-tRNA amounts are limiting, is used to regulate pausing strength. Thus, codon selection both in discrete positions but especially in proline codon pairs can tune protein copy numbers.

[1] Department of Biology I, Microbiology, Ludwig-Maximilians-Universität München, München, Germany. [2] Institute of Genomics, School of Biomedical Sciences, Huaqiao University, Xiamen, China. [3] Department of Bioinformatics, Wissenschaftzentrum Weihenstephan, Technische Universität München, Freising, Germany. [4] These authors contributed equally: Ralph Krafczyk, Fei Qi. ✉email: d.frishman@wzw.tum.de; juergen.lassak@lmu.de

Proline has a set of characteristics that is not found in other proteinogenic amino acids. It is the only n-alkyl amino acid and thus has unique chemical properties. Its pyrrolidine ring makes proline conformationally rigid and thus it can shape protein structure: depending on its configuration—*cis* or *trans*—the binding axis rotation of amide bonds changes with major consequences for folding[1]. Peptide stretches enriched in prolines can even form a distinct type of secondary structure, the so called polyproline helix[2]. However, all these unique features come at a price. Not only is peptide bond formation with proline the slowest compared to all other proteinogenic amino acid[3–5], but ribosomes can even be arrested when translating stretches of proline residues[6–8]. However, consecutive prolines occur frequently in eukaryotic and prokaryotic proteomes[9,10]. For example, in *Escherichia coli* every third protein contains at least one polyproline motif (PP-motif, at least diproline)[11] and in *Streptomyces* species there is more than one PP-motif per protein on average[12]. The explanation for this apparent oddity is the existence of a ubiquitous elongation factor (termed EF-P in bacteria and a/eIF5A in archaea/eukaryotes) that alleviates ribosome stalling[13–16]. Nevertheless, EF-P cannot fully compensate for the translational burden caused by PP-motifs[11]. Intriguingly, bacteria can even benefit from ribosomal pausing by using it to regulate translation rates[14]. PP-motifs are enriched in inter-domain linker regions, which might promote correct folding, upstream of transmembrane regions, where they could facilitate correct insertion, and close to the protein N-terminus[11]. Here, similar to rare codons[17], PP-motifs might be instrumental in generating a translational ramp and helping to avoid ribosome collisions[18].

It is well accepted that the amino acids bracketing PP-motifs influence the pausing strength[19–22], thus representing a specific regulatory mechanism of translation. The role of proline codon choice, however, has not yet been investigated, although the incorporation speed of proline into the nascent chain differs significantly depending on which of the four codons (CCA/C/G/U) (Fig. 1a) and three tRNAs (ProK/ProL/ProM) are used (Fig. 1b)[5]. Here, we have comprehensively investigated how the interplay of codon choice and tRNA abundance affect the translation of PP-motifs.

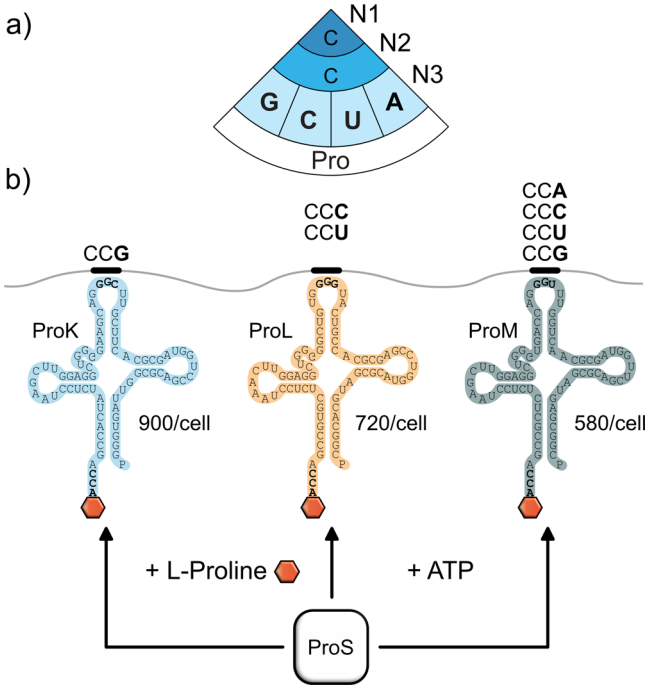



**Fig. 1 Diversity of proline codons and their corresponding tRNAs. a** The genetic code contains four codons for proline: CCG, CCC, CCU, and CCA. **b** The three tRNAs ProK, ProL, and ProM recognize distinct sets of proline codons and exhibit different levels of abundance within the cell[34]. All three prolyl-tRNAs are charged by the prolyl-tRNA synthetase ProS.

## Results

**Distribution of proline codon pairs suggests their regulatory role in translation**. Our study started with a bioinformatics analysis, in which we investigated whether codon usage differs between single prolines and proline pairs in the proteome of *E. coli* MG1655 (Figs. 2 and 3). We observed a depletion of CCC (8.1 vs. 11.6%) and CCU (12.3 vs. 15.3%) in codon pairs as compared to single prolines (Fig. 2a). Both of these codons delay diproline synthesis more ($t_{dip}$[CCC]) = ~116.3 ms; $t_{dip}$[CCU]) = ~71.4 ms) than the other two codons ($t_{dip}$[CCA]) = ~66.7 ms; $t_{dip}$[CCG]) = ~62.5 ms)[5]. Selection against slowly translating proline codon pairs is not restricted to *E. coli*: Out of 15 bacterial genomes with a broad range of GC-content values CCC and CCU are disfavored in 13 and 11 genomes, respectively (Fig. S1 and Supplementary data file S1). We next asked whether this bias might be related to codon order. Reportedly, an mRNA sequence of CC**C**/**U**-**C**/**U**CN promotes +1 ribosomal frameshifting, which is in principle counteracted by methylation of the corresponding isoacceptor tRNAs ProL and ProM at the 3′ side of the anticodon ($m^1$G37)[23,24]. However, this modification cannot fully prevent ribosome slipping, as we could demonstrate with a bioreporter in vivo (Fig. S2). Accordingly, it would be plausible that the selective pressure on proline codon pairs is most pronounced for the first codon. Indeed, our analysis unveiled strong avoidance of both CCC and CCU at the first positions, while their occurrence at the second position matches their genome-wide usage (Fig. 2b). Further, the observed bias is not restricted to proline codon pairs but also to single prolines, as long as the downstream codon starts either with "C" or "U" (Fig. 2c).

Interestingly, the negative selection of CCC/U in proline codon pairs is not compensated by overrepresentation of CCG, being the most optimal codon in terms of diproline synthesis rates[5]. Instead, an enrichment of CCA (18.2 vs. 25%) was found. Ranking proteins with proline pairs according to their translation efficiency (Fig. 2d) revealed a preference for CCG in the top 20%. These findings imply a potential regulatory role of the relative CCA accumulation in PP-motifs, e.g., to slow down translation for proper membrane insertion or at the protein start to generate a translational ramp as a late stage of translation initiation thereby reducing ribosomal traffic jams[17]. In fact, non-CCG proline codons are enriched in these regions, further supporting the idea (Fig. 2e, f).

PP-motifs can be classified into "weak", "intermediate", and "strong" pausing motifs according to their interference with translation[11,20,22]. These differences result from the preceding amino acid. We were therefore interested whether specific proline codon biases exist within these subgroups of PP-motifs. Thus, we dissected PP-motifs accordingly (Fig. 3a–d and Supplementary data file S2). The most pronounced difference to single prolines was again the CCA usage (Fig. 3d). This codon represents 23.2% of all proline codons associated with weak pausing compared to 27.2% and 27.1% for intermediate and strong pausing, respectively. This difference is significant according to a two-sided $Z$ test ($p$ value = 7.0e−3). Thus, the differences between CCA and CCG in terms of pausing strength might be an additional mechanism to tune the translation efficiency.

**An in vivo reporter system to quantify translational pausing**. In order to measure codon effects on translational efficiency, we

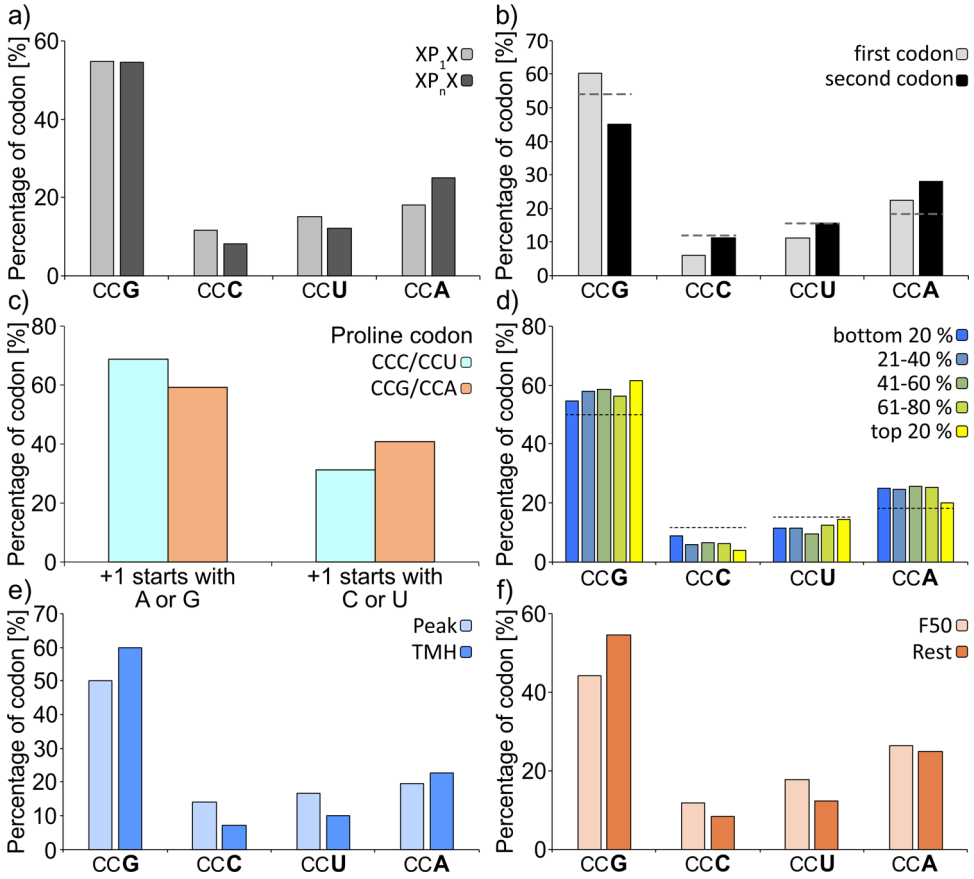



**Fig. 2 Bioinformatic analysis of proline codon bias in *E. coli*.** **a** Codon usage of either single ($XP_1X$) or consecutive ($XP_nX$) prolines (with X being any amino acid except proline and $n > 1$). $p$ value = 1.7e−30, chi-squared test. **b** Codon usage of the first and second proline in PP-motifs. Only PP-motifs with two consecutive proline residues were included in this analysis. The dashed lines indicate the codon usage for single prolines. $p$ value < 2.2e−16, chi-squared test. **c** Codon usage for amino acids in the +1-position downstream CCC/CCU (cyan) or CCG/CCA (orange) encoded single prolines. $p$ value < 2.2e−16, chi-squared test. **d** Correlation between proline codon usage in PP-motifs and translation efficiency from least efficiently translated proteins (dark blue) to most efficiently translated proteins (yellow). The dashed lines indicate the codon usage for single prolines. **e** Difference between proline codon usage of PP-motifs in the peak region (light blue, amino acids 49–59 from the TMH start where PP-motifs are enriched to facilitate the efficient insertion of TMH into the membrane) and TMHs (blue; transmembrane helices in which PP-motifs are depleted for proper folding of transmembrane segments[11]. $p$ value = 0.13, chi-squared test. **f** Proline codon usage in PP-motifs in the first 50 codons (light orange) compared with the rest of proteins (orange). $p$ value = 2.14e−7, chi-squared test.

established a reporter system that is capable of determining translational pausing strength within living cells. The system hijacks the attenuation mechanism of the histidine biosynthesis operon *hisGDCBHAF* (Fig. 4a)[25]. Here, translational speed of the preceding His-leader peptide (HisL) controls expression of the downstream structural genes[26]. Naturally this peptide contains seven consecutive histidines. When charged histidyl-tRNA is present in excess, ribosomes translate HisL non-stop, which in turn results in the formation of an mRNA attenuator stem loop that prevents transcription of *hisGDCBHAF*. When histidine concentrations are limiting, HisL translation is decelerated due to a lack of charged histidyl-tRNAs and an alternative mRNA stem loop is formed, which in turn permits transcription of the histidine biosynthesis genes. We fused the 5′ untranslated region (5′ UTR) of *hisGDCBHAF* as well as the preceding *hisL* with the *luxCDABE* operon of *Photorhabdus luminescens*[27] and integrated the resulting construct via single homologous recombination into the *E. coli* chromosome (Fig. 4b)[28,29]. Monitoring of light emission over 16 h of growth showed a maximal output of only around 500 RLU, demonstrating that almost no pausing takes place under standard growth conditions in complex medium (LB). This was expected as LB contains about 1 mM of histidine[30], which means an excess of about 100-fold[31].

To assess the potential of our reporter to measure ribosome pausing we generated HisL variants encompassing PP-motifs of varying strength (Fig. 4c). Specifically, we substituted His1 through His4 by TPPP, FPPP, or RPPP being representatives of weak, intermediate, and strong pausers, respectively[22]. As a positive control, we placed a stop codon in the position corresponding to His4. As a negative control, we chose RPAP, which does not reduce translational speed[14]. As codon for alanine, we selected GCG being highly similar to the proline codon CCG. This choice was made to minimize putative effects of mRNA structural alterations.

To delineate codon effects from those caused by the peptide sequence all prolines were encoded only by CCG. The maximal light output of the corresponding *E. coli* strains $^{RPAP}$CC**G**, $^{TPPP}$CC**G**, $^{FPPP}$CC**G**, and $^{RPPP}$CC**G** increased from 390 RLU to 44,000 RLU to 106,000 and 530,000 RLU, respectively (Fig. 4c). The results obtained here perfectly match published datasets based on completely different experimental principles[14,19,20,22,32]. Accordingly, the outcome of our assay is a result of ribosome pausing that is determined by sequence identity but not mRNA structure. Notably, the positive control HHH* reached a maximum RLU of 336,000, which was in the same range as the reporter activity of the RPPP construct. We can therefore

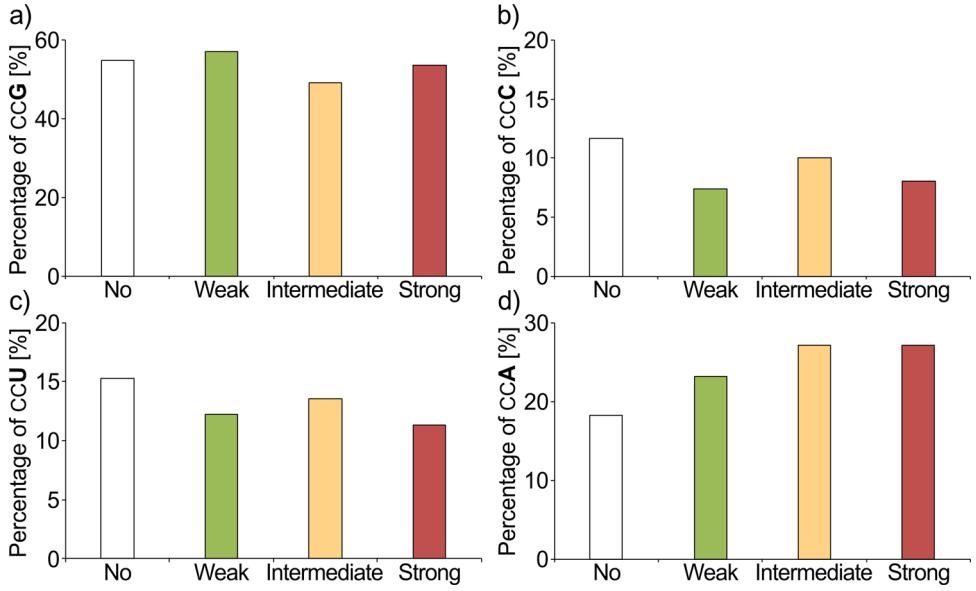



**Fig. 3 Codon usage in PP-motifs of different pausing strength.** Pausing strength of PP-motifs depends on the upstream amino acid context[11,20,22] resulting in weak, intermediate, and strong pausers. The pausing strength resulting from amino acid context is indicated by colored bars (no pausing—white; weak pausing—green; intermediate pausing—yellow; strong pausing—red). Codon usage in differently strong pausing motifs is shown for CCG (**a**), CCC (**b**), CCU (**c**), and CCA (**d**) codons. The difference is significant according to chi-squared test, $p$ value = 4.2e−3.

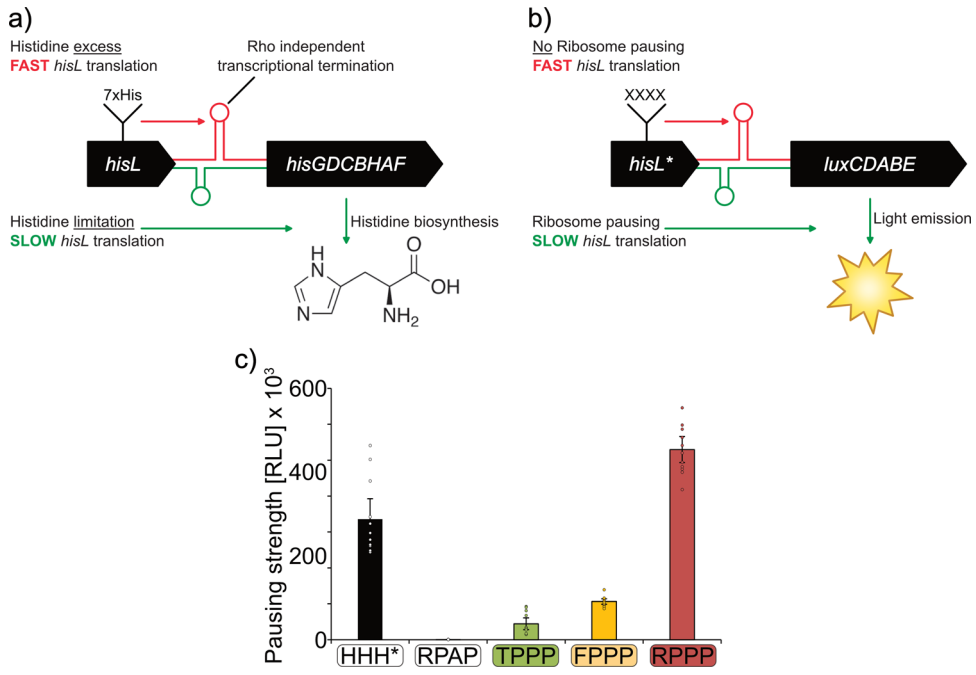



**Fig. 4 The His-pausing system for in vivo measurement of pausing strength. a** Architecture of the histidine biosynthesis operon in *E. coli*. In its native state, the histidine biosynthesis gene cluster (*hisGDCBHAF*) is regulated by the His-leader peptide (*hisL*). This peptide contains seven consecutive histidines. At high histidine/histidyl-tRNA levels, translation efficiently proceeds through the His-leader peptide, resulting in the formation of an attenuator stem loop (red) that prevents transcription of the downstream genes. At low histidine and histidyl-tRNA levels translation is slowed down allowing for transcription and translation of the structural genes and synthesis of histidine (green). **b** Architecture of the His-pausing operon. An engineered His-leader peptide (*hisL**) precedes the structural genes of the lux operon (*luxCDABE*). Here, His1 through His4 are exchanged by artificial sequence motifs (XXXX). In case of non-consecutive proline motifs (e.g., RPAP) there is no pausing, resulting in the formation of an attenuator stem loop (red) that prevents transcription of the downstream genes and low light emission. In the presence of motifs that contain consecutive prolines (e.g., RPPP) translation is slowed down allowing for transcription and translation of the structural genes and thus increased light emission (green). **c** Maximal luminescence emission at PP-motifs with increasing pausing strength. HisL*_Lux operons carrying a stop codon at the position corresponding to His4 (HHH*), non-consecutive (RPAP) or consecutive prolines of varying known pausing strength at the *hisL** position (Weak: TPPP; green. Intermediate: FPPP; yellow. Strong: RPPP; red) were chromosomally integrated in *E. coli* BW25113 and tested for maximal luminescence emission. Threonine, phenylalanine, and arginine were encoded by ACC, TTT, and CGC, respectively. CCG was used as proline codon in all constructs. $n = 12$, Error bars indicate 95% confidence intervals.

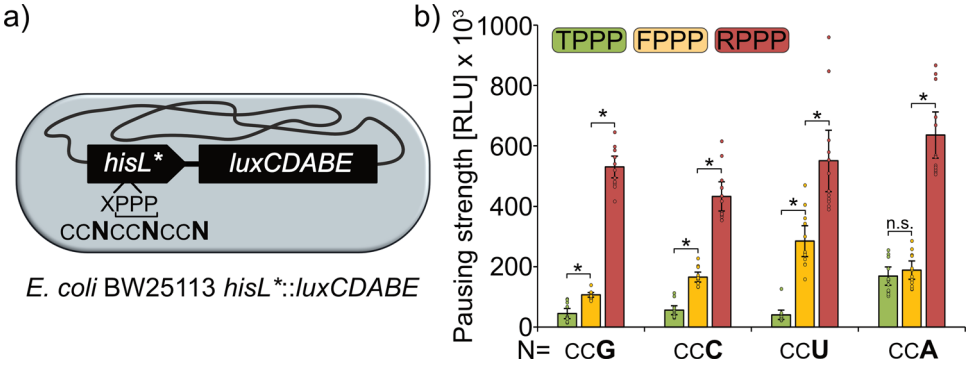



**Fig. 5 Codon-dependent pausing strength at weak, intermediate, and strong PP-motifs. a** Genomic organization of the HisL*_Lux reporter. Synthetic His-Leader peptides (HisL*) preceding the *lux* genes (*luxCDABE*) were genomically integrated at the *his*-locus. In *hisL**, His1 one was replaced by a variable amino acid (X) to modulate pausing strength[16]. His2 through His4 were replaced by proline. In this regard several reporter strains (Supplementary data file S3) were generated with *hisL** varying in the proline codon usage and are denoted as $^{\underline{X}PPP}$CC**N** where the underlined X designates the preceding amino acid and the bold **N** designates the wobble base used for encoding the proline residues. **b** HisL*_Lux carrying PP-motifs of varying pausing strength (weak—TPPP: green; intermediate—FPPP: yellow; strong—RPPP: red) with different proline codon usage were chromosomally integrated in *E. coli* BW25113 and tested for maximal luminescence emission. $n = 12$, Error bars indicate 95% confidence intervals. Data for CCG codons are duplicated from Fig. 5 for better overview. Statistically significant differences according to unpaired two-sided *t*-tests (*p* value < 0.05) are indicated by asterisks.

conclude that ribosome pausing induced by strong stallers is comparable to a stop caused by a termination signal.

Of particular interest is, that the measurements were conducted in an *E. coli* wild-type strain where stalling at consecutive prolines is alleviated by EF-P[14]. Thus, we have a tool in hand to determine pausing strength in vivo. Using the system, we unambiguously demonstrate that the burden associated with PP-motifs is an inherent translational feature and explains the strong selective pressure causing the proteome shaping[11].

**Codon choice modulates pausing strength at consecutive proline motifs**. To investigate whether the statistical tendencies of codon usage in PP-motifs can be attributed to physiological differences we conducted a systematic in vivo analysis. To this end we constructed a series of 4 × 4 HisL*_Lux reporter strains (Fig. 5a and Supplementary data file S3). Utilizing the strong pauser RPPP, CCG, CCA, and CCU were indistinguishable from each other, each producing a maximal light output of over 525,000 RLUs (Fig. 5). Only with CCC codons we found around 1.2-fold reduced maximal light emission. When testing a motif with intermediate strength (FPPP) a different pattern was obtained. In this case, CCU stretches produced significantly more light than the other codons, whereas emission using CCG was significantly decreased. CCC and CCA ranged in the middle of both. Interestingly, the most pronounced effect of codon choice on pausing strength occurred with the weak pauser TPPP. The luminescence with $^{TPPP}$CC**A** was significantly elevated by at least threefold compared to the strains encoding $^{TPPP}$CC**C**, $^{TPPP}$CC**G**, or $^{TPPP}$CC**U**. Notably, such an increase is equivalent to a step in the pausing strength from weak to intermediate pausing (Figs. 5b and S3A). This result is also in perfect agreement with our genome scale analysis (Fig. 3a–d) and explains the strong selection against CCA in weak pausers. On the other hand, the bias in favor of CCA in intermediate and strong pausers might be attributed to a regulatory role that requires a further slowdown of translation. To exclude that the observed effects derive from mRNA structure alterations, we conducted another analysis utilizing a second reporter series XPPP with X being N (AAC) for weak, L (CTG) for intermediate and W (TGG) for strong[22] (Fig. S3B, C). Expectedly, the activities are congruent with the T/F/RPPP derived data including the CCA effect in the weak pausing context. Taken together, these results demonstrate that codon choice in PP-motifs is capable of influencing ribosome pausing.

**tRNA abundance influences pausing strength at all proline codons**. The variations in proline codon bias of PP-motifs of varying strength, particularly the one of CCA, raised the question whether tRNA abundance might contribute to pausing strength. In *E. coli*, three tRNAs—ProK, ProL, and ProM—are responsible for decoding of proline codons (Fig. 1). ProM represents a general tRNA that is capable of recognizing them all[33], while ProL and ProK are more specialized and decode CCC/U and CCG, respectively[5]. These differences have a quantitative effect on the reading probabilities of the individual codons. Taking the copy numbers of ProK (900/cell), ProL (720/cell), and ProM (580/cell) into account, CCG has the highest number of the corresponding tRNAs (900 + 580 = 1480/cell)[34] and thus matches very well to the general codon usage in the *E. coli* genome, where more than 50% of all prolines are encoded by CCG (Fig. 2). CCA is the other extreme, being recognized solely by ProM and accordingly only 580 tRNA copies per cell are available for translation.

To assess an effect of prolyl-tRNA copy numbers on pausing strength, we unbalanced the native ratios in favor of either ProK, ProL, and ProM (ProX$^{++}$) by ectopically expressing them from $P_{proL}$. Beforehand, the 5′ upstream sequences of *proK* ($5'_{proK}$), *proL* ($5'_{proL}$), and *proM* ($5'_{proM}$) were tested on promoter activity, by generating an artificial operon with *lacZ* (Fig. 6a). As expected, no β-galactosidase activity could be measured when utilizing $5'_{proM}$, as *proM* is part of the *argX* polycistronic operon (*argX_hisR_leuT_proM*)[35,36]. From the remaining two regions—$5'_{proK}$ and $5'_{proL}$—the latter gave a higher reporter signal and was therefore chosen as constitutive promoter for all three prolyl-tRNAs.

The effect of tRNA copy number increase was first assessed in the four reporter strains which harbor a HisL-TPPP variant each encoded by a series of one of the four distinct proline codons ($^{TPPP}$CC**G**, $^{TPPP}$CC**C**, $^{TPPP}$CC**U**, $^{TPPP}$CC**A**) (Fig. 6b). The CCG-specific ProK had a positive but only mild influence on pausing strength, solely when translating $^{TPPP}$CC**G**. One plausible explanation is that the native copy number of 900/cell is already close to saturating levels and accordingly overexpression does not substantially add to pausing strength reduction. With ProL we observed significantly reduced pauses when testing $^{TPPP}$CC**C** and $^{TPPP}$CC**U**, being again in line with the tRNA codon specificity. Interestingly, an increase in copy number of the general tRNA ProM had no major impact on reporter activity of the $^{TPPP}$CC**G**/**C**/**U** strains, indicative of a selection in favor of the more specialized tRNAs (ProK and ProL). Conversely, we saw a

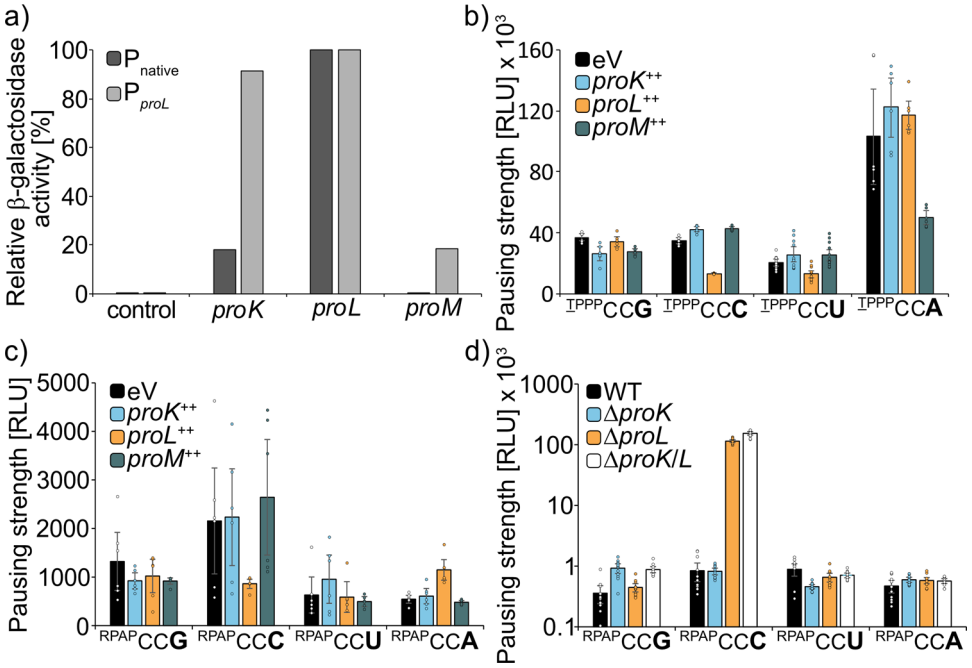



**Fig. 6 Influence of prolyl-tRNA copy number on the codon-dependent pausing strength at PP-motifs. a** Approximation of *E. coli* BW25113 cells carrying the weak HisL*_Lux operon (TPPP) with different proline codon usage were transformed with pBBR1 MCS4-*lacZ* plasmids encoding ProK, ProL, or ProM under the control of their corresponding native promoters. $n = 4$. **b** *E. coli* BW25113 cells carrying the weak HisL*_Lux operon (TPPP) were transformed with pBBR1 MCS4-*lacZ* plasmids encoding for ProK, ProL, or ProM under control of $P_{proL}$ and tested for bioluminescence emission. $n = 6$. **c** *E. coli* BW25113 cells carrying the "non-PP" HisL*_Lux operon (RPAP) were transformed with pBBR1-MCS4-*lacZ* plasmids encoding for ProK, ProL, or ProM under control of $P_{proL}$ and tested for bioluminescence emission. $n = 6$. **d** The "non-PP" HisL*Lux operon (RPAP) was genomically integrated in *E. coli* BW25113 deletion strains lacking either *proK* (Δ*proK*), *proL* (Δ*proL*), or both (Δ*proK/L*) and cells were tested for bioluminescence emission. $n = 12$, Error bars indicate 95% confidence intervals.

significant pausing strength reduction (>2-fold) for the $^{TPPP}$CC**A**, which can only be decoded by ProM.

Second, to separate PP-motif specific effects from those also occurring only with single prolines, a reference reporter set encoding RPAP-HisL variants was included into our study (Fig. 6c). Here, the previously observed minor alleviating effect at CCG codons on translational pausing upon ProK overexpression was lost. On the contrary, an increase in the copy number of ProL still significantly reduced pausing strength at CCC codons, yet no reduction of reporter activity for $^{RPAP}$CC**U** was observed. CCC codons are reportedly translated the slowest[5], which is in line with a general increase in luminescence compared to all HisL variants encoded by other proline codons. However, this does not explain the stimulatory effect on translational speed when overexpressing ProL: In their in vitro study on dipeptide synthesis with proline Pavlov et al. always employed bulk tRNA when measuring incorporation speed[5]. Accordingly, tRNA abundance effects were neglected. Our findings now indicate that the observed differences in dipeptide synthesis time might be partially due to ProL limitation. This idea is supported by the fact that the in vitro experiments in Pavlov et al. revealed CCU after CCC as the slowest codon to be decoded[5].

The tRNA abundance effect that differs most between consecutive and single prolines is for CCA (Fig. 6b, c). While translational pausing is alleviated by a factor of around three upon overexpression of ProM, we hardly found any changes when analyzing the $^{RPAP}$CC**A** reporter. Thus, our findings provide a rationale for the CCA codon bias in PP-proteins.

Third, we performed the converse experiment by deleting the two non-essential tRNA genes *proK* and *proL*[33], both individually —Δ*proK*, Δ*proL*—and in combination—Δ*proK/L*. These strains (Supplementary data file S3) were investigated on growth and cell morphology (Fig. S4) as well as on the effect they have on pausing strength (Fig. 6d). When analyzing the effects of *proK* and *proL* deletions on luminescence we saw the expected increase at $^{RPAP}$CC**G**, when *proK* is missing. The most striking results were obtained upon *proL* deletion. The light output at CCC significantly increased almost by a factor of 100, whereas translation of the $^{RPAP}$CC**U** reporter remained unaffected. This led us to conclude that the general tRNA ProM is a good decoder of CCU codons, as it can compensate for the lack of ProK. By contrast, the strong increase in pausing strength with CCC in Δ*proL* strains explains the necessity for a more specialized tRNA, which can outperform the ProM decoding capabilities at this specific codon. We therefore speculate that nature has evolved ProL predominantly to read CCC codons in order to compensate for its reduced translational speed. Additional reading of the "U" in the wobble position was acquired later, as a consequence of a mild advantage (Fig. 6b). This idea is also congruent with the identity of the ProL anticodon, which is GGG.

Taken together, we could show that tRNA abundance is a major driving force for the efficient translation of single and consecutive prolines.

**Proline codon choice finetunes protein copy number of the pH sensor CadC**. Based on our results, we hypothesized that codon choice within PP-motifs can be used as a regulatory means to tune the pausing strength according to stoichiometric requirements. In this regard, counterselection of certain codons would occur in order to prevent modulation of the pausing strength predetermined by the amino acid context. To test this hypothesis, we investigated codon choice in the PP-motif of the transcriptional activator CadC.

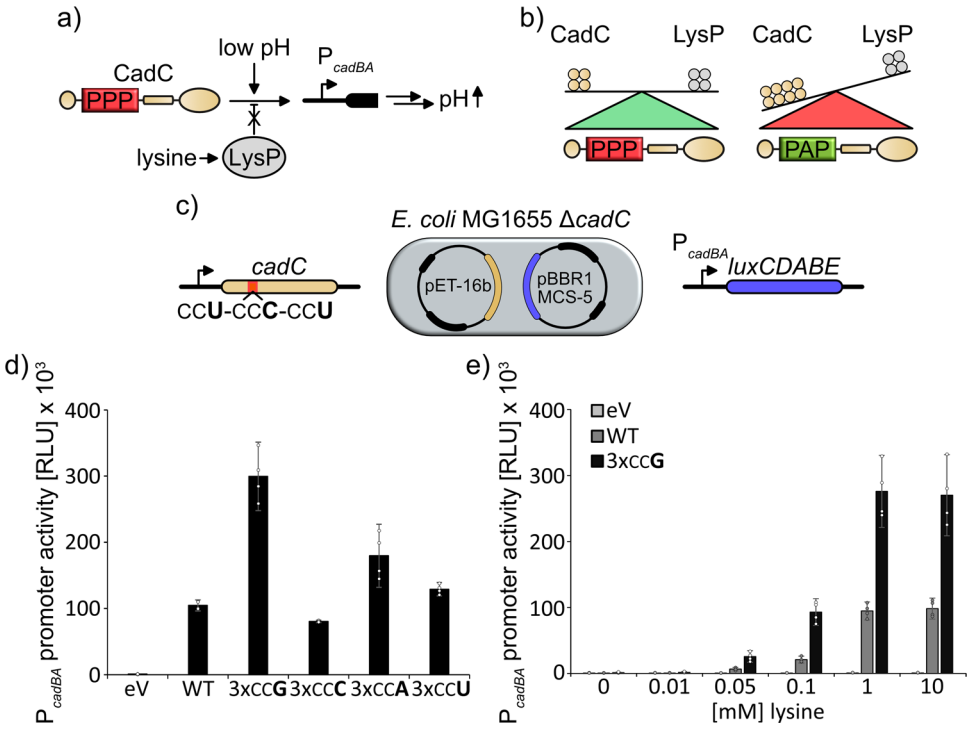



**Fig. 7 Codon choice modulates protein expression and ensures physiological protein stoichiometry of the Cad system. a** The Cad system. CadC is a pH sensor that induces expression of its target genes at low pH by binding to the *cadBA* promoter ($P_{cadBA}$). Expression of the corresponding gene products ultimately leads to an increase in pH. The lysine dependency of the acid stress response depends on stoichiometric expression of CadC and the co-sensor LysP. **b** The equilibrium of the protein copy numbers of CadC and LysP is ensured by a triproline motif within the CadC primary structure. Absence of the triproline results in deregulation of the acid stress response due to increased CadC copy number. **c** Reporter system used to test the *cadC* translation efficiency. *E. coli* MG1655 Δ*cadC* cells were transformed with pET-16B vectors encoding for wild type or proline codon-exchanged variants of CadC. Cells were cotransformed with pBBR1MCS-5 vectors carrying the *lux* genes under control of the $P_{cadBA}$ promoter. $P_{cadBA}$ promoter activity was assessed by measuring luminescence emission and used as a proxy for CadC copy number[14]. **d** $P_{cadBA}$ promoter activity under inducing conditions (pH = 5.8; 10 mM lysine) upon expression of wild-type CadC or proline codon-exchanged CadC variants where all proline codons in the pausing motif have been substituted by the same codon. $n = 4$. **e** $P_{cadBA}$ promoter activity at increasing external lysine concentrations. $P_{cadBA}$ induction when *cadC* contains the natural codon composition is shown in dark gray. $P_{cadBA}$ induction when *cadC* contains only CCG codons at the relevant PP-motif is shown in black. $n = 4$, Error bars indicate 95% confidence intervals.

CadC is a membrane-bound transcriptional regulator and part of the *E. coli* acid stress response[37–39]. The two external stimuli, mild acidic pH (<6.5) and lysine are needed to activate expression of the *cadBA* operon. While acidic conditions are sensed by CadC directly, lysine is recognized by a coregulator—the permease LysP. LysP directly interacts with CadC and a specific equilibrium between both proteins is crucial for an adequate transcriptional response (Fig. 7a)[14]. This equilibrium is strictly dependent on a triproline motif (aa120-122) within CadC[37] that is decoded from $CCU_{P120}$-$CCC_{P121}$-$CCU_{P122}$[14,32] and preceded by a serine (TCG).

As expression of *cadC* from pET-16b leads to physiological protein levels and an adequate pH-stress response in *E. coli* MG1655 cells[14], we generated plasmid-encoded CadC variants in which we unified the codons within the triproline motif (Fig. 7b). These were tested with a *lux* reporter controlled by $P_{cadBA}$ (Fig. 7c)[40]. Upon monitoring the maximal light output during 16 h of growth in minimal medium under CadC inducing conditions (pH 5.8 and supplemented with 10 mM lysine) we detected a threefold increase in $P_{cadBA}$ activity with CCG stretches in the CadC open reading frame while 3 × CCA, 3 × CCC, and 3 × CCU resulted in only subtle changes in light emission compared to the wild-type protein (Fig. 7d). As previously shown these changes in promoter activity reflect fluctuating copy numbers of the regulator[14]. Of note, only the changes between the CadC variants with 3 × CCA and 3 × CCG can be directly compared in terms of translation efficiency as a consequence of differences in tRNA abundance and codon anticodon pairing[5]. For the variants with 3 × CCC and 3 × CCU, the additional effect of ribosome slipping also causes decreased protein output.

To test for physiological repercussions of the elevated protein production with $CadC^{CCG}$, we performed the same experiment again, but tested different lysine concentrations (Fig. 7e). In this setup both the wild type and the $CadC^{CCG}$ variant reached the highest induction level at 1 mM lysine but the latter showed a threefold increased maximal light output. More importantly, $CadC^{CCG}$ turned on *cadBA* transcription already at 100 µM lysine. This concentration, however, is insufficient for pH neutralization. Thus, codon choice within the CadC triproline motif is crucial to maintain an optimal ratio between CadC and LysP in order to achieve an adequate stress response.

## Discussion

The theory of codon bias postulates the correlation between preferred codons and abundances of their iso-accepting tRNAs[41], thereby increasing translation efficiency[42] and accuracy[43]. Although the "tRNA abundance" theory also applies to proline codons (Fig. S1)[34], the strong correlation with incorporation velocities seems to be more important[5]. This explains, for example, why CCC is a rather neglected codon in *E. coli* as it interacts least efficiently with the $tRNA^{Pro}$–EF-Tu–GTP ternary

complex[5]. Moreover, a pair of CCC/U codons promotes ribosomal frameshifting (Fig. S2)[24].

Generally, proline pairs are difficult to translate as they cause ribosome stalling[7]. Their frequent occurrence in nature points to a selective advantage that outweighs the concomitant translational burden[16] and has even favored the emergence of a specialized elongation factor EF-P to aid in translation[14]. This advantage is due to the unique properties of polyprolines affecting protein structure[2] and function[10]. Although there is an evolutionary trend to reduce the translational load, we have previously identified specific regions where pausing by PP-motifs is favored to limit translation rates and to facilitate proper membrane insertion and correct folding[11].

Our previous work focused on the PP-motifs and their amino acid context. We have now extended our study to the transcript level, which led to several new insights into the relation between codon pairs and tRNA abundance.

First, we found that the codon bias in consecutive prolines differs significantly from that in single prolines, which helps to avoid slippery sequences (CCC/U-CCN) and to boost translation efficiency. Only in the regions where increased pausing time might be beneficial, such as the vicinity of the translational start and downstream of transmembrane helices, more slowly translating codons are favored (Fig. 2). Moreover, we have demonstrated the physiological importance of codon choice on one prominent example—the pH sensor CadC. Here, the silent mutation of prolines of $CCU_{P120}$-$CCC_{P121}$-$CCU_{P122}$ into 3 × CCG led to a deregulation of the acid stress response as a result of an increased protein copy number. Thus, there is a concerted adjustment of both codon usage within PP-motifs and their amino acid context, in turn allowing for a precise adjustment of protein copy numbers. We note, that additional factors such as mRNA structure or stability might also contribute to this effect.

Second, we have uncovered the specific effects associated with isoacceptor $tRNA^{Pro}$. Overall, we found that both overexpression and deletion of each individual prolyl-tRNA gene—*proK*, *proL*, or *proM*—affected translation at their cognate codons, regardless of their amino acid context. The most pronounced effect was observed with ProL and on CCC, whereas the benefit for the other target codon CCU was comparatively small. One reason for this might be the ProL anticodon—GGG, which could lead to different affinities between both of the recognized codons. Generally, dipeptide synthesis is slowest with CCC and CCU, which can also explain their scarceness in the genome. This rare usage was also one reason for having included *proL* into the pRARE plasmid in order to augment the yield and fidelity of heterologously produced proteins[44]. Our data now show that even under natural conditions ProL is limiting and thus increasing its copy number might have a positive effect on endogenously produced proteins (Fig. 6b, c). Especially, decoding of CCC benefits from ProL overproduction (Fig. 6) and thus heterologous expression of genes from GC-rich organisms such as *Streptomyces* species might lead to an increased yield. In this regard, it is notable that, e.g., *S. venezuelae* encodes a second copy of *proL*, presumably to circumvent this limitation (CCG: 52%, CCC: 43%, CCU: 3% CCA: 2%). Moreover, tRNA abundance explains also the selective pressure against CCA in weak PP-motifs. One might therefore speculate that recruitment of ProM to the ribosome is the rate limiting step in the weak context. Interestingly, increased copy number of ProM did not result in a decrease of ribosome pausing at any other codon than CCA. For CCC the reason might be in the poor interaction between the cmo5U34 modified base and the 3′ cytosine of the CCC codon. Besides that, even under control of $P_{proL}$, proM was less efficiently transcribed than the other tRNAs, indicating that the relative titers compared to the more specialized ProK and ProL for translation of CCG/C/U were not as strongly affected. The preference for CCG for which the cognate tRNA levels are close to saturation is consistent with this idea (Fig. 6). Further, CCG is enriched in PP-motifs at the expense of CCA in the top 20% of proteins in terms of translation efficiency (Fig. 2). In general, CCG seems to be the "best" proline codon in bacteria, when it comes to translation efficiency of codon pairs. This also explains why especially this codon is avoided in the CadC proline codon triplet, as here an extremely low copy number is crucial for a regulated acid stress response[14].

Thus, codon choice in proline codon pairs represents an elegant strategy to control translation efficiency and finetune protein copy numbers in bacteria.

## Material and methods

**Plasmid and strain construction**. All strains, plasmids, and oligonucleotides used in this study are listed and described in Supplementary data files S3–S5, respectively. All kits and enzymes were used according to manufacturer's instructions. Plasmid DNA was isolated using the Hi Yield® Plasmid Mini Kit from Süd Laborbedarf. DNA fragments were purified from agarose gels using the Hi Yield® Gel/PCR DNA fragment extraction kit from Süd Laborbedarf. All restriction enzymes, DNA modifying enzymes, and the Q5® high fidelity DNA polymerase for PCR amplification were purchased from New England BioLabs.

The pNPTS-138-R6KT_*hisL_luxCDABE* vector was generated by amplification of *hisGDCBHAF* operon leader peptide *hisL* from *E. coli* BW25113 genomic DNA and ligation into pNPTS-138-R6KT_$P_{BAD}$_*luxCDABE* after restriction with *Sph*I and *Nco*I. All variants of *hisL* (*hisL**, Supplementary data file S4) were generated by overlap extension PCR with mutagenized primers (Supplementary data file S5) from pNPTS-138-R6KT_*hisL_luxCDABE* and subsequent cut/ligation into pNPTS-138-R6KT_$P_{BAD}$_*luxCDABE* as described above. HisL*_lux reporter strains (Supplementary data file S3) were generated by single homologous recombination as described previously[29]. Briefly, *E. coli* WM3064 cells were transformed with pNPTS-138-R6KT vectors[28] carrying the *lux* operon preceded by either native or synthetic His-leader peptides (Supplementary data file S4). The vectors were transferred into the target *E. coli* BW25113 or Δ*efp* cells by conjugation. Transformants were selected from LB agar plates supplemented with kanamycin sulfate. PCR (Pf: HisL_chk_fw Pr: LuxC_chk_rev, Supplementary data file S5) and subsequent sequencing of the amplicon were used to verify incorporation of the correct *hisL**.

tRNA deletion strains (Supplementary data file S3) were generated according to the "Quick and Easy *E. coli* Gene Deletion Kit by Red®/ET® Recombination" protocol (Gene Bridges). In short, primers containing 50 base-pair overhangs corresponding to the tRNA loci (Supplementary data file S5) were used to amplify linear FRT-side-flanked resistance cassettes from either FRT-PGK-gb2-neo-FRT or FRT-PGK-gb2-cat-FRT (Supplementary data file S4) using PCR. *E. coli* BW25113 cells transformed with pRED/ET were transferred from a thick overnight culture into a fresh culture in LB by 1:100 dilution, which was grown at 37 °C for about 2 h, until an optical density at 600 nm ($OD_{600}$) of 0.3 was reached. Cells were then harvested and washed in 10% glycerol three times. The cells were subsequently transformed with the linear fragment by electroporation. Successful integration was confirmed by selective growth on LB plates containing either kanamycin sulfate or chloramphenicol and by PCR. Loss of the temperature sensitive pRED/ET plasmid was confirmed by selective growth on LB plates containing carbenicillin sodium salt or no antibiotic. To remove the chromosomally integrated resistance cassettes, the corresponding strains were transformed

with the 707-FLPe plasmid (Supplementary data file S4) and transformants were subsequently inoculated in LB and grew at 30 °C for 2 h before shifting the temperature to 37 °C for overnight incubation. On the next day, cells were streaked out on LB plates and incubated overnight at 37 °C. Successful removal of the resistance cassettes and the temperature sensitive 707-FLPe plasmid was confirmed by selecting cells on plates with and without antibiotic and subsequent sequencing of the corresponding loci after colony PCR.

Plasmids for expression of *E. coli* tRNAs under control of their native promoters were generated by amplification of the corresponding genes and putative regulatory regions from *E. coli* BW25113 genomic DNA using specific primers (Supplementary data file S5) and subsequent cut/ligation into the pBBR1-MCS4-*lacZ* vector[29] (Supplementary data file S4). Plasmids for expression of *E. coli* tRNAs under control of the *proL* promoter were generated using primers with a 70 BP overhang corresponding to $P_{proL}$ (Supplementary data file S5) and subsequent cut/ligation into pBBR1-MCS4-*lacZ*.

Plasmids for quantification of ribosome slipping were generated by overlap extension PCR using primers with the sequence ATTAACCATGGGGNNN*TAG***G**ACTAAAAAAATTTCATTC (Supplementary data file S5) and pBAD_HisA-*luxCDABE*[27] (Supplementary data file S4) as template. The first underlined sequence of the primer designates the initial open reading frame coding for a short peptide that stops at the TAG codon (italic). NNN designates the mutagenized region coding for the slipping sequence. The single base **G** (bold) allows the +1 frameshift into the *luxCDABE* open reading frame which is represented by the second underlined sequence.

**Growth conditions**. *E. coli* cells were routinely grown in Miller modified Lysogeny Broth (LB)[45,46] at 37 °C aerobically under agitation, if not indicated otherwise 1.5% (w/v) agar were used to solidify media when required. Antibiotics were added at the following concentrations: 100 µg/ml carbenicillin sodium salt, 50 µg/ml kanamycin sulfate, 20 µg/ml gentamycin sulfate. Plasmids carrying $P_{BAD}$[47] were induced with L-arabinose at a final concentration of 0.2% (w/v).

**Measurement of pausing strength in vivo**. Pausing strength at PP-motifs was determined by measuring light output of the *lux* operon under the control of a synthetic His-leader peptide (HisL*) (Figs. 4c–6). Cells carrying the reporter were inoculated in 96-well plates (Sarstedt TC-Plate 96-Well, Standard d, F) with each well containing 200 µl of LB supplemented with kanamycin sulfate and incubated in an Eppendorf Thermomixer comfort at 37 °C and 550 rpm for at least 16 h. When expressing *E. coli* tRNAs from MCS4 plasmids, carbenicillin sodium salt was also added to the medium. On the next morning, Corning® 96-well flat clear bottom black polystyrene TC-treated microplates containing 200 µl of LB—supplemented with either kanamycin sulfate alone or in combination with carbenicillin sodium salt—were inoculated with 2 µl of overnight culture. The plates were directly transferred to a Tecan Spark® plate reader. Absorption at 600 nm (Number of flashes: 10; Settle time: 50 ms) and luminescence emission (Attenuation: none; Settle time: 50 ms; Integration time: 200 ms) were determined in between 10-min cycles of agitation (orbital, 180 rpm, amplitude: 3 mm) for around 16 h.

**β-Galactosidase activity assay**. *E. coli* HisL* reporter strains (Supplementary data file S3) containing plasmids for expression of *E. coli* tRNA (Supplementary data file S4) were inoculated in 1.5 ml LB containing kanamycin sulfate and carbenicillin sodium salt and cultivated overnight in an Eppendorf Thermomixer comfort at 37 °C under microaerobic conditions and agitation at 650 rpm. On the next day, the optical density ($OD_{600}$) was determined in 1 ml volumes containing 0.5 ml overnight culture and 0.5 ml of fresh LB medium. In total, 0.5 ml of overnight culture were transferred to a new 2 ml Eppendorf reaction tube. Cells were harvested by centrifugation and subsequently resuspended in 1 ml Buffer Z (0.06 M $Na_2HPO_4$, 0.04 M $NaH_2PO_4$, 0.01 M KCl, 0.001 M $MgSO_4$). In total, 0.1 ml Chloroform and 0.05 ml 0.1 % SDS were added and the suspension was mixed by vortexing. Samples were preincubated at 30 °C for 5 min. The reaction was started by adding 0.2 ml of *ortho*-Nitrophenyl-β-galactoside solution (4 mg/ml in Buffer Z) and stopped by adding 0.5 ml 1 M $Na_2CO_3$ when yellow color formation was observed or after 5 min of incubation. The time between starting and stopping the reaction was noted in seconds. The samples were centrifuged at 20,000 × *g* for 10 min and 1 ml of the reaction solution was transferred to a cuvette. Absorbance at 420 nm was determined and Miller units (MU) were calculated as $MU = 1000 \times Abs_{420} \times t^{-1} \times V^{-1} \times Abs_{600}{}^{-1}$,[48].

**Measurement of *cadBA* promoter activity in vivo**. Activity of the *cadBA* promoter upon exchange of proline codons within the *cadC* gene (Fig. 7) was assessed using a luminescence reporter as described before[40]. *E. coli* MG1655 Δ*cadC* cells were cotransformed with the reporter plasmid pBBR1-MCS5-$P_{cadBA}$-*lux* (Supplementary data file S4) and a pET16B vector for ectopic expression of either the wild-type *cadC* (pET16B-*cadC*) or a copy with silent mutations in the proline codon triplet $CCU_{P120}$-$CCC_{P121}$-$CCU_{P122}$ leading to pET16B-*cadC*_3xCCG, pET16B-*cadC*_3xCCC, pET16B-*cadC*_3xCCU, and pET16B-*cadC*_3xCCA (Supplementary data file S4). As control the reporter plasmid was cotransformed with pET16B. Transformants were incubated in 200 µl of minimal medium developed by Epstein and Kim[49] pH 7.6 supplemented with gentamycin sulfate, carbenicillin sodium salt, and 0.2% glucose (w/v) in 96-well plates in a Eppendorf Thermomixer comfort at 37 °C and agitation of 550 rpm overnight. On the next day, 2 µl of overnight culture were transferred to 200 µl of fresh medium supplemented with gentamycin sulfate and carbenicillin sodium salt in a Corning® 96-well flat clear bottom black polystyrene TC-treated microplate. Here, KE pH 5.8, 0.2% (w/v) glucose with varying concentrations of lysine was used. Bioluminescence emission (Attenuation: none; Settle time: 50 ms; Integration time: 200 ms) and growth (Wavelength: 600 nm; Number of flashes: 10; Settle time: 50 ms) were monitored in a Tecan Spark® in 10-min intervals during agitation (orbital, 180 rpm, amplitude: 3 mm) for around 16 h.

**Quantification of +1 translational frameshifting in vivo**. *E. coli* BW25113 cells were transformed with plasmids containing pBAD-HisA-*luxCDABE* plasmids (Supplementary data file S4) in which the *luxC* gene was cloned out of frame as described above. In total, 200 µl LB containing the carbenicillin sodium salt were inoculated with 2 µl of an overnight culture of the desired transformants. To induce expression of the slipping vector arabinose was added to a final concentration of 0.2% (w/v). The measurement was performed in a Tecan Spark® reader in Corning® 96-well flat clear bottom black polystyrene TC-treated microplates. Bioluminescence emission (Attenuation: none; Settle time: 50 ms; Integration time: 200 ms) and growth (Wavelength: 600 nm; Number of flashes: 10; Settle time: 50 ms) were monitored in a Tecan Spark® in 10-min intervals during agitation (orbital, 180 rpm, amplitude: 3 mm) for around 16 h.

### Bioinformatic analyses

*cDNA and protein sequences from* E. coli. The cDNA and protein sequences of 4352 *E. coli* K-12 MG1655 genes were downloaded from the OMA database[50]. The cDNA and protein sequences of genes from the other 15 bacteria were downloaded from the Ensembl Bacteria database (Supplementary data file S1)[51].

*Identification of PP-motifs in protein sequences*. PP-motifs in protein sequences were identified using the *fuzzpro* program from the EMBOSS package[52]. The PP-motifs were defined as in[11], i.e., XX-*n*P-X where $n \geq 2$ and X could be any non-proline amino acid.

*Protein abundance and translation efficiency*. We obtained the protein abundance and translation efficiency values for *E. coli* genes as described previously[11]: protein abundance data covering 2163 *E. coli* genes was from[53,54]; transcription levels of 2710 *E. coli* genes under standard growth conditions were downloaded from the ASAP database[55]. For each of the 1743 genes present in both datasets, we calculated the translation efficiency as the ratio between its protein abundance and transcription level.

*Transmembrane segments of the* E. coli *proteins*. Sequence positions of 5672 transmembrane segments within 912 α-helical transmembrane proteins were downloaded from the Uniprot database[56]. Data for the *E. coli* K-12 strain (taxonomy ID 83333) were used instead of *E. coli* K-12 MG1655 (taxonomy ID 511145), since the reviewed data of the latter are unavailable in the Uniprot database[56].

*Statistics and reproducibility*. Sample size: sample size in biochemical experiments was chosen to be at least $n = 4$. This sample size was calculated from Lehr's formula where the effect size was at least twice the standard deviation of experiments using wild-type cells. Biological replicates were defined as single colonies derived from culture plates. No data were excluded from the analysis. Replication: initial experiments using the His-Leader system (Fig. 4c) were conducted as technical replicates both in a Tecan Spark and a Tecan F500 reader showing qualitatively comparable results. Experiments on codon choice variation (Fig. 5b) were conducted both in 200 and 150 µl showing quantitatively comparable results.

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

## Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information. Source data underlying graphs presented in the main figures are available in Supplementary data file S6. No datasets were generated during this study.

Received: 4 September 2020; Accepted: 16 April 2021;
Published online: 17 May 2021

## References

1. Wedemeyer, W. J., Welker, E. & Scheraga, H. A. Proline cis-trans isomerization and protein folding. *Biochemistry* **41**, 14637–14644 (2002).
2. Adzhubei, A. A., Sternberg, M. J. & Makarov, A. A. Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.* **425**, 2100–2132 (2013).
3. Muto, H. & Ito, K. Peptidyi-prolyl-tRNA at the ribosomal P-site reacts poorly with puromycin. *Biochem. Bioph. Res. Co.* **366**, 1043–1047 (2008).
4. Wohlgemuth, I., Brenner, S., Beringer, M. & Rodnina, M. V. Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. *J. Biol. Chem.* **283**, 32229–32235 (2008).
5. Pavlov, M. Y. et al. Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc. Natl Acad. Sci. USA* **106**, 50–54 (2009).
6. Hayes, C. S. Proline residues at the C terminus of nascent chains induce SsrA tagging during translation termination. *J. Biol. Chem.* **277**, 33825–33832 (2002).
7. Tanner, D. R., Cariello, D. A., Woolstenhulme, C. J., Broadbent, M. A. & Buskirk, A. R. Genetic identification of nascent peptides that induce ribosome stalling. *J. Biol. Chem.* **284**, 34809–34818 (2009).
8. Huter, P. et al. Structural basis for polyproline-mediated ribosome stalling and rescue by the translation elongation factor EF-P. *Mol. Cell* **68**, 515–527 e516 (2017).
9. Morgan, A. A. & Rubenstein, E. Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *PLoS ONE* **8**, e53785 (2013).
10. Starosta, A. L. et al. A conserved proline triplet in Val-tRNA synthetase and the origin of elongation factor P. *Cell Rep.* **9**, 476–483 (2014).
11. Qi, F., Motz, M., Jung, K., Lassak, J. & Frishman, D. Evolutionary analysis of polyproline motifs in *Escherichia coli* reveals their regulatory role in translation. *PLoS Comput. Biol.* **14**, e1005987 (2018).
12. Pinheiro, B. et al. Structure and function of an elongation factor P subfamily in actinobacteria. *Cell Rep.* **30**, 4332–433 (2020).
13. Doerfel, L. K. et al. EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science* **339**, 85–88 (2013).
14. Ude, S. et al. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* **339**, 82–85 (2013).
15. Gutierrez, E. et al. eIF5A promotes translation of polyproline motifs. *Mol. Cell* **51**, 1–11 (2013).
16. Lassak, J., Wilson, D. N. & Jung, K. Stall no more at polyproline stretches with the translation elongation factors EF-P and IF-5A. *Mol. Microbiol.* **99**, 219–235 (2016).
17. Tuller, T. et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354 (2010).
18. Hersch, S. J., Elgamal, S., Katz, A., Ibba, M. & Navarre, W. W. Translation initiation rate determines the impact of ribosome stalling on bacterial protein synthesis. *J. Biol. Chem.* **289**, 28160–28171 (2014).
19. Hersch, S. J. et al. Divergent protein motifs direct elongation factor P-mediated translational regulation in *Salmonella enterica* and *Escherichia coli*. *mBio* **4**, e00180–00113 (2013).
20. Peil, L. et al. Distinct X/PP/X sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc. Natl Acad. Sci. USA* **110**, 15265–15270 (2013).
21. Elgamal, S. et al. EF-P dependent pauses integrate proximal and distal signals during translation. *PLoS Genet.* **10**, e1004553 (2014).
22. Starosta, A. L. et al. Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site. *Nucleic Acids Res.* **42**, 10711–10719 (2014).
23. Yourno, J. & Tanemura, S. Restoration of in-phase translation by an unlinked suppressor of a frameshift mutation in *Salmonella typhimurium*. *Nature* **225**, 422–426 (1970).
24. Gamper, H. B., Masuda, I., Frenkel-Morgenstern, M. & Hou, Y. M. Maintenance of protein synthesis reading frame by EF-P and m(1)G37-tRNA. *Nat. Commun.* **6**, 7226 (2015).
25. Chevance, F. F., Le Guyon, S. & Hughes, K. T. The effects of codon context on in vivo translation speed. *PLoS Genet.* **10**, e1004392 (2014).
26. Johnston, H. M., Barnes, W. M., Chumley, F. G., Bossi, L. & Roth, J. R. Model for regulation of the histidine operon of *Salmonella*. *Proc. Natl Acad. Sci. USA* **77**, 508–512 (1980).
27. Volkwein, W. et al. Switching the Post-translational modification of translation elongation factor EF-P. *Front. Microbiol.* **10**, 1148 (2019).
28. Lassak, J., Henche, A. L., Binnenkade, L. & Thormann, K. M. ArcS, the cognate sensor kinase in an atypical Arc system of *Shewanella oneidensis* MR-1. *Appl. Environ. Microbiol.* **76**, 3263–3274 (2010).
29. Fried, L., Lassak, J. & Jung, K. A comprehensive toolbox for the rapid construction of *lacZ* fusion reporters. *J. Microbiol. Methods* **91**, 537–543 (2012).
30. Sezonov, G., Joseleau-Petit, D. & D'Ari, R. *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.* **189**, 8746–8749 (2007).
31. Bertels, F., Merker, H. & Kost, C. Design and characterization of auxotrophy-based amino acid biosensors. *PLoS ONE* **7**, e41349 (2012).
32. Woolstenhulme, C. J., Guydosh, N. R., Green, R. & Buskirk, A. R. High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep.* **11**, 13–21 (2015).
33. Näsvall, S. J., Chen, P. & Bjork, G. R. The modified wobble nucleoside uridine-5-oxyacetic acid in $tRNA^{Pro}$(cmo5UGG) promotes reading of all four proline codons in vivo. *RNA* **10**, 1662–1673 (2004).
34. Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* **260**, 649–663 (1996).

35. Kröger, C. et al. An infection-relevant transcriptomic compendium for *Salmonella enterica* Serovar Typhimurium. *Cell Host Microbe* **14**, 683–695 (2013).
36. Mohanty, B. K., Petree, J. R. & Kushner, S. R. Endonucleolytic cleavages by RNase E generate the mature 3' termini of the three proline tRNAs in *Escherichia coli*. *Nucleic Acids Res.* **44**, 6350–6362 (2016).
37. Buchner, S., Schlundt, A., Lassak, J., Sattler, M. & Jung, K. Structural and functional analysis of the signal-transducing linker in the pH-responsive one-component system CadC of *Escherichia coli*. *J. Mol. Biol.* **427**, 2548–2561 (2015).
38. Schlundt, A. et al. Structure-function analysis of the DNA-binding domain of a transmembrane transcriptional activator. *Sci. Rep.* **7**, 1051 (2017).
39. Jung, K., Fabiani, F., Hoyer, E. & Lassak, J. Bacterial transmembrane signalling systems and their engineering for biosensing. *Open Biol.* **8**, 180023 (2018).
40. Brameyer, S. et al. DNA-binding directs the localization of a membrane-integrated receptor of the ToxR family. *Commun. Biol.* **2**, 4 (2019).
41. Ikemura, T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**, 13–34 (1985).
42. Ikemura, T. Correlation between the abundance of *Escherichia coli* transfer-RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that Is optimal for the *Escherichia coli* translational system. *J. Mol. Biol.* **151**, 389–409 (1981).
43. Akashi, H. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* **136**, 927–935 (1994).
44. Novy, R., Drott, D., Yaeger, K. & Mierendorf, R. Overcoming the codon bias of *E. coli* for enhanced protein expression. *Innovations* **12**, 1–3 (2001).
45. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293–300 (1951).
46. Miller, J. H. Experiments in molecular genetics. *Cold Spring Harb. Lab. Press* **221–222**, 263–274 (1972).
47. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose $P_{BAD}$ promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
48. Tetsch, L., Koller, C., Haneburger, I. & Jung, K. The membrane-integrated transcriptional activator CadC of *Escherichia coli* senses lysine indirectly via the interaction with the lysine permease LysP. *Mol. Microbiol.* **67**, 570–583 (2008).
49. Epstein, W. & Kim, B. S. Potassium transport loci in *Escherichia coli* K-12. *J. Bacteriol.* **108**, 639–644 (1971).
50. Altenhoff, A. M. et al. The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res.* **46**, D477–D485 (2018).
51. Howe, K. L. et al. Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Res.* **48**, D689–D695 (2020).
52. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
53. Wisniewski, J. R. & Rakus, D. Multi-enzyme digestion FASP and the 'Total Protein Approach'-based absolute quantification of the *Escherichia coli* proteome. *J. Proteom.* **109**, 322–331 (2014).
54. Wisniewski, J. R. & Rakus, D. Quantitative analysis of the *Escherichia coli* proteome. *Data Brief.* **1**, 7–11 (2014).
55. Glasner, J. D. et al. ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Res.* **31**, 147–151 (2003).
56. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–d515 (2019).


## Acknowledgements

We thank Dr. Kerstin Lassak and Dr. Natalia Tschowri for fruitful discussions and constructive criticism. We thank Urte Tomasiunaite for her help during the revision of this manuscript. F.Q. is supported by the National Natural Science Foundation of China (Project No. 32000462) and Scientific Research Funds of Huaqiao University. J.L. & K.J. gratefully acknowledge financial support from the DFG Research Training Group GRK2062 (Molecular Principles of Synthetic Biology). Moreover, J.L. is grateful for DFG grant LA 3658/1-1 and D.F. for DFG grant FR 1411/17-1

## Author contributions

The HisL*-Lux reporter concept was developed by J.L. R.K., A.S., J.M., and J.L. constructed HisL*-lux reporter strains and plasmids. Corresponding in vivo measurements were performed by R.K., A.S. and J.M. tRNA gene deletions were made by R.K. and J.M. Reporters for quantification of +1 frameshifting were constructed and corresponding assay performed by R.K. and J.M. R.K. constructed *cadC* mutations and recorded their effect on pH-regulation. All bioinformatic analyses were performed by F.Q. and D.F. The study was designed by F.Q., R.K., D.F. and J.L. with contributions of K.J. The manuscript was written by F.Q., R.K., and J.L. with contributions of D.F. and K.J.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02115-z.

**Correspondence** and requests for materials should be addressed to D.F. or J.L.

**Reprints and permission information** is available at http://www.nature.com/reprints

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

**Supplemental Information Chapter 2**

# Proline codon pair selection determines ribosome pausing strength and translation efficiency in bacteria

Ralph Krafczyk[1]*, Fei Qi[2,3]*, Alina Sieber[1] , Judith Mehler[1] , Kirsten Jung[1], Dmitrij Frishman[3+] & Jürgen Lassak[1+]

[1] Department of Biology I, Microbiology, Ludwig-Maximilians-Universität München, Martinsried, Germany

[2] Institute of Genomics, School of Biomedical Sciences, Huaqiao University, Xiamen, China

[3] Department of Bioinformatics, Wissenschaftzentrum Weihenstephan, Technische Universität München, Freising, Germany

*These authors contributed equally to this work.

[+]To whom correspondence should be addressed. E-mail: d.frishman@wzw.tum.de or : juergen.lassak@lmu.de

## Supplementary information

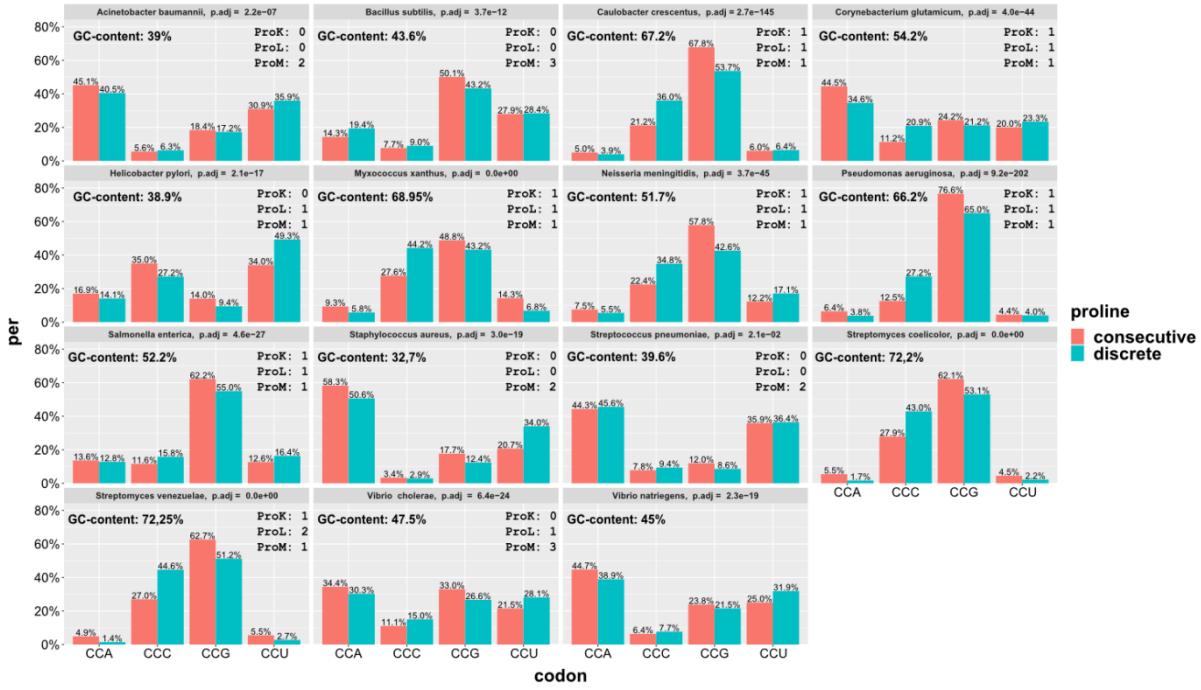



**Fig. S1.** Codon usage of single and consecutive prolines in the proteomes of 15 bacteria. The name of corresponding bacterium and the adjusted p-value of chi-squared test are shown in the title of each panel. The genomic GC-content and copy number of tRNA genes for individual bacteria are shown. tRNA copy numbers were derived from tRNA database (http://gtrnadb.ucsc.edu/)

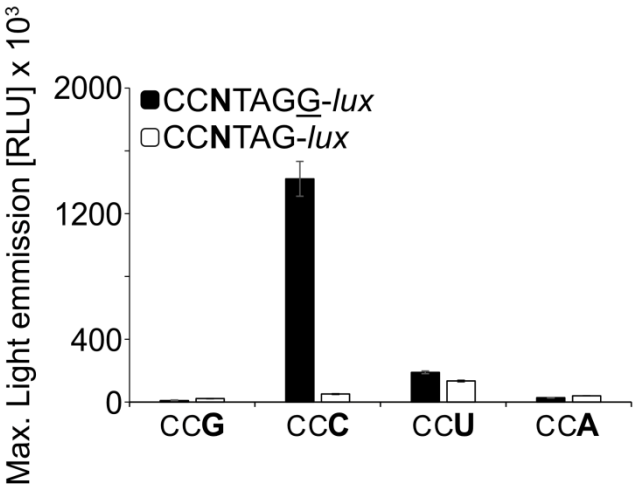



**Fig. S2. Ribosome slipping and translational readthrough at proline codons.** Maximal luminescence signals measured when the proline codons were coded in frame with the stop codon but in the -1 frame of the *lux*-operon (CC**N**TAGG-*lux*) are shown in black. Maximal luminescence signals measured when the proline codons were coded frame with the stop codon and the *lux*-operon (CC**N**TAG-*lux*) are shown in white.

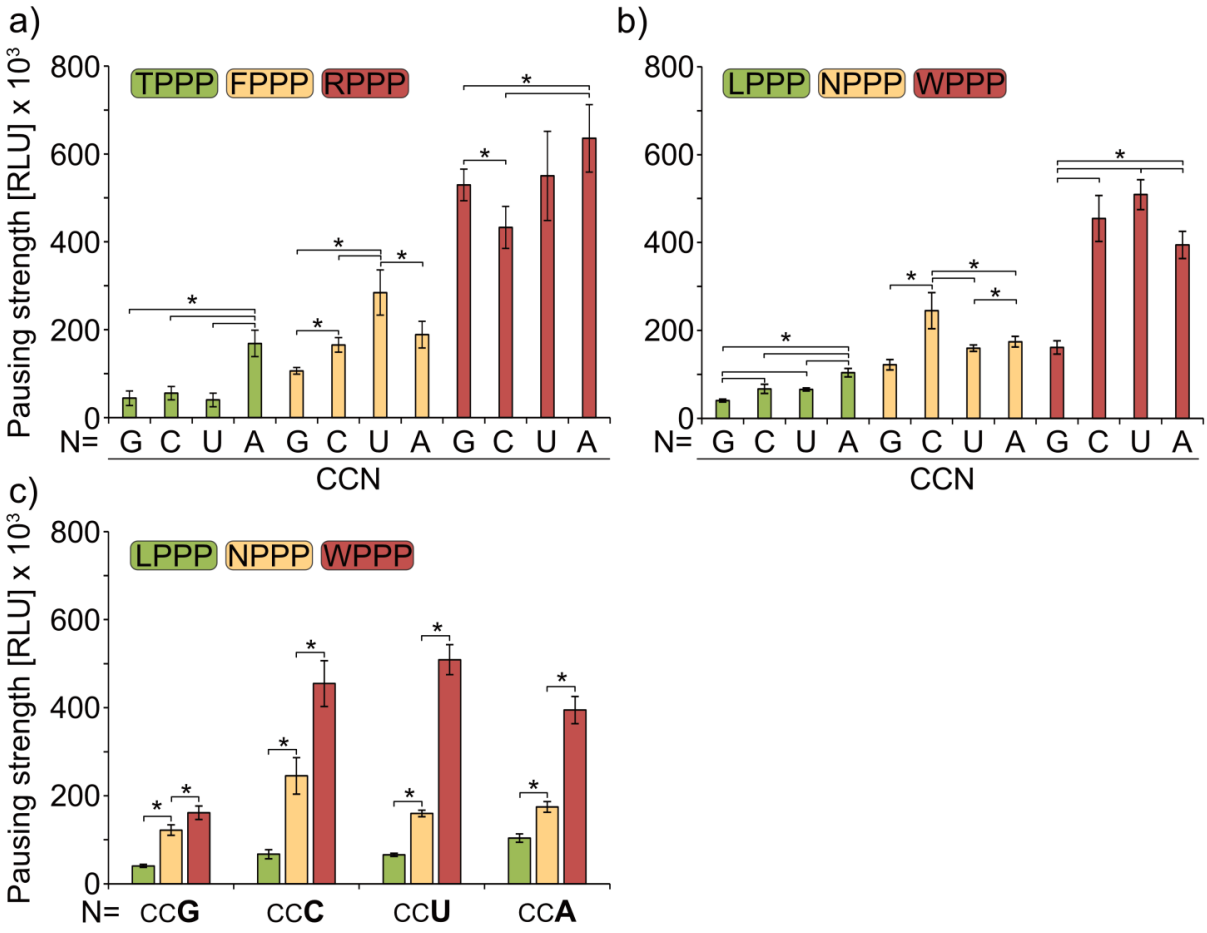



**Fig. S3. Codon-dependent pausing strength at weak, intermediate and strong PP-motifs.** HisL*_Lux carrying PP-motifs of varying pausing strength with different proline codon usage **A)** weak – TPPP: green; intermediate – FPPP: yellow; strong – RPPP: red; **B,C)** weak – LPPP: green; intermediate – NPPP: yellow; strong – WPPP: red) were chromosomally integrated in *E. coli* BW25113 and tested for maximal luminescence emission. n = 12, Error bars indicate 95% confidence intervals. Data for Figure 3A was duplicated from Figure 5B for better overview. Data for Figure 3C was duplicated from Figure 3B for better overview. Statistically significant differences according to unpaired two-sided t-tests (*p*-value < 0.05) are indicated by asterisks.

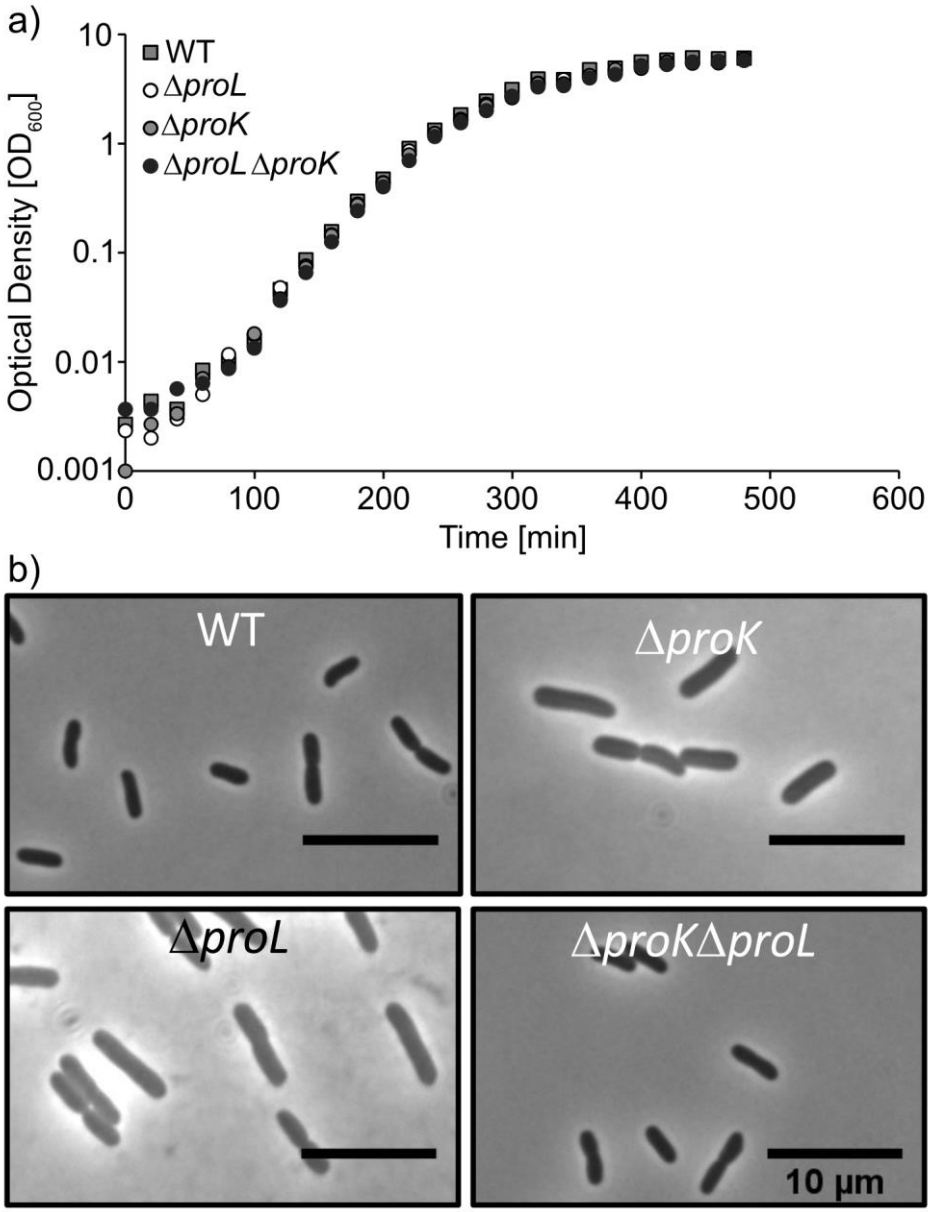



**Fig. S4. Phenotypic characterization of *E. coli* BW25113 tRNA-deletion strains.** A) Growth in LB medium at 37 °C under aerobic conditions. B) Phase contrast microscopy of exponentially growing cells.

# 3 EF-P and its paralog EfpL (YeiP) differentially control translation of proline-containing sequences

Sieber A*, Parr M*, von Ehr J, Dhamotharan K, Kielkowski P, Brewer T, Schäpers A, Krafczyk R, Qi F, Schlundt A, Frishman D & Lassak J. 2024. EF-P and its paralog EfpL (YeiP) differentially control translation of proline-containing sequences. *Nat. Commun.* **15**, 10465

*Authors contributed equally

nature communications

Article

https://doi.org/10.1038/s41467-024-54556-9

# EF-P and its paralog EfpL (YeiP) differentially control translation of proline-containing sequences

Received: 30 April 2024

Accepted: 13 November 2024

Published online: 02 December 2024

Check for updates

Alina Sieber[1,8], Marina Parr[2,8], Julian von Ehr[3,4], Karthikeyan Dhamotharan[3], Pavel Kielkowski[5], Tess Brewer[1], Anna Schäpers[1], Ralph Krafczyk[1], Fei Qi[6], Andreas Schlundt[3,7], Dmitrij Frishman[2] & Jürgen Lassak[1]

Polyproline sequences are deleterious to cells because they stall ribosomes. In bacteria, EF-P plays an important role in overcoming such polyproline sequence-induced ribosome stalling. Additionally, numerous bacteria possess an EF-P paralog called EfpL (also known as YeiP) of unknown function. Here, we functionally and structurally characterize EfpL from *Escherichia coli* and demonstrate its role in the translational stress response. Through ribosome profiling, we analyze the EfpL arrest motif spectrum and find additional sequences beyond the canonical polyproline motifs that both EF-P and EfpL can resolve. Notably, the two factors can also induce pauses. We further report that EfpL can sense the metabolic state of the cell via lysine acylation. Overall, our work characterizes the role of EfpL in ribosome rescue at proline-containing sequences, and provides evidence that co-occurrence of EF-P and EfpL is an evolutionary driver for higher bacterial growth rates.

Decoding genetic information at the ribosome is a fundamental trait shared among all living organisms. However, translation of two or more consecutive prolines leads to ribosome arrest[1–6]. To allow translation to continue, nearly every living cell is equipped with a specialized elongation factor called e/aIF5A in eukaryotes and archaea, or EF-P in bacteria[7,8]. Upon binding close to the ribosomal tRNA exiting site (E-site), EF-P stimulates peptide bond formation by stabilizing and orienting the peptidyl-tRNA$^{Pro}$[9,10]. EF-P has a three-domain structure that spans both ribosomal subunits[10,11] and consists of an N-terminal Kyprides, Ouzounis, Woese (KOW) domain and two oligonucleotide binding (OB) domains[12], together mimicking tRNA in size and shape[13]. Although this structure is conserved among all EF-P homologs[14], bacteria have evolved highly diverse strategies to facilitate proper interactions between EF-P and the CCA end of the P-site tRNA$^{Pro}$. For instance, in *Escherichia coli*, a conserved lysine K34 at the tip of the loop bracketed by two beta strands β3 and β4 (β3Ωβ4) of the KOW domain is post-translationally activated by β-D-lysylation using the enzyme EpmA[15–19]. Firmicutes such as *Bacillus subtilis* elongate lysine K32 of their EF-P by 5-aminopentanolylation[20], while e.g., in β-proteobacteria or pseudomonads, an arginine is present in the equivalent position, which is α-rhamnosylated by the glycosyltransferase EarP[14,21,22]. Among the remaining EF-P subtypes the paralogous YeiP (from now on termed EfpL for "EF-P like") sticks out, as it forms a highly distinct phylogenetic branch (Fig. 1A; Supplementary Fig. 1)[14,23]. However, to date, the molecular function of EfpL remains enigmatic[24]. Bioinformatic analyses based on AlphaFold predictions indicate that EF-P-like proteins have a three-domain structure similar to EF-P, but they only share about 30% sequence similarity. Across the

[1]Faculty of Biology, Microbiology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany. [2]Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, Freising, Germany. [3]Institute for Molecular Biosciences and Biomolecular Resonance Center (BMRZ), Goethe University Frankfurt, Frankfurt, Germany. [4]IMPRS on Cellular Biophysics, Frankfurt, Germany. [5]Department of Chemistry, Institut für Chemische Epigenetik (ICEM), Ludwig-Maximilians-Universität München, Munich, Germany. [6]State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen, China. [7]Institute of Biochemistry, University of Greifswald, Greifswald, Germany. [8]These authors contributed equally: Alina Sieber, Marina Parr. e-mail: juergen.lassak@lmu.de

https://doi.org/10.1038/s41467-024-54556-9

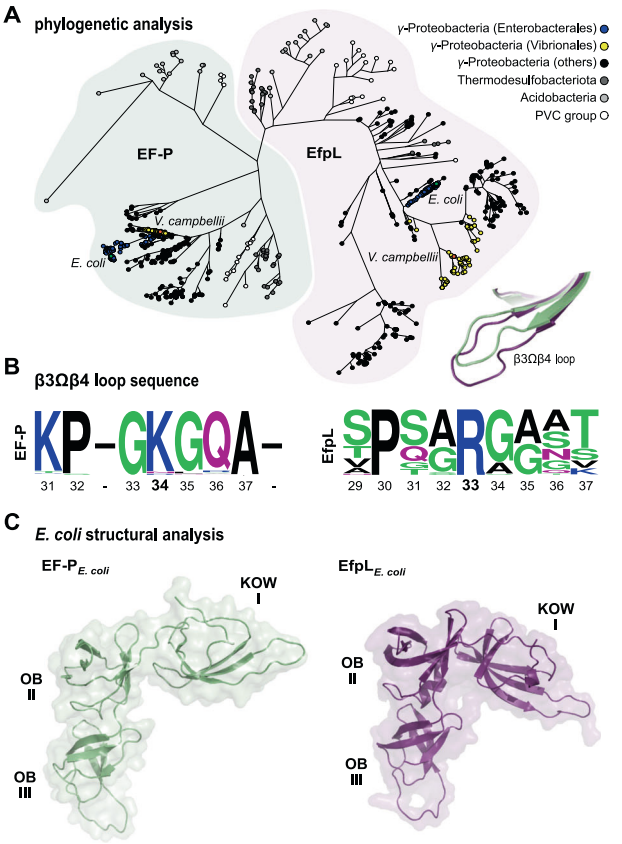



**Fig. 1 | Structural and phylogenetic analysis of the EfpL subgroup.** **A** Phylogenetic tree of EfpL (purple) and co-occurring EF-Ps (green). Colors of tip ends depict bacterial clades. Comparison of the KOW β3Ωβ4 loop in *E. coli* EF-P (taken from PDB: 6ENU; green) and EfpL (PDB: 8S8U, this study; purple). **B** Sequence logos[104] of β3Ωβ4 loop of EfpL and co-occurring EF-Ps. **C** Comparison of structures of *E. coli* EF-P (taken from PDB: 6ENU) and EfpL (PDB: 8S8U, chain B, this study) with overall fold views and three domains.

three domains, the C-terminal OB-domain shows the highest similarity between the two proteins. This domain's primary role is to interact with the small ribosomal subunit and the anticodon stem loop of the P-site tRNA. Notably, both EF-P and EfpL contain a tyrosine and an arginine in position 183 and 186, respectively (according to *E. coli* EF-P numbering), which are close enough to form hydrogen bonds with A42 of the P-site tRNA and G1338 within helix h29 of the 16S rRNA[10]. By contrast, the key residues in the KOW domain of EF-P, as well as the residue involved in specific recognition of prolyl-tRNA in stalled ribosomes, are less conserved. This in turn suggests that EfpL's role in translation diverges from those of canonical EF-Ps. In the frame of this study, we solve the structure of *E. coli* EfpL (EfpL) and uncover its role in translation of XP(P)X-containing proteins: Through ribosome profiling, we explore the EfpL arrest motif spectrum and uncover additional sequences beyond the typical polyproline motifs that both EF-P and EfpL can resolve. Additionally, these factors can also trigger translational pauses. Moreover, we demonstrate that EfpL is capable of detecting the cell's metabolic state via lysine acylation.

## Results

### Structural and phylogenetic analysis of EfpL revealed unique features in the β3Ωβ4 loop

We began our study by recapitulating a phylogenetic tree of EF-P in order to extract the molecular characteristics of the EfpL subgroup. A collection of 4736 complete bacterial genomes from a representative set that covers species diversity was obtained from the RefSeq database[25]. From these organisms, we extracted 5448 EF-P homologs and identified the branch that includes the "elongation factor P-like protein" of *E. coli*. This subfamily comprises 528 sequences (Supplementary Figs. 1 and 2; Supplementary Data 1) and is characterized by a number of unique features (Fig. 1). First, we observed that EfpL is predominantly found in Proteobacteria of the γ-subdivision but also in Thermodesulfobacteria, Acidobacteria and the Planctomycetes/Verrucomicrobia/Chlamydiae-group (Fig. 1A). This suggests a similar but more specialized role in translation than EF-P. Second, we noted that the EfpL branch is most closely affiliated but still separated from the arginine-type EF-P subgroup, which is activated by α-rhamnosylation, a reaction catalyzed by the glycosyltransferase EarP (Supplementary Fig. 1)[14,21,22]. This evolutionary connection extends beyond overall sequence similarity to the functionally significant β3Ωβ4 loop (Fig. 1A) and the arginine (R33 in *E. coli* EfpL) at its tip (Fig. 1B; Supplementary Fig. 2C)[14]. However, in contrast to these α-rhamnosylated EF-Ps, R33 in EfpL remains unmodified, as confirmed by mass spectrometry (MS) (Supplementary Fig. 3). Additionally, we discovered a strictly conserved proline three amino acids upstream of EfpL_R33—an amino acid typically absent from that position in α-rhamnosylated EF-Ps[14]. Third, EfpLs predominantly co-occur with the EF-P subfamily activated by β-D-lysylation whereas the presence of an EF-P that is α-rhamnosylated typically excludes the existence of the paralogous EfpL (Supplementary Fig. 2B)[23]. Lastly, distinguishing itself from all other EF-Ps, EfpL appears to possess a β3Ωβ4 loop extension (Fig. 1A, B; Supplementary Fig. 2D). However, the exact length of this extension remains ambiguous in the in silico models.

Accordingly, we solved the crystal structure of *E. coli* EfpL (PDB: 8S8U; Supplementary Data 2A) and compared it with other available protein structures of EF-P[10,26]. This confirmed the highly conserved fold of EF-P typed proteins in prokaryotes, both expressed by a structural overlay and respective root-mean-square deviation (r.m.s.d.) values (Supplementary Fig. 4). The EfpL structure reveals a significantly tilted KOW domain relative to the C-terminal di-domain compared to EF-P structures (Fig. 1C), certainly enabled by the flexible hinge region between the independent moieties. However, a separate alignment of KOW and OB di-domains between *E. coli* EfpL and for example, the EF-P structure resolved within the *E. coli* ribosome from Huter et al.[10], reveals low r.m.s.d. values (Supplementary Fig. 4). This suggests the relative domain arrangement is merely a consequence of the unique crystal packing. Altogether, the EfpL high-resolution structure reveals the anticipated fold and features needed for its expected functional role interacting with the ribosome, analogously to EF-P. We then took a closer look at the KOW domain β3Ωβ4 loop relevant for interacting with the tRNA. The structural alignment ultimately revealed a β3Ωβ4 loop elongation by two amino acids for EfpL, different from the canonical seven amino acids in EF-P (Fig. 1A, B; Supplementary Fig. 2D). In this way, EfpL_R33 remains apical similar to canonical EF-Ps. We reasoned that such a loop extension would enable unprecedented contacts with the CCA end of the P-site tRNA without further post-translational modification, which we set out to investigate in detail. Given the overall structural similarity with EF-P, we overlaid the EfpL KOW domain with the cryo-EM structure of EF-P bound to the ribosome[10] to analyze the position and potential contacts of EfpL_R33 with that tRNA trinucleotide. In EF-P, the modified K34 aligns with the trinucleotide backbone without obvious RNA-specific interactions, while the prolonged sidechain allows for a maximum contact site with the RNA (Supplementary Fig. 5). To allow for local adjustments in an otherwise sterically constrained frame of the ribosome, we carried out molecular docking of EfpL and the CCA trinucleotide with a local energy minimization using HADDOCK[27] (Supplementary Fig. 5; Supplementary Data 2B). As shown by the lowest-energy model, the local geometry in principle would allow the unmodified arginine of the EfpL β3Ωβ4 loop to reestablish the interaction with the tRNA trinucleotide. Furthermore, the model suggests EfpL could mediate specific interactions with the RNA as—unlike EF-P_K34—EfpL_R33 was found to stack

https://doi.org/10.1038/s41467-024-54556-9

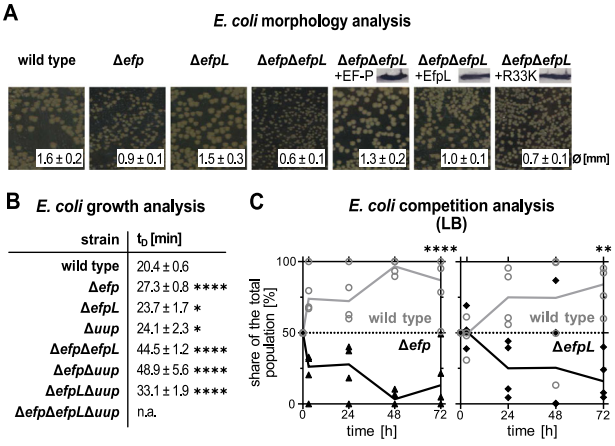



**Fig. 2 | The role of EfpL in bacterial physiology. A** Morphology analysis of *E. coli* BW23113 and isogenic mutant strains lacking *efp* (Δ*efp*), *efpL* (Δ*efpL*), or both genes (Δ*efp*Δ*efpL*). In strains overproducing EF-P (+EF-P), EfpL (+EfpL) and EfpL_R33K (+R33K) protein production was confirmed by immunoblotting utilizing the C-terminally attached $His_6$-tag and Anti-$His_6$ antibodies (α-His). Colony size was quantified by averaging the diameters (mean Ø ± standard deviation (sd)) of 30 colonies on LB agar plates after 18 h of cultivation at 37 °C. Morphology analysis on plates was repeated two more times with similar results. **B** Doubling times (mean $t_D$ ± sd) were calculated from exponentially grown cells in LB ($n \geq 6$, biological replicates). Statistically significant differences to wild-type growth according to two-way ANOVA test (*P* value (*P*) *$P < 0.0332$, ****$P < 0.0001$). **C** Growth analysis of *E. coli* cells in mixture over 72 h. An *E. coli* strain Δ*cadC* without any mutant growth phenotype under the test conditions[92] was used as wild type. *E. coli* BW25113 Δ*cadC* was mixed with either *E. coli* BW25113 Δ*efp* or Δ*efpL* and cultivated for 72 h. The share of the population was detected on LB agar plates ($n = 4$, biological replicates). Statistically significant differences to wild-type growth according to two-way ANOVA test (**$P = 0.003$, ****$P < 0.0001$). **A**–**C** Source data are provided as a Source Data file.

between the two C-bases and make polar interactions with the phosphate-sugar backbone. Hence, based on the docking model we suggest that the prolonged β3Ωβ4 loop and its central tip R33 are capable of compensating for the lack of a modified lysine. It will be interesting to see an atom-resolved proof for this interaction in future high-resolution structures that provide insight beyond the limitations of a docked model.

## *E. coli* EF-P and EfpL have overlapping functions

Based on the structural similarities (Fig. 1C), we assumed that EF-P and EfpL have a similar molecular function. However, there has been no experimental evidence supporting this hypothesis so far. Accordingly, we analyzed growth of *E. coli* wild type and mutants lacking *efp* (Δ*efp*), *efpL* (Δ*efpL*), or both genes (Δ*efp*Δ*efpL*) (Fig. 2A, B). Compared to the strong mutant phenotype in Δ*efp* ($t_d$ ~27 min), we observed a slight but still significant increase in doubling time from ~20 min in the wild type to ~24 min in Δ*efpL*. In line with this observation, a paralleling competition experiment demonstrated that wild-type cells outcompete not only Δ*efp* but also Δ*efpL* within 72 h (Fig. 2C). Further, the mild growth phenotype in Δ*efpL* becomes pronounced in the double deletion mutant Δ*efp*Δ*efpL*, which impairs growth beyond the loss of each single gene ($t_d$ ~45 min). This implies a cooperative role in the translation of polyproline proteins, which is almost masked by EF-P in Δ*efpL* cells. The overproduction of either EF-P or EfpL, but not the substitution of the functional important R33 at the β3Ωβ4 loop tip in the EfpL_R33K variant, completely or partially eliminates the growth defect. However, this effect vanishes when the functional important R33 at the β3Ωβ4 loop tip is substituted in the EfpL_R33K variant, demonstrating the significance of R33 for the molecular function of EfpL. It is also noteworthy that overproduction of EfpL in Δ*efp*Δ*efpL* reduced doubling time below that of Δ*efp* (~27 min) (Supplementary Fig. 6). We hypothesize that ectopic expression partially compensates for the comparatively low-copy number of EfpL per cell (EfpL: ~4500 vs. EF-P: ~40,000 in complex medium[28]) (Supplementary Fig. 7A, B). A similar phenotypic pattern was observed for *efp* and *efpL* deletions when examining the same strains in terms of the CadABC-dependent pH stress response (Supplementary Fig. 8)[29], whose regulator CadC has a polyproline motif[3].

Parallel to our work another player in ribosome rescue at proline-containing arrest motifs was described: an ABCF ATPase termed Uup in *E. coli* and YfmR in *B. subtilis*[30–32]. Notably, while the phenotypic consequences of losing *yfmR* or *efp* hardly affect vegetative growth of *B. subtilis*, their simultaneous deletion dramatically impacts viability and was even suggested to be synthetically lethal. However, there is no ortholog of EfpL in *B. subtilis*. We consequently asked what happens when we delete *uup* in our previously introduced *efp* and *efpL* mutant strains (Fig. 2A). We were able to construct the two double deletions Δ*efp*Δ*uup*, and Δ*efpL*Δ*uup* but we failed to generate a triple deletion Δ*efp*Δ*efpL*Δ*uup* (Fig. 2B). This only succeeded in the presence of a plasmid-encoded, arabinose-inducible copy of *efpL* (Δ*efp*Δ*efpL*Δ*uup* +EfpL). Subsequent growth analyses confirmed that the presence of the inducer allowed *E. coli* to reach cell numbers similar to those of the wild type (and all single and double deletion strains) (Supplementary Fig. 9). By contrast, repression of *efpL* transcription reduced the viable cell counts of *E. coli* Δ*efp*Δ*efpL*Δ*uup* +EfpL by five orders of magnitude. Altogether, this led us to conclude that all three proteins have an overlapping arrest spectrum, and that EfpL becomes essential for ribosome rescue at consecutive prolines when *efp* and *uup* are absent. To confirm this latter hypothesis and pinpoint EfpL's molecular functions in relieving ribosome arrest on diprolines, we used our recently described reporter assay (Fig. 3A)[33]. This assay allows positive correlation of translational pausing strength with bioluminescence. Deletion of either *efp* or *efpL* leads to an increased light emission, and for Δ*efp*Δ*efpL*, we observed a cumulative effect. Again, the phenotype of Δ*efp*Δ*efpL* was trans-complemented by wild-type copies of the respective genes. A parallel quantitative in vitro assay employing NanoLuc® variants with and without polyproline insertion (Fig. 3B) confirmed the results of the previous in vivo experiments with EfpL and its substitution variant EfpL_R33K (Fig. 3C). Unlike in the in vivo analyses with Δ*efpL* and Δ*efp* strains, there are no significant differences in the rescue efficiency between EF-P and EfpL at the tested diproline motif PPN.

## *E. coli* EF-P and EfpL alleviate ribosome stalling at distinct XP(P)X motifs with differences in rescue efficiency

To elucidate the EfpL arrest motif spectrum, a ribosome profiling analysis (RiboSeq) was conducted. Here an *E. coli* wild type was compared with Δ*efp* and Δ*efpL* strains. Importantly, we also included Δ*efp* cells in which EfpL was overproduced. As indicated by our previous analyses (Figs. 2 and 3) this compensates for the relatively low natural copy number of the factor and might uncover motifs that are otherwise masked by the presence of EF-P. We used PausePred[34] to predict pauses in protein translation in the respective strains. Subsequently, we calculated the frequencies of amino acid triplet residues occurring at the sites of predicted pauses (Fig. 4A; Supplementary Data 3A). In line with the molecular function of EF-P, diproline motifs were heavily enriched at pause sites in Δ*efp*[3,5,6]. As already suspected by the mild mutant phenotype of the *efpL* deletion (Figs. 2 and 3) we did not see a significant difference between Δ*efpL* and the wild type. However, in stark contrast, overproduction of EfpL alleviated ribosome stalling at many but not all arrest motifs identified in Δ*efp*. Further, in line with EF-P function, our comparative metagene analysis revealed no noticeable effects on initiation or termination for EfpL (Supplementary Figs. 10 and 11)[35]. Together this corroborates the idea that EfpL has evolved to assist EF-P in translational rescue. Our analysis further

https://doi.org/10.1038/s41467-024-54556-9

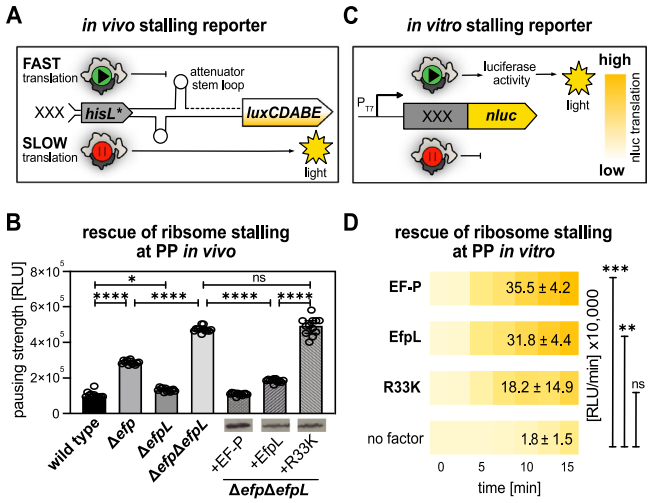



**Fig. 3 | The function of EfpL in alleviation of ribosome stalling. A** Scheme of the in vivo stalling reporter system[33]. The system operates on the histidine biosynthesis operon of *E. coli*. In its natural form, the histidine biosynthesis gene cluster is controlled by the His-leader peptide (HisL), which comprises seven consecutive histidines. In our setup, the original histidine residues (His1 through His4) were replaced by artificial sequence motifs (XXX). Non-stalling sequences promote the formation of an attenuator stem loop (upper part) that impedes transcription of the downstream genes, thus ultimately preventing light emission. Conversely, in the presence of an arrest motif, ribosomes pause and hence an alternative stem loop is formed that does not attenuate transcription of the *luxCDABE* genes of *Photorhabdus luminescens*. **B** In vivo comparison of pausing at PPN in *E. coli* (for strain labeling and immunoblotting details see (**A**)). Pausing strength is given in relative light units (RLU) ($n = 12$, biological replicates, mean with sd indicated as error bars). Statistically significant differences according to an ordinary one-way ANOVA (*$P < 0.0332$, ****$P < 0.0001$, ns not significant). **C** Scheme of the in vitro cell-free stalling reporter assay. The system is based on *nanoluc* luciferase (nluc®) which is preceded by an artificial sequence motif (XXX). DNA is transcribed from a T7 promoter ($P_{T7}$) using purified T7 polymerase (NEB). Pausing strength is proportional to light emission. **D** In vitro transcription and translation of the nLuc® variant nLuc_PPN. The absence (no factor) or presence of the respective translation elongation factors of *E. coli* (EF-P, EfpL, EfpL_R33K) is shown. Translational output was determined by measuring bioluminescence in a time course of 15 min and endpoints are given in relative light units (RLU/min±sd) ($n \geq 3$, technical replicates). Statistically significant differences to the control (no factor) according to ordinary one-way ANOVA (**$P = 0.0015$, ***$P = 0.0005$, ns not significant). **B**, **D** Source data are provided as a Source Data file.

revealed that among the top 29 stalling motifs are not only XPPX but also many XPX motifs and one motif completely lacking a proline (Fig. 4A; Supplementary Fig. 12). The RiboSeq findings were confirmed with our in vivo luminescence reporter (Fig. 4B; Supplementary Fig. 13) by testing 12 different arrest motifs as well as in vitro by quantifying production of two NanoLuc® Luciferase (nLuc) variants comprising IPW and PAP (Fig. 4C; Supplementary Fig. 14). Together, our data demonstrate that while a P-site proline is almost always a prerequisite for ribosome rescue by EF-P/EfpL, in rare cases motifs lacking proline can also be targeted.

An arrest spectrum extension beyond diprolines has only been reported for IF5A thus far[35,36] although there are weak indications in the literature that EF-P[4] and similarly EfpL might assist in synthesis of the XPX containing sequence of the leader peptide MgtL[37–40], which we able to substantiate (Supplementary Fig. 15). To further explore EfpL's contribution to gene-specific translational rescue, we focused on the top 29 motifs as done before for eIF5A[35] and looked at the frequency of ribosome occupancy before and after the pause sequence. The ratio between these values gives an asymmetry score (AS) and provides a good measure for stalling strength[6]. EF-P and EfpL dependency was determined by comparing with the AS from the wild type. We were thus able to recapitulate the data from previous RiboSeq analyses for the Δ*efp* samples (Supplementary Data 3B, C). Moreover, with this approach, we were able to find EfpL targets not only in the Δ*efp* +EfpL sample but also in Δ*efpL*. In line with our phenotypic analyses (Fig. 2; Supplementary Figs. 8 and 15), most of these proteins are also targeted by EF-P (Fig. 4D; Supplementary Data 3). While in the majority of cases, the rescue efficiency was better with EF-P, we found some proteins where EfpL seems to be superior. We even identified a few candidates that were only dependent on EfpL. The proteins targeted by EfpL are frequently involved in amino acid metabolism and transport (Fig. 4D; Supplementary Data 3D). This provides a potential explanation for the growth phenotype we observed in Lysogeny broth (LB), where amino acids constitute the major source of nutrients (Fig. 2A–C). Notably, when we swapped to glucose as dominant C-source and compared growth in LB and LB supplemented with 20 mM Glucose, indeed the cumulative growth defect of Δ*efp*Δ*efpL* was gone (Supplementary Fig. 6). Moreover, while wild-type *E. coli* outcompetes Δ*efp* under these conditions, the proportion of the Δ*efpL* population remained constant within 72 h (Fig. 4E). Thus, our data support the assumption that EF-P functions as a housekeeping factor whereas EfpL exerts its role depending on the available nutrients. We hypothesize, that the structural differences between the two factors lead to different efficiencies in resolving ribosome stalling at specific motifs (Supplementary Data 3)[30–32,41]. A sequence logo based on translations modulated exclusively by EfpL (Supplementary Fig. 12C) shows a clear overrepresentation of DPA, PPV and DPN (Supplementary Data 3C) and, presumably depending on the amino acid context of the arrest motif[42,43], this factor will become superior in resolving the stall.

## A guanosine in the first position of the E-site codon as recognition element for EF-P and EfpL

The chemical nature of the X residues in XP(P)X in the top 29 stalling motifs (Fig. 4A; Supplementary Fig. 12A, B) is highly diverse and does not provide a cohesive rationale for the arrest motif spectrum: besides the negatively charged residues aspartate and glutamate, we found especially the hydrophobic amino acids isoleucine and valine as well as small ones, like glycine for X at the XP(P) position (Supplementary Fig. 12C). Consequently, we extended our view to the codon level. EF-P and accordingly EfpL can interact with the E-site codon utilizing the first loop in the C-terminal OB-domain (d3 loop I)[10,11]. We did not see any preference for a specific base in the wobble position. By contrast, we revealed a strong bias for guanosine in the first position of the E-site codon in the sequence logos (Supplementary Fig. 12D) of EF-P- and EfpL-targeted XPP motifs, where X ≠ P. Notably, we observed no clear trend when we looked at the X in (P)PX in motifs (Supplementary Fig. 12C). When bound to the ribosome, EF-P establishes contacts with the first and second position of the E-site codon through d3 loop I residues G144–G148, with sidechain-to-base specific contacts involving D145 and T146[10] (Supplementary Fig. 16). However, in the available high-resolution structure, ribosomes are arrested at a triproline motif and thus, the E-site codon (CCN) does not contain a guanosine. Referring to our observation, we replaced the cytosine in the structure by guanosine in silico, followed by an additional docking and energy minimization of the loop-RNA interface (Supplementary Fig. 16A, B). The resulting complex suggests additional contacts that in principle could appear between guanosine and EF-P as compared to cytosine (Supplementary Fig. 16C). Despite potential biases of the docking procedure, a preference for G would be supported by an extended interface with sequential contacts up to residue G151. As suggested by the model, this could per se involve the entire d3 loop I. Based on the motif analysis, we thus conclude that especially guanosine in the first position of the E-site codon promotes EF-P and EfpL binding to the ribosome, which is additionally supported by the in silico comparison.

https://doi.org/10.1038/s41467-024-54556-9

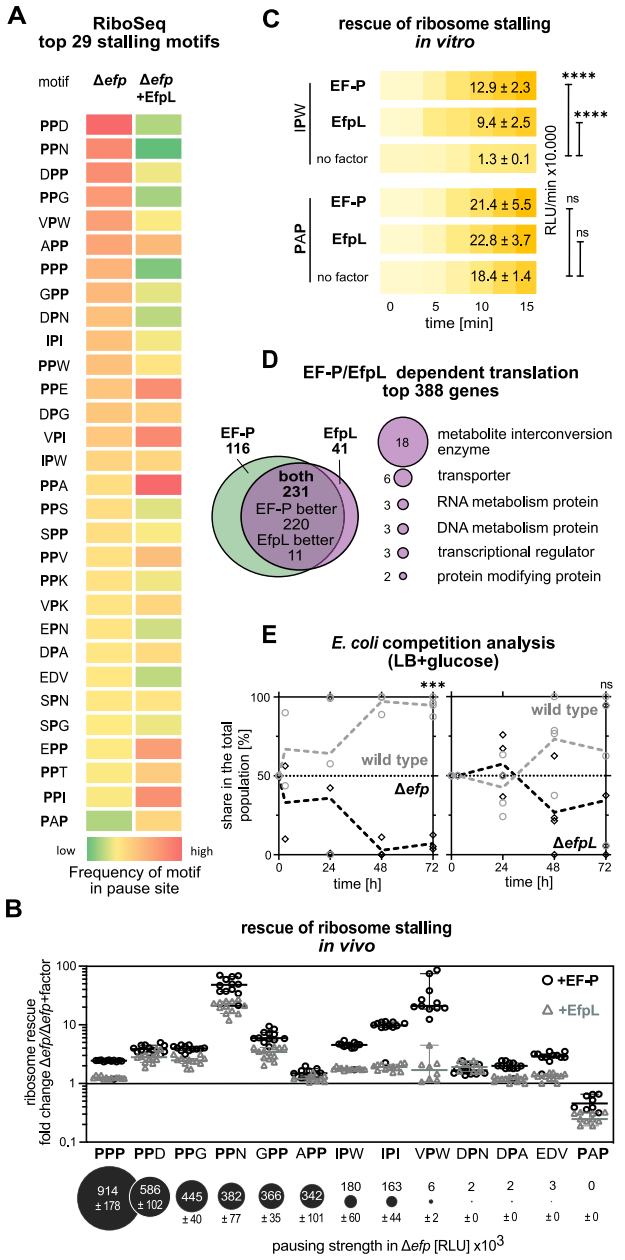



**Fig. 4 | The target spectrum of EF-P and EfpL. A** Color code of the heat map corresponds to frequency of the motif to occur in pause site in the ribosome profiling analysis predicted with PausePred[34] (From green to red = from low to high). First column: Top 29 motifs whose translation is dependent on EF-P and the control motif PAP in the ribosome profiling analysis comparing *E. coli* BW25113 with the *efp* deletion mutant (Δ*efp*). Second column: Comparison of profiling data of Δ*efp* and Δ*efp* cells overexpressing *efpL* (Δ*efp* +EfpL) at these motifs. **B** In vivo comparison of rescue efficiency of a set of stalling motifs and the control PAP. Given is the quotient of relative light units measured in Δ*efp* and corresponding trans-complementations by EF-P (+EF-P) and EfpL (+EfpL). Motifs are sorted according to pausing strength determined with our stalling reporter ($n = 12$, biological replicates, mean with sd indicated as error bars). **C** In vitro transcription and translation of *nLuc*® variants nLuc_3xRIPW (IPW) and nLuc_3xRPAP (PAP). The absence (no factor) or presence of the respective translation elongation factors of *E. coli* (EF-P/EfpL) is shown. Translational output was determined by measuring bioluminescence in a time course of 15 min and is given in relative light units measured at the end of the reaction (RLU/min ± sd) ($n \geq 3$, technical replicates). Statistically significant differences to control (no factor) according to ordinary one-way ANOVA (****$P < 0.0001$, ns not significant). **D** Left part: Venn diagram of top 388 genes, whose translation depends on EF-P and EfpL. Dependency was determined by comparing asymmetry scores from genes encompassing top 29 stalling motifs listed in (**A**). Right part: Enriched protein classes to which EfpL-dependent genes belong[126]. **E** Growth analysis of *E. coli* cells in mixture over 72 h in LB with 40 mM glucose. A Δ*cadC* strain without growth phenotype[92] was used as the wild type. *E. coli* BW25113 Δ*cadC* was mixed with either *E. coli* BW25113 Δ*efp* or Δ*efpL*. The share of the population was detected on LB agar plates ($n = 4$, biological replicates). Statistically significant differences to wild-type growth according to two-way ANOVA (***$P = 0.0006$, ns not significant). **A**–**C**, **E** Source data are provided as a Source Data file.

## EF-P and EfpL can induce translational pauses

We found the unique recognition elements of an EF-P/EfpL-dependent arrest motif to be the P-site tRNA$^{Pro}$ and the E-site codon, in agreement with past studies[9,10,44]. We therefore wondered whether XP—regardless of being part of a stalling motif or not—promotes binding of EF-P and similarly EfpL to the ribosome. If so, such "off-binding" might induce pausing at non-stalling motifs instead of alleviating it. Although weak, we indeed saw that loss of *efp* increases pausing with our PAP non-stalling control (Fig. 4B), which comprises two XPX motifs namely RPA and APH. Conversely, *efp* and *efpL* overexpression showed the opposite effect. Thus, our study provides evidence that the translation factors EF-P and EfpL can induce pausing, presumably by blocking tRNA translocation to the E-site. Our hypothesis was confirmed by showing that one can also induce pausing at a clean APH motif (Supplementary Fig. 17A). Either such an apparently deleterious effect is accepted, as the positive influence on arrest motifs outweighs the negative one, or translational pauses at XP(P)X might also have positive effects on, for example, buying time for domain folding or membrane insertion[45]. We were further curious whether we see codon-specific effects and tested the non-stalling motif RPH, in which the E-site codon starts with C (R is encoded by CGC) (Supplementary Fig. 17A). Congruent with our previous findings EF-P could no longer increase pausing strength and with EfpL the effect was less pronounced, while an R33K substitution had no inhibitory effect. In summary, our findings indicate that EF-P (and EfpL) may be able to bind to the ribosome whenever a proline is translated, with binding being further promoted by the E-site codon. This idea is in line with earlier work from Mohapatra et al.[46]. The authors reported that EF-P binds to ribosomes during many or most elongation cycles. Our data may now provide a rationale for this (at the time) unexpectedly high binding frequency, which by far exceeds the number of XPPX arrest motifs. In addition to these weak pauses induced at XPX, we observed in our RiboSeq data that EF-P might also bind non-productively at certain motifs as evidenced by asymmetry scores that are higher in Δ*efp* samples than in the wild type (Supplementary Data 3B, C). While such events are predominantly weak and only rarely observed in our Δ*efp* RiboSeq data, their frequency and strength increased when we overproduced EfpL in the Δ*efp* +EfpL sample (Supplementary Fig. 17; Supplementary Data 3B, C). This supports the idea that the structural differences of the two factors differentially align and stabilize the P-site tRNA$^{Pro}$. We thus reasoned that the presence of a constitutive EF-P and a more specialized EfpL, would provide the cell with a lever to intentionally delay or accelerate translation gene specifically. However, this would require regulation. Following indications from a global analysis, *efpL* expression is regulated by carbon catabolite repression (Supplementary Fig. 18A)[47]. It was predicted that $P_{efpL}$ is a class II cAMP response protein (CRP)-dependent promoter. However, the putative CRP binding site deviates significantly from the consensus motif of the regulator. Consequently, we reinvestigated the hypothesized regulation analogous to previous studies[48] but did not observe any measurable effect (Supplementary Fig. 18B, C). Subsequently, we extended our dataset to include conditions such as nutrient availability, acetylphosphate levels, heat, cold, acidic and alkaline pH, as well as high and low osmolarity (Supplementary Fig. 18D, E). Under all tested conditions, the promoter activities of $P_{efp}$ and $P_{efpL}$ maintained a constant

ratio of about 10:1. Our findings are also consistent with "The quantitative and condition-dependent *E. coli* proteome"[28], which shows that the protein copy number patterns of EF-P and EfpL perfectly match and follow other ribosomal factors (Supplementary Fig. 18F). Accordingly, post-transcriptional control of the respective *efp* and *efpL* mRNAs is attributed to maintaining the balance in protein copy number between the two proteins[49–52].

### *E. coli* EfpL is deactivated by acylation

As an alternative to protein copy number control, post-translational modifications provide a means to adjust EF-P activity to cellular needs. Since we were able to demonstrate that—unlike many other EF-P subtypes—the EfpL β3Ωβ4 loop tip is unmodified, we extended our view to the entire protein sequence. The idea arose as the activity and subcellular localization of the eukaryotic EF-P ortholog eIF5A is regulated by phosphorylation and acetylation, respectively[53,54]. A literature search revealed that *E. coli* EfpL is acylated at four different lysines (K23, K40, K51, and K57) in the KOW domain (Fig. 5A)[55–58]. Notably, a sequence comparison with EF-P shows that a lysine is found only in the position equivalent to K57, and there is no evidence of modification[55–58]. Possible acylations of EfpL encompass not only acetylation but also malonylation and succinylation (Fig. 5A). As a consequence, the positive charge of lysine can either be neutralized or even turned negative. To investigate the impact of acylation on EfpL we generated protein variants in which we introduced $N_\varepsilon$-acetyllysine (AcK) by amber suppression[59] at each individual position where acylation was previously reported (EfpL_K23AcK, EfpL_K40AcK, EfpL_K51AcK, and EfpL_K57AcK). Testing of purified protein variants in the established in vitro assay revealed that K51 acetylation impairs EfpL's function, significantly (Fig. 5B; Supplementary Fig. 19). We argue that charge alterations at these lysines, as well as subsequent steric constraints, will impair ribosomal interactions. To this end, we modeled the EfpL KOW domain to the ribosome by structural alignment with EF-P in order to investigate the effects of acetylation visualized by respective in silico modifications (Supplementary Fig. 20). In line with the rescue experiments, the in silico data show that compared to all other modification sites K51 is most sterically impaired by acetylation. Longer sidechain modifications at K51 such as succinylation will most likely prevent EfpL from binding to the ribosome. To confirm the in silico and in vitro data on EfpL inactivation by acylation, we sought to validate these findings in vivo. Acylation is predominantly a non-enzymatic modification influenced by the cell's metabolic state (Supplementary Fig. 21A, B), specifically by internal levels of acetyl-phosphate[55,56]. Consequently, different growth conditions can either promote or inhibit acylation levels. For instance, glucose or acetate utilization increases acetylation of EfpL due to higher levels of the acetyl group donor, acetyl phosphate, which is particularly significant for K51[55,56,60]. We used *E. coli* cells that dependent on EfpL as the sole ribosome rescue system for stalls at XPX and XPPX (Δ*efp*Δ*uup*), and tested growth in acetate medium, expressing EfpL K51 substitutions (Fig. 5C). Arginine (K51R) was used to mimic the non-acetylated state, glutamine (K51Q) served as an acetyllysine mimic, and glutamate (K51E) introduced a negative charge similar to malonylation and succinylation. All variants were expressed from a low-copy number plasmid[61] under the control of the native *efpL* promoter ($P_{efpL}$). Under these conditions, only the K51R culture grew comparable to Δ*efp*Δ*uup* cells ectopically expressing *efp*. By contrast, the culture with the K51Q variant turned only slightly turbid and we did not observe an increase in culture density for the K51E variant nor with wild-type EfpL. This, in turn, confirmed our previous assumptions, demonstrating that both chain length and charge at EfpL position 51 are crucial for protein activity. In conclusion, our combined in vivo, in vitro, and in silico data clearly demonstrate that EfpL is inactivated by acylation. It has recently been shown that acetylation of ribosomal proteins in general inhibits translation and increases the proportions of dissociated 30S and 50S ribosomes[62]. In addition to this scheme, we have now uncovered, that in *E. coli* the activity of EfpL is regulated by acylation. In this way, the protein acts as a sensor for the metabolic state to regulate translation of specific XP(P)X proteins.

### The presence of EfpL is associated with faster bacterial growth

Paralogous proteins evolve to diversify functionality and enable species-specific regulation[63]. In this regard, we found that in enterobacteria, the four acylation sites of EfpL in *E. coli* remain largely invariable, whereas in others, such as *Vibrio* species, they show less conservation (Fig. 5D; Supplementary Fig. 22). Most importantly, lysine in position 51 is an arginine in the EfpLs of e.g., *Vibrio cholerae, Vibrio natriegens* and *Vibrio campbellii*. Moreover, we found that expression levels of *efpL V. campbellii* ($efpL_{Vca}$) are much higher than in *E. coli* and equal those of $efp_{Vca}$, together suggesting a broader role for EfpL in this organism (Supplementary Fig. 23). We compared the rescue efficiency of $EfpL_{Vca}$ with those of selected *Enterobacteriaceae* (Fig. 5E) and found that overproduction of $EfpL_{Vca}$ was superior over all tested enterobacterial EfpLs. In fact, the protein could most effectively counteract the translational arrest at PPN not only in vitro but also in vivo (Fig. 5E, F; Supplementary Fig. 24). Next, we investigated the effect of an $efpL_{Vca}$ deletion. Similar to *E. coli* we did not find any growth phenotype. However, in stark contrast, a deletion of $efp_{Vca}$ also had no consequences for growth speed. Only the simultaneous deletion of both genes (Δ*efp*Δ*efpL*) diminished growth in *V. campbelli*, suggesting that $EfpL_{Vca}$ and $EF\text{-}P_{Vca}$ can fully compensate for the absence of the other. To exclude a species-specific behavior, we further included *V. natriegens*, the world record holder in growth speed (doubling time is less than 10 min under optimal conditions) (Fig. 5G)[64]. Similar to *V. campbellii* both proteins seem to be of equal importance. Therefore, we conclude, that the role of EfpL in ribosomal rescue of XP(P)X is more general in *Vibrio* species compared to Enterobacteria.

We were ultimately curious, whether there might be a universal benefit for bacteria in encoding EfpL. To this end, we estimated doubling times of a reference dataset of γ-proteobacteria using a codon usage bias-based method (Supplementary Fig. 25)[65]. Then we categorized them according to presence or absence of an EfpL paralog. To minimize differences resulting from phylogenetic diversity we focused specifically on γ-proteobacteria encoding an EF-P that is activated by EpmA (Fig. 5H). Notably, bacteria with EfpL are predicted to grow faster than those lacking it. Thus, we conclude that the concomitant presence of EF-P and EfpL might be an evolutionary driver for faster growth. We speculate that microorganisms with both proteins benefit from their unique capabilities to interact with the P-site $tRNA^{Pro}$, which in turn helps to increase overall translation efficiency.

## Discussion

Proline is the only secondary amino acid in the genetic code. The pyrrolidine ring can equip proteins with unique properties[66] and the polyproline helix is just one expression for the structural possibilities[67]. However, all this comes at a price. The rigidity of proline decelerates the peptidyl transfer reaction with $tRNA^{Pro}$. Not only is it a poor A-site peptidyl acceptor, but also proline is a poor peptidyl donor for the P-site[68,69]. Nevertheless, arrest-inducing polyprolines occur frequently in pro- and eukaryotic genomes[45,70]. This, in turn, shows that the benefits of such sequence motifs outweigh the corresponding drawbacks and explain why nature has evolved the universally conserved EF-P to assist in translation elongation at XP(P)X[33]. To promote binding to the polyproline stalled ribosome EF-P specifically interacts with the D-loop of the P-site $tRNA^{Pro}$[9], the L1 stalk, and the 30S subunit[11] and the mRNA[10], with the latter being the only variable in this equation. Accordingly, in the ideal case, the EF-P retention time on the ribosome could be modulated according to the motif's arrest strength. Indeed, the dissociation rate constant of EF-P from the ribosome differs

https://doi.org/10.1038/s41467-024-54556-9

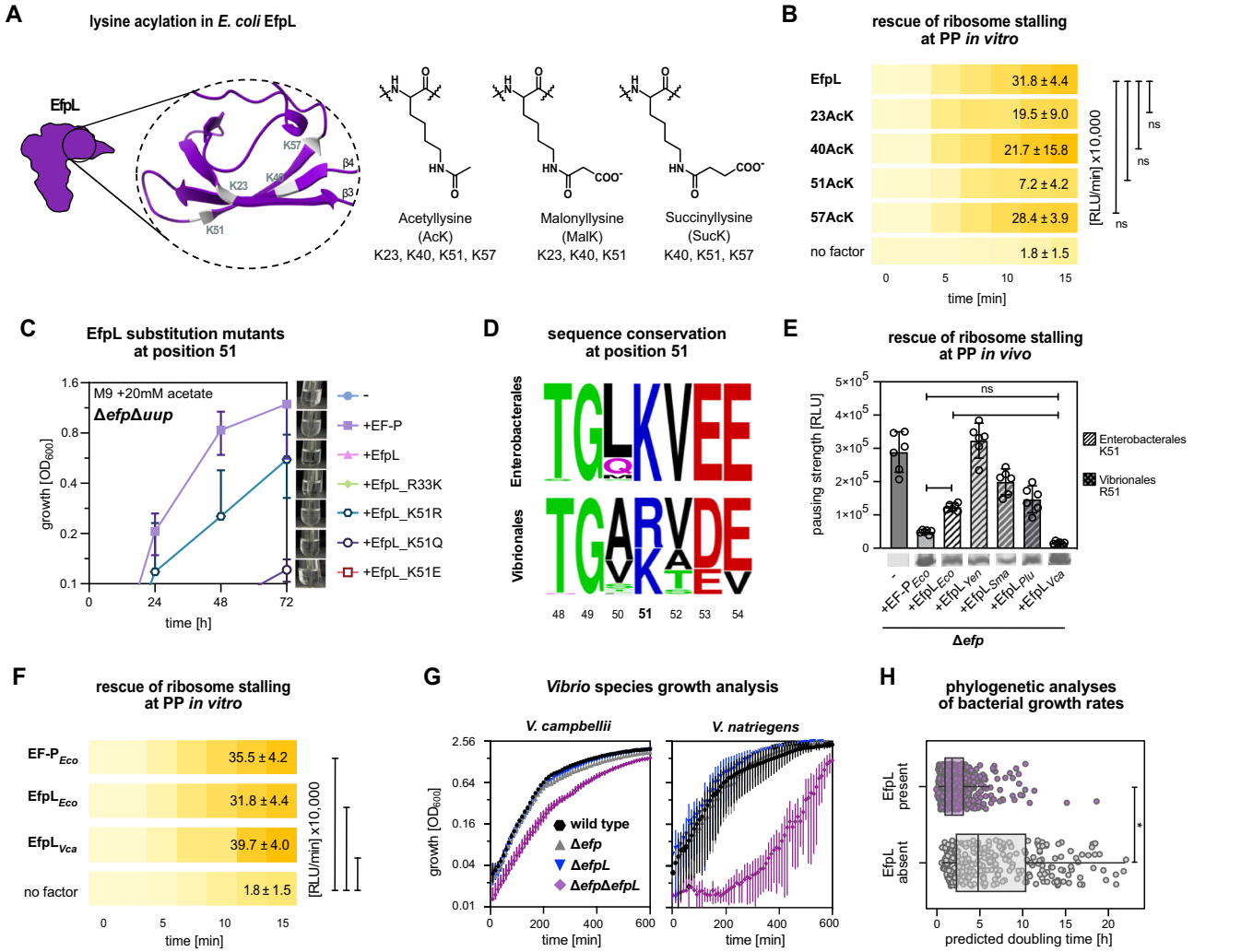



**Fig. 5 | EfpL acylation and its regulation in distinct bacteria. A** EfpL acylations according to refs. 55–58. Acylated lysines are depicted as part of a polypeptide, represented by the wavy line. **B** In vitro transcription and translation of the *nLuc*® variant nLuc_PPN. The absence (no factor) or presence of *E. coli* EF-P or EfpL and substitution variants EfpL_K23AcK, EfpL_K40AcK, EfpL_51AcK, EfpL_K57AcK is shown. Translational output was determined by measuring bioluminescence in a 15 min time course and is given in relative light units (RLU/min±sd) ($n \geq 3$, technical replicates). Statistically significant differences according to ordinary one-way ANOVA (*$P = 0.0364$, ns not significant). **C** Growth analysis of *E. coli* BW25113 Δ*efp*Δ*uup* trans-complemented with *efp* (+EF-P), *efpL* (+EfpL) or *efpL* substitution mutants (+EfpL_R33K/_K51R/_K51Q/_K51E) in M9-medium with 20 mM acetate as sole carbon source. Images of growth media were taken after 48 h ($n = 3$, biological replicates, mean with sd indicated as error bars). **D** Sequence logos[104] of position 51 ± 3 amino acids in EfpL in Enterobacterales and Vibrionales. **E** In vivo comparison of pausing at PPN in *E. coli* Δ*efp* cells and trans-complementations with EF-P/EfpL of *E. coli* (+EF-$P_{Eco}$/+$EfpL_{Eco}$), *Yersinia enterocolitica* (+$EfpL_{Yen}$), *Serratia marcescens* (+$EfpL_{Sma}$), *P. luminescens* (+$EfpL_{Plu}$), *Vibrio campbellii* (+$EfpL_{Vca}$). Pausing strength is given in relative light units (RLU) ($n = 6$, biological replicates, mean with sd indicated as error bars). Statistically significant differences according to one-way ANOVA (*$P = 0.0152$, ***$P = 0.0002$, ns not significant). **F** In vitro analysis as in (**B**). The absence (no factor) or presence of elongation factors of *E. coli* (EF-$P_{Eco}$/$EfpL_{Eco}$) and *V. campbellii* ($EfpL_{Vca}$) is shown. ($n \geq 3$, technical replicates) (statistics as in (**B**), ****$P = 0.0001$, **$P = 0.0015$, ***$P = 0.0005$). **G** Growth analysis of *V. campbellii* (in LM) and *Vibrio natriegens* (in LB) with corresponding deletions of *efp* (Δ*efp*), *efpL* (Δ*efpL*), or both genes (Δ*efp*Δ*efpL*) ($n = 11$; biological replicates, mean with sd indicated as error bars). **H** Phylogenetic analysis of predicted γ-proteobacterial growth rates comparing absence or presence of EfpL. Doubling times were predicted using codon usage bias in ribosomal proteins. ($n = 786$ genomes, median with top and bottom boundaries representing 1st and 3rd quartiles and whiskers indicating 1.5 times inter-quartile range). Statistically significant difference according to phylogenetic ANOVA ($P = 0.029$, $P$ value based on 1000 permutations). **B**, **C**, **E**–**G** Source data are provided as Source Data file.

depending on the E-site codon[44]. Our data support the hypothesis that amino acids encoded by a codon beginning with a guanosine induce a particularly strong translational arrest in XP(P) motifs (Supplementary Fig. 12). As EF-P is an ancient translation factor being already present before phylogenetic separation of bacteria and eukaryotes/archaea[71], we wondered whether there is a connection to the evolution of the genetic code. Remarkably, all six amino acids encoded by GNN (Gly, Ala, Asp, Glu, Val, Leu) are included among the standard amino acids that can be produced under emulated primordial conditions[72]. One might therefore speculate that in the early phase of life, EF-P/IF5A were essential to assist in nearly every peptide bond formation with proline in the P-site and thus reading the E-site codon by a second OB-domain was especially beneficial.

The importance to alleviate ribosome stalling at prolines is further underlined by the existence of additional rescue systems namely YebC1 and YebC2[73] (orthologous to the mitochondrial TACO1[74]) and the ATP-Binding Cassette family-F (ABCF) protein Uup in *E. coli* and its ortholog YfmR in *B. subtilis*[30–32,41,75]. In interplay with EF-P, Uup/YfmR, YebC, and EfpL can facilitate translation of XP(P)X-containing proteins. The different modes of action and structural characteristics of the four factors enabled specialization. In case of EfpL, the protein is superior in ribosome rescue at specific genes (Fig. 4D; Supplementary Data 3).

This, in turn, might be an evolutionary driving force for translational speed and hence higher growth rates as indicated by our phylogenetic analysis (Fig. 5H). Alternatively, an EF-P paralog opens additional regulatory possibilities. In contrast to EfpL-encoding bacteria, some lactobacilli, for instance, have two copies of *efp* in the genome (Supplementary Data 1)[14]. One might speculate that here one *efp* is constitutively expressed and the second copy is transcriptionally regulated according to the translational needs. Although relying only on one EF-P, such regulation was reported for Actinobacteria, in which polyproline-containing proteins are concentrated in the accessory proteome[76]. Here EF-P accumulates during early stationary phase and might boost secondary metabolite production as evidenced for *Streptomyces coelicolor*. By contrast, for *E. coli* EfpL there is no evidence for such copy number control, as it simply mirrors the expression pattern of other ribosomal proteins[28]. Instead, the protein seems to fulfill a dual role in this organism. On the one hand, it is essential for full growth speed (Figs. 2A and 5H). On the other hand, it acts as sensor of the metabolic state (Supplementary Fig. 6). The combination of multiple sites of acylation[55–58] and the chemical diversity of this modification type[60] lead to a highly heterogenic EfpL population, which could fine-tune translation in each cell differently. We speculate that regulation of translation by acylation[62] in general and of EfpL in particular adds to phenotypic heterogeneity and thus might contribute to survival of a population under changing environmental conditions[77]. Such a scenario is particularly important for bacteria that colonize very different ecological niches, such as many enterobacteria including *E. coli* do. Depending on whether they are found e.g., in the soil/water or in the large intestine, the nutrient sources they rely on change. Therefore, it is plausible to assume that fine-tuning metabolic responses by acylating and deacylating, EfpL gives enterobacteria an advantage to thrive in the gastrointestinal tract.

Compared to the eukaryotic and archaeal IF5A, EF-P diversity is much greater[8,78]. Especially the functionally significant β3Ωβ4 has undergone significant changes. Starting with the catalytic residue at the loop tip, which is not restricted to lysine as for eukaryotes/archaea. Instead, one also finds asparagine, glutamine, methionine, serine and glycine, besides arginine[23]. These changes extend to the overall sequence composition of β3Ωβ4 to either increase stiffness[79] or, in the case of EfpL, to prolong the loop, as shown in this study by an EfpL high-resolution structure. The latter two strategies functionalize the protein without modification. Notably, the EfpL subgroup is phylogenetically linked most closely to the EF-P branch being activated by α-rhamnosylation[14,21]. This raises the question about the evolutionary origin of EfpL. Starting from a lysine-type EF-P[71], we speculate that upon gene duplication and sequence diversification, an early EfpL arose, and cells benefitted from improved functionality in a subset of XP(P)X arrest peptides. Further evolutionary events could include the shrinkage of the loop back to the canonical seven amino acids and eventually the phylogenetic recruitment of EarP. Such phylogenetic order is supported by an invariant proline upstream of the catalytically active loop tip residue which is found in EfpLs and lysine-type EF-Ps, but is absent in EarP-type EF-Ps (Fig. 1).

Lastly, EF-P diversity holds also potential for synthetic biology applications. Reportedly, EF-P can boost peptide bond formation with many non-canonical amino acids (ncAA)[80–83]. This includes not only proline derivatives but also D- and β-amino acids. However, in all studies, *E. coli* EF-P was used. Given the structural differences between EfpL and EF-P and the resulting differences in the rescue spectrum, we speculate that use of EfpL might be especially beneficial for genetic code expansion for certain ncAA. Collectively, our structural and functional characterization of the EfpL subfamily not only underscores the importance of ribosome rescue at XP(P)X motifs but also adds another weapon to the bacterial arsenal for coping with this type of translational stress. We further illustrate how different bacteria utilize this weapon to gain evolutionary advantages and give an outlook on how EfpL can potentially be used as a molecular tool.

## Methods

### Plasmid and strain construction

All strains, plasmids, and oligonucleotides used in this study are listed and described in Supplementary data files (Supplementary Data. 4), respectively. Kits and enzymes were used according to manufacturer's instructions. Plasmid DNA was isolated using the Zyppy® Plasmid Miniprep Kit from Zymo Research. DNA fragments were purified from agarose gels using the Zymoclean® Gel DNA Recovery Kit or from PCR reactions using the DNA Clean & Concentrator®-5 DNA kit from Zymo Research. All restriction enzymes, DNA modifying enzymes, and the Q5® high fidelity DNA polymerase for PCR amplification were purchased from New England BioLabs.

Plasmids for expression of C-terminally $His_6$-tagged *efp* and *efpL* genes under the control of an inducible promoter were generated by amplification of the corresponding genes from genomic DNA using specific primers and subsequent cut/ligation into the pBAD33 vector[84]. Plasmids for expression of SUMO-tagged *efpL* genes were generated with the Champion™ pET-SUMO Expression System from Invitrogen™ according to manufacturer's instructions. HisL*_lux reporter strains were generated according to Krafczyk et al.[33]. Briefly, an upstream fragment (containing desired mutations) of the *hisLGDCBHAFI* operon was amplified via PCR using the respective primer pairs. After purification, these fragments were isolated from an agarose gel, digested with specific restriction enzymes, and then ligated into the suicide vector. The resulting plasmids were introduced into *E. coli* BW25113 by conjugative mating with *E. coli* WM3064 as the donor strain on LB medium supplemented with meso-α,ε-Diaminopimelic acid (DAP). Single-crossover integration mutants were selected on LB plates containing kanamycin and lacking DAP. Finally, the resulting mutations were confirmed by Sanger sequencing.

Deletions and chromosomal integrations of $His_6$-tagged encoding genes using RecA-mediated homologous recombination with pNPTS138-R6KT of *efp* and *efpL* were made according to Lassak et al.[85,86]. To achieve this, ~500 bp long upstream and downstream fragments of the desired gene region were amplified via PCR using the respective primer pairs. After purification, these fragments were combined through overlap PCR. The final product was isolated from an agarose gel, digested with specific restriction enzymes, and then ligated into the suicide vector. The resulting plasmids were introduced into *E. coli* BW25113 or *Vibrio* species by conjugative mating with *E. coli* WM3064 as the donor strain on LB medium supplemented with DAP. Single-crossover integration mutants were selected on LB plates containing kanamycin and lacking DAP. Single colonies were cultured overnight in LB without antibiotics and subsequently plated onto LB containing 10% (wt/vol) sucrose to select for plasmid excision. Kanamycin-sensitive colonies were screened for targeted deletions through sequencing using primers flanking the site of mutation.

Genetic manipulations via Red®/ET® recombination were done with the Quick & Easy *E. coli* Gene Deletion Kit (Gene Bridges, Heidelberg, Germany). Reporter plasmid constructions with pBBR1-MCS5-TT-RBS-lux were made according to Gödeke et al.[87]. Briefly, the upstream region of genes of interest was cloned 5′ to the *luxCDABE* operon.

### Growth conditions

*E. coli* cells were routinely grown in Miller-modified LB[88,89], super optimal broth (SOB)[90] or M9 minimal medium supplemented with 20 mM of Glucose[86] at 37 °C aerobically under agitation unless indicated otherwise. *V. campbellii* cells were grown in Luria marine (LM) medium (LB supplemented with an additional 10 g/l NaCl)[91] at 30 °C aerobically. *V. natriegens* cells were grown in LB at 30 °C aerobically. Growth was recorded by measuring the optical density at a wavelength

of 600 nm ($OD_{600}$). When required 1.5% (w/v) agar was used to solidify media. Alternative carbon sources and media supplements were added and are indicated. If needed, antibiotics were added at the following concentrations: 100 μg/ml carbenicillin sodium salt, 50 μg/ml kanamycin sulfate, 20 μg/ml gentamycin sulfate, 30 μg/ml chloramphenicol. Plasmids carrying pBAD[84] or Lac promoter were induced with L(+)-arabinose at a final concentration of 0.2% (w/v) or Isopropyl-β-D-thiogalactopyranosid (IPTG) at a final concentration of 1 mM, respectively.

### In vivo promotor activity assay

*E. coli* cells harboring the plasmids pBBR1-MCS5-$P_{efp}$-*luxCDABE* or pBBR1-MCS5-$P_{efpL}$-*luxCDABE* versions were inoculated in LB with appropriate antibiotics. The next day, 96-well microtiter plates with fresh LB with supplements or M9 minimal media with mentioned carbon sources and were inoculated with the cells at an $OD_{600}$ of 0.01. The cells were grown aerobically in the CLARIOstar® PLUS at 37 °C, 25 °C or 42 °C. $OD_{600}$ and luminescence were recorded in 10 min intervals over the course of 16 h. Light emission was normalized to $OD_{600}$. Each measurement was performed in triplicates as a minimum.

### LDC assay

Cells were cultivated in LDC indicator medium (indicator: bromothymol blue) for 16 h and the pH increase was shown qualitatively as a color change[3].

### MgtL reporter assay

*E. coli* cells harboring the plasmids pBBR1-MCS5-*mgtL_luxCDABE* were inoculated in M9 minimal supplemented with the appropriate antibiotics and grown aerobically at 37 °C. The next day, a microtiter plate with fresh M9 minimal medium initially leaving out $Mg^{2+}$ ($Mg^{2+}$-free M9). Indicated concentrations of $Mg^{2+}$ (added as $MgSO_4$) were added subsequently. Cells were inoculated with a starting $OD_{600}$ of 0.01. Then cells were grown aerobically in the CLARIOstar® PLUS at 37 °C. $OD_{600}$ and luminescence was recorded in 10 min intervals over the course of 16 h. Light emission was normalized to $OD_{600}$. Each measurement was performed in triplicates as a minimum.

### Measurement of pausing strength in vivo

The pausing strength of different motifs was determined according to Krafczyk et al.[33] by measuring absorption at 600 nm (Number of flashes: 10; Settle time: 50 ms) and luminescence emission (Attenuation: none; Settle time: 50 ms; Integration time: 200 ms) with a Tecan Infinity® or ClarioStar plate reader in between 10-min cycles of agitation (orbital, 180 rpm, amplitude: 3 mm) for around 16 h.

### Competition experiments

For a direct comparison of *E. coli* lacking either *efp* or *efpL* with *E. coli* expressing both, different mixtures of *E. coli* BW25113 strains were analyzed over a time-course experiment. An *E. coli* strain Δ*cadC* without any mutant growth phenotype under the test conditions[92] was used as wild type. Single strains were incubated overnight at 37 °C shaking, washed in LB the following day, and resuspended to an $OD_{600}$ of 1. *E. coli* BW25113 Δ*cadC* was mixed with either *E. coli* JW4107-1 (BW25113 Δ*efp::Kan*$^R$) or JW5362-1 (BW25113 Δ*efpL::Kan*$^R$) and *E. coli* JW4094-5 (BW25113 Δ*cadC::Kan*$^R$) was mixed with either *E. coli* BW25113 Δ*efp* or BW25113 Δ*efpL*, to a starting $OD_{600}$ of 0.01 and cultivated for 3 h in LB or LB with 20 mM glucose. 100 cells from each mixture were plated on LB and LB with 50 μg/ml Kanamycin agar plates, respectively. Cultures were diluted to an $OD_{600}$ of 0.001 and cultivated for 24 h. On the next day colonies on the plates were counted, and the share of the population was calculated. The process was repeated as necessary.

### Protein overproduction and purification

For in vitro studies, C-terminally $His_6$-tagged EF-P and EfpL variants were overproduced in *E. coli* LMG194 harboring the corresponding pBAD33 plasmid. C-terminally $His_6$-tagged EfpL with acetyllysine instead of lysine at position 23, 40, 51 or 57 were overexpressed from pBAD33_*efpL*K23Amber_$His_6$, pBAD33_*efpL*K40Amber_$His_6$, pBAD33_*efpL*K51Amber_$His_6$, or pBAD33_*efpL*K57Amber_$His_6$ in *E. coli* LMG194 which contained the additional plasmid pACycDuet_AcKRST[59]. This allowed for amber suppression utilizing the acetyllysine-tRNA synthetase (AcKRS) in conjunction with PylT-tRNA. LB was supplemented with 5 mM $N^{\varepsilon}$-acetyl-L-lysine and 1 mM nicotinamide to prevent deacetylation by CobB[93]. During exponential growth, 0.2% (w/v) L(+)-arabinose was added to induce gene expression from pBAD vectors, and 1 mm IPTG served to induce gene expression of the pACycDuet-based system. Cells were grown overnight at 18 °C and harvested by centrifugation on the next day. The resulting pellet was resuspended in HEPES buffer (50 mM HEPES, 100 mM NaCl, 50 mM KCl, 10 mM $MgCl_2$, 5% (w/v) glycerol, pH 7.0). Cells were then lysed using a continuous-flow cabinet from Constant Systems Ltd (Daventry, UK) at 1.35 kbar. The resulting lysates were clarified by centrifugation at 4 °C at 234 998 × *g* for 1 h. The $His_6$-tagged proteins were purified using Ni-NTA beads (Qiagen, Hilden, Germany) according to the manufacturer's instructions, using 20 mM imidazole for washing and 250 mM imidazole for elution. In the final step, the purified protein was dialyzed overnight against HEPES buffer to remove imidazole from the eluate.

For MS analysis cells with chromosomally encoded $His_6$-tagged EfpL were grown in SOB until mid-exponential growth phase and harvested by centrifugation. To overproduce EfpL proteins LMG194 harboring a pBAD33 plasmid with C-terminally $His_6$-tagged EfpL were grown in SOB and supplemented with 0.2% (w/v) L(+)-arabinose during exponential growth phase ($OD_{600}$). Cells were grown overnight at 18 °C and harvested by centrifugation on the next day. Pellets were resuspended in 0.1 M sodium phosphate buffer, pH 7.6. Cells were then lysed using a continuous-flow cabinet from Constant Systems Ltd. (Daventry, UK) at 1.35 kbar. The resulting lysates were clarified by centrifugation at 4 °C at 234 998 × *g* for 1 h. The $His_6$-tagged proteins were purified using Ni-NTA beads (Qiagen, Hilden, Germany) according to the manufacturer's instructions. For washing and elution, a gradient of imidazole (10, 25, 50, 75, 100, 150, 200, 250 mM) was used. The purified protein was dialyzed overnight against in 0.1 M sodium phosphate buffer, pH 7.6 to remove imidazole from the eluate.

For crystallization, *E. coli* BL21 cells harboring a pET-SUMO plasmid were grown in SOB and supplemented with 1 mM IPTG during the exponential growth phase. Cells were grown overnight at 18 °C and harvested by centrifugation on the next. Pellets were resuspended in 0.5 M Tris-HCl buffer, pH 7.0. Cells were then lysed using a continuous-flow cabinet from Constant Systems Ltd. (Daventry, UK) at 1.35 kbar. The resulting lysates were clarified by centrifugation at 4 °C at 234 998 × *g* for 1 h. The $His_6$-tagged proteins were purified using Ni-NTA beads (Qiagen, Hilden, Germany) according to the manufacturer's instructions, using 20 mM imidazole for washing and 250 mM imidazole for elution. The purified protein was dialyzed overnight against 0.5 M Tris-HCl buffer, pH 7.0 to remove imidazole from the eluate. 0.33 mg SUMO-protease per 1 mg protein were added and incubated overnight at 4 °C. SUMO-protease and SUMO-tag were captured using Ni-NTA beads (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The protein was additionally purified via size exclusion chromatography on a Superdex 75 10/300 Increase column (Cytiva) in 20 mM Tris-HCl, 50 mM NaCl and 1 mM DTT at pH 8.0. Fractions with the protein of interest were concentrated and further subjected to anion exchange chromatography on a Resource Q (Bio-Rad) 6 ml-column to remove remaining contaminants with a NaCl salt gradient from 50 to 500 mM. The protein eluted at ~200 mM NaCl. The final sample was buffer-adjusted to 50 mM NaCl for crystallization.

### SDS–PAGE and western blotting

For protein analyses cells were subjected to 12.5% (w/v) sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE)[94]. To visualize proteins by UV light 2,2,2-trichloroethanol was added to the polyacrylamide gels[95]. Subsequently, the proteins were transferred onto nitrocellulose membranes, which were then subjected to immunoblotting. In a first step, the membranes were incubated either with 0.1 μg/ml anti-6×His® antibody (Abcam) or 0.1 μg/ml anti-acetylated-lysine (SIGMA). These primary antibodies, produced in rabbit, were targeted with 0.2 μg/ml anti-rabbit alkaline phosphatase-conjugated secondary antibody (Rockland) or 0.1 μg/ml anti-rabbit IgG (IRDye® 680RD) (donkey) antibodies (Abcam). Anti-rabbit alkaline phosphatase-conjugated secondary antibody was detected by adding development solution [50 mM sodium carbonate buffer, pH 9.5, 0.01% (w/v) p-nitro blue tetrazolium chloride (NBT), and 0.045% (w/v) 5-bromo-4-chloro-3-indolyl-phosphate (BCIP)]. Anti-rabbit IgG was visualized via Odyssey® CLx Imaging System (LI-COR, Inc).

### In vitro transcription/translation assay

The PURExpress In Vitro Protein Synthesis Kit from New England Biolabs was used according to the manufacturer's instructions, but reactions were supplemented with EF-P or EfpL, respectively, and a plasmid coding for *nluc* variants (Supplementary Data 4). Luminescence was measured over time. For a 12.5 μl reaction mixture, 5 μl of PURExpress solution A and 3.75 μl of solution B, 0.25 μl of Murine RNAse inhibitor (New England Biolabs), 5 μM EF-P or EfpL, and 1 ng pET16b_nluc variants are incubated under agitation (300 rpm) at 37 °C. At various time points, a 1 μl aliquot was quenched with 1 μl of 50 mg/ml kanamycin and stored on ice. Afterward, 2 μl of Nano-Glo Luciferase Assay Reagent (Promega) and 18 μl $ddH_2O$ were added to induce luminescence development, which was detected by the Infinite F500 microplate reader (Tecan®). At least three independent replicates were analyzed, and the statistical significance of the result was determined using GraphPad prism.

### Ribosome profiling

*E. coli* strains BW25113, BW25113 Δ*efpL*, BW25113 Δ*efp* and BW25113 Δ*efp* complemented with pBAD33-*efpL*_His6 (+EfpL) were cultivated in LB or LB supplemented with 30 μg/mL chloramphenicol and 0.2% L-(+)-arabinose at 37 °C under aerobic conditions. Stranded mRNA-seq and ribosome profiling (RiboSeq) libraries were generated by EIRNA Bio (https://eirnabio.com) from stab cultures. *E. coli* strains were grown in 400 mL LB at 37 °C to an OD600 of 0.4. Cells were harvested from 200 mL of culture by rapid filtration through a Kontes 90 mm filtration apparatus with 0.45 μm nitrocellulose filters (Whatman). Cells were scraped from the filter in two aliquots (90% for RiboSeq/10% for RNA-seq) before being immediately frozen in liquid nitrogen. Total RNA was extracted from RNA-seq aliquots in trizol before mRNA was rRNA depleted, fractionated, and converted into Illumina-compatible cDNA libraries. RiboSeq aliquots were lysed in 600 μl ice-cold polysome lysis buffer (20 mM Tris pH 8; 150 mM $MgCl_2$; 100 mM $NH_4Cl$; 5 mM $CaCl_2$; 0.4% Triton X-100; 0.1% NP-40; 20 U/ml Superase*In; 25 U/mM Turbo DNase) by bead beating in a FastPrep-24 with CoolPrep Adapter−3 rounds at 6 m/s for 30 s in 2 mL cryovials containing 0.1 mm silica beads. Lysates were clarified by centrifugation at 10,000 × *g* for 5 min at 4 °C. Ribosomes for subsequently pelleted from lysates by ultracentrifugation at 370,000 × *g* for 1 h at 4 °C and resuspended in polysome digestion buffer (20 mM Tris pH 8; 15 mM $MgCl_2$; 100 mM $NH_4Cl$; 5 mM $CaCl_2$). Samples were then digested with 750 U MNase for 1 h at 25 °C and the reaction was stopped by adding EGTA to a final concentration of 6 mM. Following RNA purification and size selection of ribosome-protected mRNA fragments between 20 and 40 nt in length on 15% urea PAGE gels, contaminating rRNA was depleted from samples using EIRNA Bio's custom biotinylated rRNA depletion oligos for *E. coli* before the enriched fragments were converted into Illumina-compatible cDNA libraries.

Both stranded mRNA-seq libraries and RiboSeq libraries were sequenced in three replicates on Illumina's Nova-seq 6000 platform in 150PE mode to depths of 10 million and 30 million raw read pairs per sample respectively.

The sequence structure of the RiboSeq reads was as follows:

QQQ−rpf sequence−NNNNN−BBBBB−AGATCGGAAGAGCACACGTCTGAA

, where Q = Untemplated Addition, rpf sequence = the sequence of the read, N = UMI, a 5 nt are unique molecular identifiers (UMIs), B = Barcode, used to demultiplex (the fastq files have already been demultiplexed) and AGATCGGAAGAGCACACGTCTGAA is the sequence of the adapter. Cutadapt[96] was used with parameters -u 3 and -a AGATCGGAAGAGCACACGTCTGAA to remove untemplated addition and linker sequence. Untrimmed reads and those shorter than 30 nt after trimming were discarded. Next, the UMI and Barcode were removed and the UMI was used to remove duplicate sequences using a custom Python script. Both the RiboSeq and RNA-seq reads were next mapped to rRNA and tRNA sequences using Bowtie version 1.2[97]. Five RiboSeq samples were sequenced with two sequencing runs. These samples (WT_Rep3, DELTAefpL_Rep2, DELTAefpL_Rep3, DELTAefp_Rep3, and DELTAefp_plus_efpL_Rep1) were concatenated at this stage. Next, the reads were aligned to BW25113 *E. coli* genome (RefSeq accession number NZ_CP009273.1) with Bowtie using parameters (-m 1 -l 25 -n 2 -S). BAM file containing read alignments are available at the SRA archive (ID PRJNA1092679).

The A-site offset in the RiboSeq reads was estimated to be 11 nucleotides upstream of the 3′ of the mapped reads. For both RiboSeq and RNAseq reads this "A-site" position was used to indicate the genomic location of reads. Pause prediction was carried out on all RiboSeq samples using PausePred[34] with a minimum fold-change for a pause score set at 20 within two sliding window sizes of 1000 nt with a minimum coverage of 5% in the window. The analysis was carried out on aggregated alignment files that included all replicates for each strain. The frequencies of occurrence of trimers of amino acid residues at the locations identified to be pauses were calculated for all possible trimers of amino acid residues. For each trimer of amino acid residues, its frequency to be covered by the ribosome in the pause sites was calculated and normalized by dividing by the averaged frequency of the corresponding trimer to occur in the whole ribosome-protected fragments.

### Sequence data and domain analysis

HMMER v.3.4 was used to search for Pfam[98] domains "EFP_N" (KOW-like domain, PF08207.12), "EFP" (OB-domain, PF01132.20), and "Elong-fact-P_C" (C-terminal, PF09285.11) in the protein sequences of 5257 complete representative or reference bacterial genomes (RefSeq)[25]. We identified 5448 proteins from 4736 genome assemblies that contained all three domains mentioned above (e-value cutoff 0.001) and no other PFAM domains. Sequences of "EFP_N" domains from these proteins were multiply aligned using Clustal Omega v.1.2.4[99] with all default parameters, shown in a multiple sequence alignment (MSA1) (Supplementary Data 5). Phylogenetic tree (Extended Data Fig. 1) was inferred by IQ-TREE 2.0.7[100] with branch support analysis performed in ultrafast mode[101] using 1000 bootstrap alignments. LG+R8 was chosen to be the best-fit model[102] for the tree. The phylogenetic tree in Newick format is available in the Supplementary Materials (Supplementary Data 6). The MSA region comprising positions 40–52 corresponds to the β3Ωβ4 loop region KPGKGQA of the EF-P protein from *E. coli* str. K-12 substr. MG1655 (accession number NP_418571.1)[22]. The sequence of the EfpL protein (NP_416676.4) from *E. coli* str. K-12 substr. MG1655 has an extended β3Ωβ4 loop SPTARGAAT with the R residue at the tip. The phylogenetic tree was annotated according to the length of the β3Ωβ4 loop and the nature of the residue at the tip of the β3Ωβ4 loop.

Those 528 sequences that have an extended β3Ωβ4 loop of more than 7 residues and R at the tip of it formed one branch in the phylogenetic tree. Among the sequences belonging to this branch 474 are annotated as "EfpL" or "YeiP" (synonym of EfpL) proteins in the RefSeq database and no other sequences from the list (Supplementary Data 1) have this annotation. Sequences with an extended β3Ωβ4 loop of more than seven residues and the R residue at the tip of it are referred to as EfpL. The remaining 4920 sequences constituted the set of EF-P sequences. The dataset covers 4777 genomes: 4111 of them contain only one sequence with the three domains mentioned above, 660 genomes contain two such sequences, and 6 genomes—three such sequences (Supplementary Data 1). In a separate analysis step, Clustal Omega v.1.2.4[99] with all default parameters were used to multiply align the sequences of KOW-like domains of the EfpL and EF-P proteins from the EfpL-containing genomes (MSA2) (Supplementary Data 5). IQ-TREE 2.0.7[100] with LG+G4 found to be the best-fit model[102] was used to build a phylogenetic tree (Fig. 1A). The branch support analysis in ultrafast mode[103] was performed using 1000 bootstrap alignments. The phylogenetic tree in Newick format, including bootstrap values, can be found in Supplementary Data (Supplementary Data 6). We used the ggtree R package[103] to visualize the phylogenetic trees and annotate them. Sequence logos were built using Weblogo[104].

EF-P-containing genomes were scanned for the EpmA, EarP and YmfI proteins. EpmA and EarP proteins were defined as single-domain proteins containing the "tRNA-synt 2" (PF00152.20) and "EarP" (PF10093.9)[14] domains, respectively. Using HMMER v.3.4 searches we identified these proteins in 1230 and 565 genomes, respectively. Orthologs of the YmfI protein (UniProt ID: O31767) from *Bacillus subtilis*[20,78] were obtained using the procedure described in Brewer and Wagner[23]. Briefly, this involved BLASTP(ref) searches using the *B. subtilis* YmfI as the query sequence, followed by manual bitscore cutoff determination due to the homology of this protein to other broadly conserved proteins.

### EfpL structure determination

Initial crystallization trials were performed in 96-well SWISSCI plates at a protein concentration of 4.8 mg/ml using the C3 ShotGun (SG1) crystallization screen (Molecular Dimension). Rod-shaped crystals grew after 7 days at 293 K. Diffracting crystals were obtained in 100 mM Sodium-HEPES, 20% (w/v) PEG 8000 and 10 mM Hexaamminecobalt (III) chloride conditions. The crystals were cryoprotected in mother liquor supplemented with 20% (v/v) glycerol and snap-frozen at 100 K. Datasets from cryo-cooled crystals were collected at EMBL P13 beamlines at the PETRA III storage ring of the DESY synchrotron[105]. The crystals belonged to space group P 1 21 1, with unit cell dimensions of $a = 60.71$, $b = 53.46$, and $c = 64.95$ Å. Preprocessed unmerged datasets from autoproc+STARANISO[106] were further processed in CCP4cloud[107]. Phases were obtained from molecular replacement using the AlphaFold2 model[102,108] deposited under ID AF-P0A6N8-F1. The structure was built using the automatic model building pipeline ModelCraft[109], optimized using PDB-REDO[110], refined in REFMAC5[111], and BUSTER[112] with manual corrections in Coot[113]. The quality of the built model was validated with the MolProbity server[114]. The final model was visualized in PyMOL version 2.55 (Delano Scientific). The asymmetric unit contained two molecules of EfpL. In chain A, residues 145–149 in OB-III loop showed weaker electron density and thus were not built. For depiction and comparison to other structures as well as for the HADDOCK procedure, chain B was chosen based on completeness and quality of the model. The data collection and refinement statistics are shown in Supplementary Data (Supplementary Data 2A).

### Docking and modeling of EF-P and EfpL complexes

For the comparative analysis of EF-P with the E-site codons CCG or GCG through a loop in its C-terminal OB-domain, we used the available PDB entry 6ENU[10] as a starting structure. In accordance with prior definitions by the authors, we directly analyzed and visualized available contacts for the EF-P d3 loop1 around the conserved motif $_{144}$GDT$_{146}$ with the present $_{-3}$CCG$_{-1}$ trinucleotide of the peptidyl-tRNA$^{Pro}$. For contacts with a putative GCG, we initially replaced the initial C nucleotide by G in silico using PyMol (Delano Scientific) and monitored the novel contacts using the implemented tools. For a more thorough analysis, we extracted both the GCG trinucleotide and EF-P from the structure and used the two components for an in silico docking followed by energy minimization using HADDOCK[27]. Here, we defined protein residues 146, 147, and 151 as active granting full flexibility to the structure and using automated secondary structure recognition and retainment. RNA residue G-3 was defined as active to enable seed contacts. From a total of 116 structures used by for clustering by HADDOCK 49 were found in the best-scoring cluster 1 (Supplementary Data 2B). Because of very low remaining restraint violation energies, we integrated the best four models to create an average structure used to analyze contacts between EF-P and RNA.

To analyze and compare interactions of EfpL and EF-P KOW domains with the P-site codon CCA through the β3Ωβ4 loop we looked at the available contacts of the loop as given in the PDB entry 6ENU[10]. For a model of EfpL with the trinucleotide, we aligned the EfpL KOW domain as found in our crystal structure with EF-P from PDB entry 6ENU[10]. We extracted the $_{74}$CCA$_{76}$trinucleotide from the latter and used the two components as starting structures for a docking and energy minimization procedure as described above. Nucleotides 74 and 75 were defined as active, and KOW domain residues 30–35 were set as fully flexible with R33 defined as explicitly active. One hundred and ninety-eight out of the 200 structures provided by HADDOCK were found in the same cluster with no measurable violations (Supplementary Data 2B).

For all HADDOCK runs, we implemented the following settings and restraints in context of the spatial and energetic constraints of the natural ribosome environment: Protein N- and C-termini were kept uncharged and no phosphates were left at nucleic acid termini. No particular RNA structure restraints have been applied and only polar hydrogens were installed in both components. For the 0th iteration, components were kept at their original positions for an initial energy-minimizing docking step. No random exclusion of ambiguous restraints was included during docking. Passive residues were defined automatically from the non-active ones using a surface distance threshold of 6.5 Å. We used a minimum percentage of relative solvent accessibility of 15 to consider a residue as accessible. In all runs 1000 initial structures were used in rigid body docking over five trials (excluding 180°-rotations of the ligand), from which the best 200 were subjected to an energy minimization step including short molecular dynamics simulations in explicit water. Default settings were used in advanced sampling parameters of the it1 and final solvated steps (Kyte–Doolittle), respectively. Standard HADDOCK settings were applied for clustering of the 200 final structures with a minimum cluster size of 4.

For the in silico analysis of modified lysines, respective sidechains were acetylated based on the EfpL crystal structure using PyMol with no further adjustments of rotamers. The modified KOW domain was then structurally aligned with EF-P in PDB entry 6ENU[10].

### Mass spectrometry for identification of modification status

For top-down EfpL measurements, the 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 −7 CV. We measured multiple spectra from the same protein sample. The MS spectra were acquired with at least 120,000 FWHM, AGC target 100 and 2–5 microscans. The spectra were deconvoluted in Freestyle 1.8 SP2 (Thermo Fisher Scientific) using the Xtract Deconvolution algorithm.

### Predicted growth rates

We used a set of 871 genomes from the class γ-proteobacteria from the Integrated Microbial Genomes database[115]. These genomes were selected to maximize diversity by including only one genome per Average Nucleotide Identity cluster. We used CheckM[116] v1.0.12 to assess the quality of each genome and retained only those that were predicted to be at least 90% complete and contain less than 5% contamination. We re-assigned taxonomy using the Genome Taxonomy Database and GTDB-Tool kit (GTDB-Tk)[117] version 0.2.2 and removed genomes where the user-reported species did not agree with GTDB (removed 2 genomes). For example, we removed a genome with a user-reported species of *Serratia marcescens 1822* which was sorted to the genus *Rouxiella* by GTDB-Tk. We also removed 14 genomes of endosymbionts from consideration, mainly from the genus *Buchnera*.

We further subset for only those genomes which contained both genes for *epmA* and *epmB* (removed 62 genomes), contained at least one *efp* gene (removed 2 genomes) and had predicted doubling times under 24 h (removed 15 genomes). This left 786 genomes for our analysis. We identified the genes for epmA, epmB, efp, and efpL (yeiP) using a combination of different functional databases. We identified *epmA* and *epmB* by searching for the COG[118] function ids COG2269 and COG1509, respectively. We identified *efp* by searching for the Pfam[119] domain pfam01132. We identified the gene for *efpL* (*yeiP*) by searching for the TIGRfam[120] annotation TIGR02178. Next, we estimated the doubling time associated with each remaining genome using the R package gRodon[65] version 1.8.0. gRodon estimates doubling times using codon usage bias in ribosomal proteins. We used phylogenetic ANOVAs to test differences in predicted doubling times between genomes that encode EfpL and those that don't. Specifically, we used the phylANOVA function from the R package phytools[121] version 2.0.3, with $p$ values based on 1000 permutations. We made the phylogenetic tree required for this function using 43 concatenated conserved marker genes generated by CheckM. We aligned these sequences using MUSCLE[122] v3.8.1551 and built the phylogenetic tree using IQ-TREE[123] v1.6.12. We used the model finder feature[124] included in IQ-TREE to determine the best-fit substitution model for our tree (which was the LG+R10 model). For this section, we performed all statistical analyses and plotting in R version 4.3.2 and created plots using ggplot2[125] version 3.4.4.

### Reporting summary

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

## Data availability

The crystal structure of $EfpL_{E.\ coli}$ generated in this study have been deposited in the PDB database under accession code 8S8U. The structure of $EF\text{-}P_{E.\ coli}$ from Huter et al.[10] was taken from the PDB database under accession code 6ENU. The ribosome profiling data generated in this study are available at SRD ID PRJNA1092679. Data on the acylation status of EfpL under the tested conditions can be found in the following publications by Kuhn et al.[55], Weinert et al.[56], Weinert et al.[57], and Qian et al.[58]. Quantitative *E. coli* proteome analysis data of Schmidt et al.[28] was used to compare protein concentrations in different conditions. Source data are provided with this paper.

## Code availability

R scripts and all files needed to reproduce the analyses on predicted growth rates are available at: https://github.com/tessbrewer/EfpL. An archived version of this repository has been generated and is accessible via Zenodo: https://doi.org/10.5281/zenodo.13897372.

## References


1. Tanner, D. R., Cariello, D. A., Woolstenhulme, C. J., Broadbent, M. A. & Buskirk, A. R. Genetic identification of nascent peptides that induce ribosome stalling. *J. Biol. Chem.* **284**, 34809–34818 (2009).
2. Doerfel, L. K. et al. EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science* **339**, 85–88 (2013).
3. Ude, S. et al. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* **339**, 82–85 (2013).
4. Hersch, S. J. et al. Divergent protein motifs direct elongation factor P-mediated translational regulation in *Salmonella enterica* and *Escherichia coli*. *mBio* **4**, e00180–00113 (2013).
5. Peil, L. et al. Distinct X/PP/X sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc. Natl. Acad. Sci. USA* **110**, 15265–15270 (2013).
6. Woolstenhulme, C. J., Guydosh, N. R., Green, R. & Buskirk, A. R. High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep.* **11**, 13–21 (2015).
7. Gutierrez, E. et al. eIF5A promotes translation of polyproline motifs. *Mol. Cell* **51**, 1–11 (2013).
8. Lassak, J., Wilson, D. N. & Jung, K. Stall no more at polyproline stretches with the translation elongation factors EF-P and IF-5A. *Mol. Microbiol.* **99**, 219–235 (2016).
9. Katoh, T., Wohlgemuth, I., Nagano, M., Rodnina, M. V. & Suga, H. Essential structural elements in $tRNA^{Pro}$ for EF-P-mediated alleviation of translation stalling. *Nat. Commun.* **7**, 11657 (2016).
10. Huter, P. et al. Structural basis for polyproline-mediated ribosome stalling and rescue by the translation elongation factor EF-P. *Mol. Cell* **68**, 515–527.e516 (2017).
11. Blaha, G., Stanley, R. E. & Steitz, T. A. Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. *Science* **325**, 966–970 (2009).
12. Hanawa-Suetsugu, K. et al. Crystal structure of elongation factor P from *Thermus thermophilus* HB8. *Proc. Natl. Acad. Sci. USA* **101**, 9595–9600 (2004).
13. Katz, A., Elgamal, S., Rajkovic, A. & Ibba, M. Non-canonical roles of tRNAs and tRNA mimics in bacterial cell biology. *Mol. Microbiol.* **101**, 545–558 (2016).
14. Lassak, J. et al. Arginine-rhamnosylation as new strategy to activate translation elongation factor P. *Nat. Chem. Biol.* **11**, 266–270 (2015).
15. Bailly, M. & de Crecy-Lagard, V. Predicting the pathway involved in post-translational modification of elongation factor P in a subset of bacterial species. *Biol. Direct* **5**, 3 (2010).
16. Yanagisawa, T., Sumida, T., Ishii, R., Takemoto, C. & Yokoyama, S. A paralog of lysyl-tRNA synthetase aminoacylates a conserved lysine residue in translation elongation factor P. *Nat. Struct. Mol. Biol.* **17**, 1136–1143 (2010).
17. Navarre, W. W. et al. PoxA, YjeK, and elongation factor P coordinately modulate virulence and drug resistance in *Salmonella enterica*. *Mol. Cell* **39**, 209–221 (2010).
18. Roy, H. et al. The tRNA synthetase paralog PoxA modifies elongation factor-P with (*R*)-β-lysine. *Nat. Chem. Biol.* **7**, 667–669 (2011).
19. Peil, L. et al. Lys34 of translation elongation factor EF-P is hydroxylated by YfcM. *Nat. Chem. Biol.* **8**, 695–697 (2012).

20. Rajkovic, A. et al. Translation control of swarming proficiency in *Bacillus subtilis* by 5-amino-pentanolylated elongation factor P. *J. Biol. Chem.* **291**, 10976–10985 (2016).
21. Li, X. et al. Resolving the α-glycosidic linkage of arginine-rhamnosylated translation elongation factor P triggers generation of the first $Arg^{Rha}$ specific antibody. *Chem. Sci.* **7**, 6995–7001 (2016).
22. Krafczyk, R. et al. Structural basis for EarP-mediated arginine glycosylation of translation elongation factor EF-P. *mBio* **8** https://doi.org/10.1128/mBio.01412-17 (2017).
23. Brewer, T. E. & Wagner, A. Horizontal Gene Transfer of a key Translation Factor and its Role in Polyproline Proteome Evolution. *Mol. Biol. Evol.* **41**, https://doi.org/10.1093/molbev/msae180 (2024).
24. Mudryi, V., Peske, F. & Rodnina, M. 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 (2023).
25. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
26. Golubev, A. et al. NMR and crystallographic structural studies of the Elongation factor P from *Staphylococcus aureus*. *Eur. Biophys. J.* **49**, 223–230 (2020).
27. van Zundert, G. C. P. et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
28. Schmidt, A. et al. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* **34**, 104–110 (2016).
29. Jung, K., Fabiani, F., Hoyer, E. & Lassak, J. Bacterial transmembrane signalling systems and their engineering for biosensing. *Open Biol.* **8**, https://doi.org/10.1098/rsob.180023 (2018).
30. Hong, H.-R., Prince, C. R., Tetreault, D. D., Wu, L. & Feaga, H. A. YfmR is a translation factor that prevents ribosome stalling and cell death in the absence of EF-P. *Proc. Natl. Acad. Sci. USA* **121**, e2314437121 (2024).
31. Chadani, Y. et al. The ABCF proteins in *Escherichia coli* individually cope with 'hard-to-translate' nascent peptide sequences. *Nucleic Acids Res.* **52**, 5825–5840 (2024).
32. Takada, H., Fujiwara, K., Atkinson, G. C., Chiba, S. & Hauryliuk, V. Resolution of ribosomal stalling by EF-P and ABCF ATPases YfmR and YkpA/YbiT. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkae556 (2024).
33. Krafczyk, R. et al. Proline codon pair selection determines ribosome pausing strength and translation efficiency in bacteria. *Commun. Biol.* **4**, 589 (2021).
34. Kumari, R., Michel, A. M. & Baranov, P. V. PausePred and Rfeet: webtools for inferring ribosome pauses and visualizing footprint density from ribosome profiling data. *RNA* **24**, 1297–1304 (2018).
35. Schuller, A. P., Wu, C. C., Dever, T. E., Buskirk, A. R. & Green, R. eIF5A functions globally in translation elongation and termination. *Mol. Cell* **66**, 194–205.e195 (2017).
36. Pelechano, V. & Alepuz, P. eIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic Acids Res.* **45**, 7326–7338 (2017).
37. Gall, A. R. et al. Mg2+ regulates transcription of *mgtA* in *Salmonella Typhimurium* via translation of proline codons during synthesis of the MgtL peptide. *Proc. Natl. Acad. Sci. USA* **113**, 15096–15101 (2016).
38. Nam, D., Choi, E., Shin, D. & Lee, E. J. $tRNA^{Pro}$-mediated down-regulation of elongation factor P is required for *mgtCBR* expression during *Salmonella* infection. *Mol. Microbiol.* **102**, 221–232 (2016).
39. Takada, H., Fujiwara, K., Atkinson, G. C., Chiba, S. & Hauryliuk, V. Resolution of ribosomal stalling by ABCF ATPases YfmR and YkpA/YbiT. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkae556 (2024).
40. Wang, M. *Proteome Characterization of Escherichia coli Cells Evolved to Tolerate Elongation Factor P Deletion*. Doctoral thesis, Technische Universität Berlin (2017).
41. Ousalem, F. et al. Comparative genetic, biochemical, and biophysical analyses of the four *E. coli* ABCF paralogs support distinct functions related to mRNA translation. *bioRxiv* https://doi.org/10.1101/2023.06.11.543863 (2023).
42. Starosta, A. L. et al. Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site. *Nucleic Acids Res.* **42**, 10711–10719 (2014).
43. Elgamal, S. et al. EF-P dependent pauses integrate proximal and distal signals during translation. *PLoS Genet.* **10**, e1004553 (2014).
44. Mudryi, V. *Elongation Factor P: Mechanism of Action and Opportunities for Drug Design*. Doctoral thesis, Georg-August Universität Göttingen (2023).
45. Qi, F., Motz, M., Jung, K., Lassak, J. & Frishman, D. Evolutionary analysis of polyproline motifs in *Escherichia coli* reveals their regulatory role in translation. *PLoS Comput. Biol.* **14**, e1005987 (2018).
46. Mohapatra, S., Choi, H., Ge, X., Sanyal, S. & Weisshaar, J. C. Spatial distribution and ribosome-binding dynamics of EF-P in live *Escherichia coli*. *mBio* **8**, https://doi.org/10.1128/mBio.00300-17 (2017).
47. Raghavan, R., Sage, A. & Ochman, H. Genome-wide identification of transcription start sites yields a novel thermosensing RNA and new cyclic AMP receptor protein-regulated genes in *Escherichia coli*. *J. Bacteriol.* **193**, 2871–2874 (2011).
48. Graf von Armansperg, B. et al. Transcriptional regulation of the $N_ε$-fructoselysine metabolism in *Escherichia coli* by global and substrate-specific cues. *Mol. Microbiol.* **115**, 175–190 (2021).
49. Deana, A., Celesnik, H. & Belasco, J. G. The bacterial enzyme RppH triggers messenger RNA degradation by 5′ pyrophosphate removal. *Nature* **451**, 355–358 (2008).
50. Richards, J., Luciano, D. J. & Belasco, J. G. Influence of translation on RppH-dependent mRNA degradation in *Escherichia coli*. *Mol. Microbiol.* **86**, 1063–1072 (2012).
51. Lee, C. R., Kim, M., Park, Y. H., Kim, Y. R. & Seok, Y. J. RppH-dependent pyrophosphohydrolysis of mRNAs is regulated by direct interaction with DapF in *Escherichia coli*. *Nucleic Acids Res.* **42**, 12746–12757 (2014).
52. Luciano, D. J., Levenson-Palmer, R. & Belasco, J. G. Stresses that raise $Np_4A$ levels induce protective nucleoside tetraphosphate capping of bacterial RNA. *Mol. Cell* **75**, 957–966.e958 (2019).
53. Ishfaq, M. et al. Acetylation regulates subcellular localization of eukaryotic translation initiation factor 5A (eIF5A). *FEBS Lett.* **586**, 3236–3241 (2012).
54. Chung, J., Rocha, A. A., Tonelli, R. R., Castilho, B. A. & Schenkman, S. Eukaryotic initiation factor 5A dephosphorylation is required for translational arrest in stationary phase cells. *Biochem. J.* **451**, 257–267 (2013).
55. Kuhn, M. L. et al. Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. *PLoS ONE* **9**, e94816 (2014).
56. Weinert, B. T. et al. Acetyl-phosphate is a critical determinant of lysine acetylation in *E. coli*. *Mol. Cell* **51**, 265–272 (2013).
57. Weinert, B. T. et al. Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. *Cell Rep.* **4**, 842–851 (2013).
58. Qian, L. et al. Global profiling of protein lysine malonylation in *Escherichia coli* reveals its role in energy metabolism. *J. Proteome Res.* **15**, 2060–2071 (2016).

59. Volkwein, W., Maier, C., Krafczyk, R., Jung, K. & Lassak, J. A versatile toolbox for the control of protein levels using $N_\varepsilon$-acetyl-l-lysine dependent amber suppression. *ACS Synth. Biol.* **6**, 1892–1902 (2017).
60. Lassak, J., Sieber, A. & Hellwig, M. Exceptionally versatile take II: post-translational modifications of lysine and their impact on bacterial physiology. *Biol. Chem*. https://doi.org/10.1515/hsz-2021-0382 (2022).
61. Kovach, M. E. et al. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**, 175–176 (1995).
62. Feid, S. C. et al. Regulation of translation by lysine acetylation in *Escherichia coli*. *mBio* **13**, e0122422 (2022).
63. Perez, J. C. & Groisman, E. A. Evolution of transcriptional regulatory circuits in bacteria. *Cell* **138**, 233–244 (2009).
64. Eagon, R. G. *Pseudomonas natriegens*, a marine bacterium with a generation time of less than 10 minutes. *J. Bacteriol.* **83**, 736–737 (1962).
65. Weissman, J. L., Hou, S. & Fuhrman, J. A. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc. Natl. Acad. Sci. USA* **118**, https://doi.org/10.1073/pnas.2016810118 (2021).
66. Starosta, A. L. et al. A conserved proline triplet in Val-tRNA synthetase and the origin of elongation factor P. *Cell Rep.* **9**, 476–483 (2014).
67. Adzhubei, A. A., Sternberg, M. J. & Makarov, A. A. Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.* **425**, 2100–2132 (2013).
68. Pavlov, M. Y. et al. Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc. Natl. Acad. Sci. USA* **106**, 50–54 (2009).
69. Wohlgemuth, I., Brenner, S., Beringer, M. & Rodnina, M. V. Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. *J. Biol. Chem.* **283**, 32229–32235 (2008).
70. Mandal, A., Mandal, S. & Park, M. H. Genome-wide analyses and functional classification of proline repeat-rich proteins: potential role of eIF5A in eukaryotic evolution. *PLoS ONE* **9**, e111800 (2014).
71. Volkwein, W. et al. Switching the post-translational modification of translation elongation factor EF-P. *Front. Microbiol.* **10**, 1148 (2019).
72. Novozhilov, A. S. & Koonin, E. V. Exceptional error minimization in putative primordial genetic codes. *Biol. Direct* **4**, 44 (2009).
73. Ignatov, D. et al. Novel RNA-binding protein YebC enhances translation of proline-rich amino acid stretches in bacteria. *bioRxiv*, 2024.2008.2026.607280 https://doi.org/10.1101/2024.08.26.607280 (2024).
74. Brischigliaro, M. et al. The human mitochondrial translation factor TACO1 alleviates mitoribosome stalling at polyproline stretches. *Nucleic Acids Res.* **52**, 9710–9726 (2024).
75. Singh, S. et al. Cryo-EM studies of the four *E. coli* paralogs establish ABCF proteins as master plumbers of the peptidyl-transferase center of the ribosome. *bioRxiv*, 2023.2006.2015.543498 https://doi.org/10.1101/2023.06.15.543498 (2023).
76. Pinheiro, D. F. B. *The Translation Elongation Factor P in Actinobacteria*. Dissertation, Ludwig-Maximilians-Universität München (2020).
77. Veening, J. W. et al. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc. Natl. Acad. Sci. USA* **105**, 4393–4398 (2008).
78. Hummels, K. R. & Kearns, D. B. Translation elongation factor P (EF-P). *FEMS Microbiol. Rev.* **44**, 208–218 (2020).
79. Pinheiro, B. et al. Structure and function of an elongation factor P subfamily in actinobacteria. *Cell Rep.* **30**, 4332–4342.e4335 (2020).
80. Doerfel, L. K. et al. Entropic contribution of elongation factor P to proline positioning at the catalytic center of the ribosome. *J. Am. Chem. Soc.* **137**, 12997–13006 (2015).
81. Katoh, T., Tajima, K. & Suga, H. Consecutive elongation of D-amino acids in translation. *Cell Chem. Biol.* **24**, 46–54 (2017).
82. Katoh, T. & Suga, H. Ribosomal incorporation of consecutive β-amino acids. *J. Am. Chem. Soc.* **140**, 12159–12167 (2018).
83. Katoh, T. & Suga, H. Translation initiation with exotic amino acids using EF-P-responsive artificial initiator tRNA. *Nucleic Acids Res.* **51**, 8169–8180 (2023).
84. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose $P_{BAD}$ promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
85. Lassak, J., Henche, A. L., Binnenkade, L. & Thormann, K. M. ArcS, the cognate sensor kinase in an atypical Arc system of *Shewanella oneidensis* MR-1. *Appl. Environ. Microbiol.* **76**, 3263–3274 (2010).
86. Lassak, J., Bubendorfer, S. & Thormann, K. M. Domain analysis of ArcS, the hybrid sensor kinase of the *Shewanella oneidensis* MR-1 Arc two-component system, reveals functional differentiation of its two receiver domains. *J. Bacteriol.* **195**, 482–492 (2013).
87. Gödeke, J., Heun, M., Bubendorfer, S., Paul, K. & Thormann, K. M. Roles of two *Shewanella oneidensis* MR-1 extracellular endonucleases. *Appl. Environ. Microbiol.* **77**, 5342–5351 (2011).
88. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293–300 (1951).
89. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, 1972).
90. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).
91. Nandi, B. et al. Rapid method for species-specific identification of *Vibrio cholerae* using primers targeted to the gene of outer membrane protein OmpW. *J. Clin. Microbiol.* **38**, 4145–4151 (2000).
92. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006-0008 (2006).
93. Gallego-Jara, J., Écija Conesa, A., de Diego Puente, T., Lozano Terol, G. & Cánovas Díaz, M. Characterization of CobB kinetics and inhibition by nicotinamide. *PLoS ONE* **12**, e0189689 (2017).
94. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
95. Ladner, C. L., Yang, J., Turner, R. J. & Edwards, R. A. Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal. Biochem.* **326**, 13–20 (2004).
96. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* **17**, 3 (2011).
97. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
98. Finn, R. D. et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
99. Sievers, F. & Higgins, D. G. The Clustal Omega Multiple Alignment Package. *Methods Mol. Biol.* **2231**, 3–16 (2021).
100. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
101. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
102. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
103. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an R package for visualization and annotation of phylogenetic trees

with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
104. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
105. Cianci, M. et al. P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. *J. Synchrotron Radiat.* **24**, 323–332 (2017).
106. Vonrhein, C. et al. Advances in automated data analysis and processing within autoPROC, combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using STARANISO. *Acta Crystallogr. A Found. Adv.* **74**, A360 (2018).
107. Krissinel, E., Uski, V., Lebedev, A., Winn, M. & Ballard, C. Distributed computing for macromolecular crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 143–151 (2018).
108. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
109. Bond, P. S. & Cowtan, K. D. ModelCraft: an advanced automated model-building pipeline using Buccaneer. *Acta Crystallogr. D Struct. Biol.* **78**, 1090–1098 (2022).
110. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).
111. Kovalevskiy, O., Nicholls, R. A., Long, F., Carlon, A. & Murshudov, G. N. Overview of refinement procedures within REFMAC5: utilizing data from different sources. *Acta Crystallogr. D Struct. Biol.* **74**, 215–227 (2018).
112. Bricogne, G., Blanc, E., Brandl, M. T., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., & Vonrhein, C. BUSTER version 2.10.3 (Global Phasing Ltd, Cambridge, UK, 2011).
113. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
114. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
115. Chen, I. A. et al. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* **49**, D751–d763 (2021).
116. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
117. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).
118. Galperin, M. Y. et al. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **49**, D274–D281 (2020).
119. Mistry, J. et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* **49**, D412–d419 (2021).
120. Li, W. et al. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Res.* **49**, D1020–D1028 (2021).
121. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
122. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
123. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
124. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
125. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* (Springer Science+Business Media, LLC, 2016).
126. Mi, H. et al. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* **14**, 703–721 (2019).

## Acknowledgements

We thank Kirsten Jung, Kerstin Lassak and Wolfram Volkwein for fruitful discussions and constructive criticism. We also thank Giovanni Gallo, Anja Michl, Nina Kim Hartmann, Bhavna Menon, Irem Niran Cagil and Sabine Peschek for technical assistance. Moreover, Jürgen Lassak is grateful for DFG grant LA 3658/5-1. This work was supported by Liebig fellowship from VCI to Pavel Kielkowski. Fei Qi is supported by the National Natural Science Foundation of China (grant No. 32000462). Andreas Schlundt acknowledges support through funds SCHL2062/2-1 and 2-2 from the German Research Council (DFG) and the Johanna Quandt Young Academy at Goethe (grant number 2019/AS01). Access to synchrotron beamtime at DESY Hamburg was enabled by the Block Allocation Group grant MX939.

## Author contributions

A. Sieber, R.K., A. Schäpers, and J.L. constructed strains and plasmids. A. Sieber and R.K. performed the biochemical in vivo/in vitro characterization of EfpL. A. Schäpers performed qPCR experiments. A. Sieber purified all proteins used for in vitro assays, MS analysis and X-ray crystallography. MS experiments and analysis were done by P.K. The crystallization screen was set up by J.v.E., K.D., and A. Schlundt; and J.v.E., K.D., and A. Schlundt solved the crystal structure of $EfpL_{E.\ coli}$. A. Schlundt performed in silico interaction analyses. All bioinformatic analyses were performed by M.P., T.B., F.Q., and D.F. Ribosome profiling analyses were done by M.P. and D.F. M.P. and D.F. performed phylogenetic analyses of the EF-P subgroups and T.B. performed phylogenetic analyses of bacterial growth rates. The study was designed by J.L. with contributions from R.K. and D.F. The manuscript was written by A. Sieber, M.P., and J.L. with contributions from A. Schlundt, T.B. and D.F.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interest.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-54556-9.

**Correspondence** and requests for materials should be addressed to Jürgen Lassak.

**Peer review information** *Nature Communications* thanks Vasili Hauryliuk, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

**Supplemental Information Chapter 3**

# SUPPLEMENTARY INFORMATION

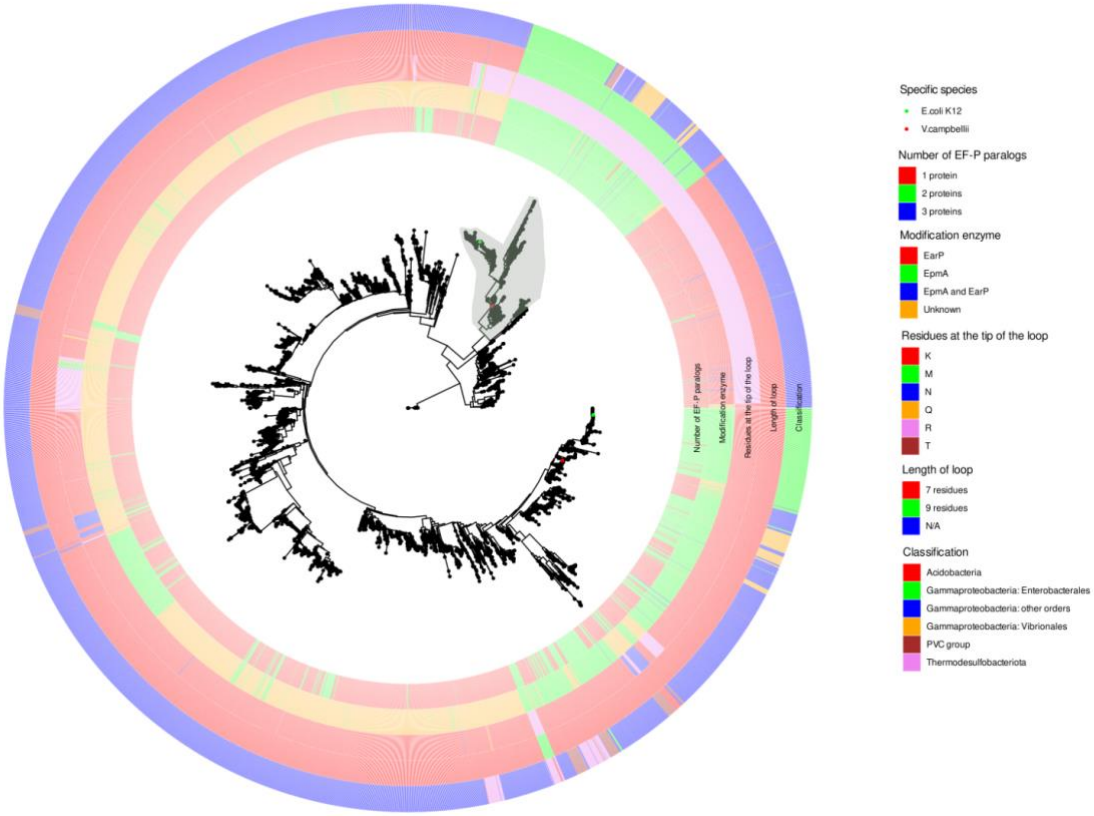
## EF-P and its paralog EfpL (YeiP) differentially control translation of proline-containing sequences

*Alina Sieber[1]*, Marina Parr[2]*, Julian von Ehr[3,4], Karthikeyan Dhamotharan[3], Pavel Kielkowski[5], Tess Brewer[1], Anna Schäpers[1], Ralph Krafczyk[1], Fei Qi[6], Andreas Schlundt[3,7], Dmitrij Frishman[2] & Jürgen Lassak[1+]*

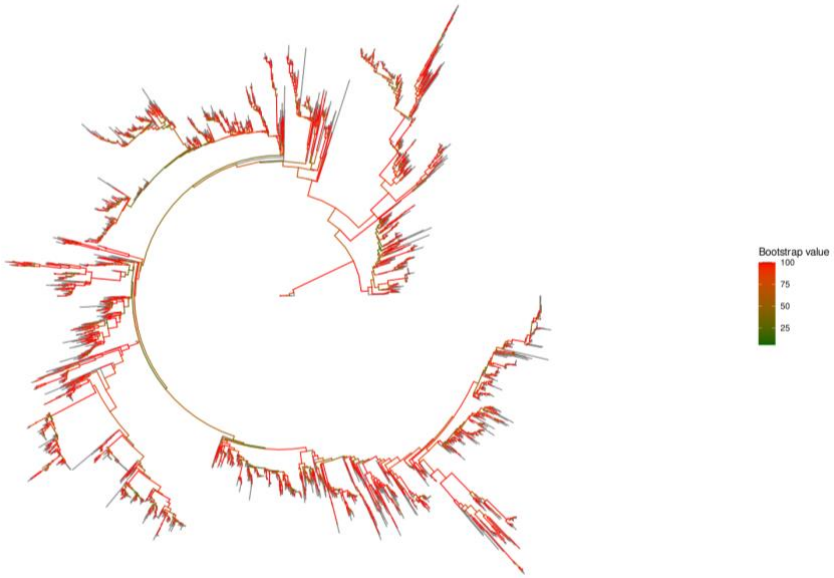
[1]Faculty of Biology, Microbiology, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany
[2]Department of Bioinformatics, Wissenschaftzentrum Weihenstephan, Technische Universität München, 85354 Freising, Germany
[3]Institute for Molecular Biosciences and Biomolecular Resonance Center (BMRZ), Goethe University Frankfurt, Max-von-Laue-Str. 7-9, 60438 Frankfurt, Germany.
[4]IMPRS on Cellular Biophysics, Max-von-Laue-Str. 7-9, 60438 Frankfurt, Germany.
[5]Department of Chemistry, Institut für Chemische Epigenetik (ICEM), Ludwig-Maximilians-Universität München, 81377 Munich, Germany
[6]State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, 361102 Xiamen, China
[7]University of Greifswald, Institute of Biochemistry, Felix-Hausdorff-Str. 4, 17489 Greifswald, Germany

*These authors contribute equally; [+]corresponding author: juergen.lassak@lmu.de

**A**

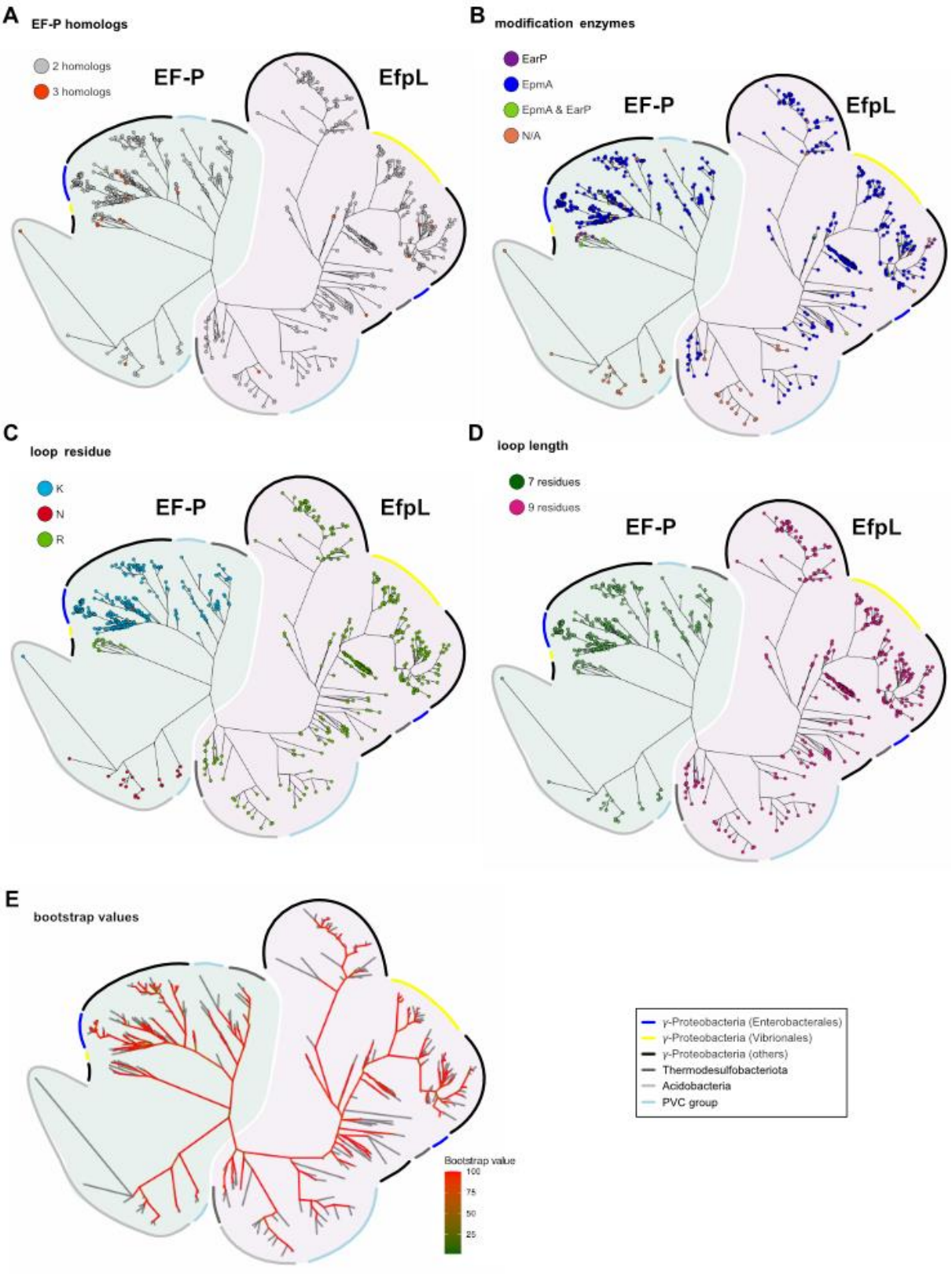



**B**



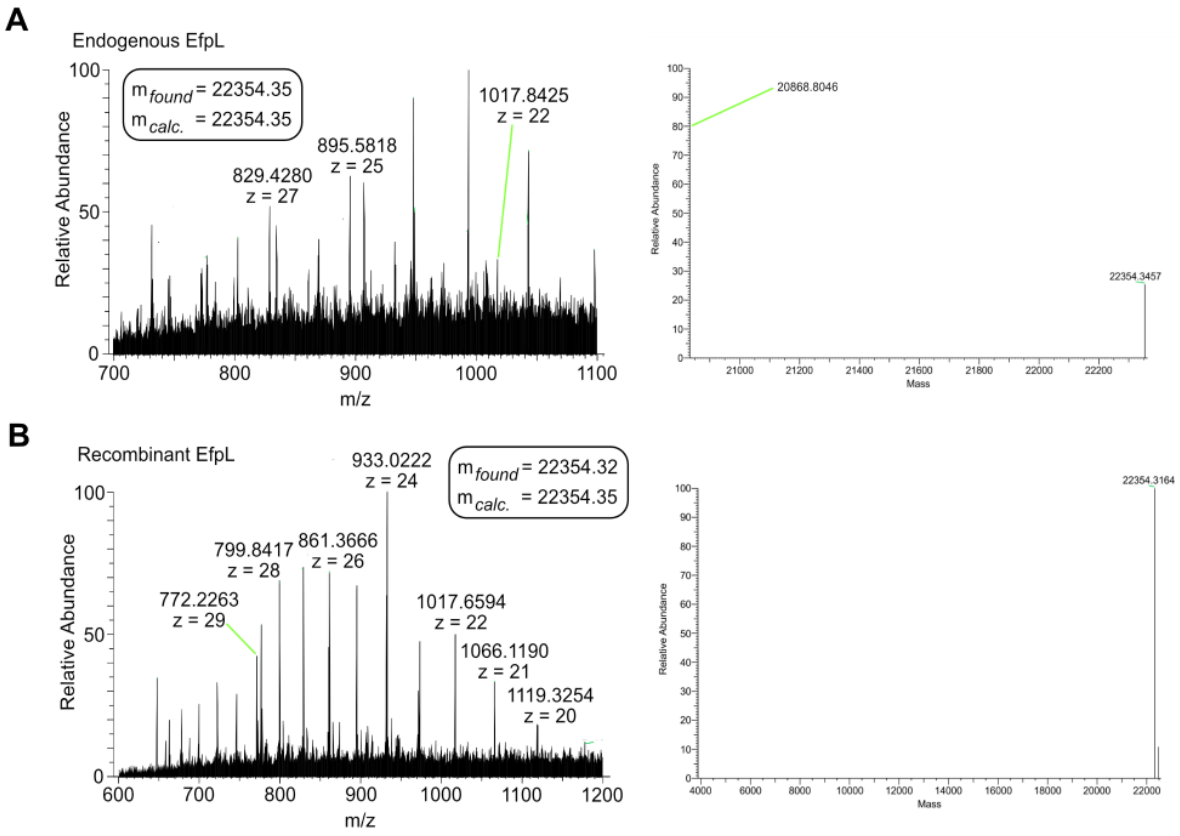

**Supplementary Fig. 1: Phylogenetic analysis of EF-P subgroups**

Phylogenetic tree was built using the multiply aligned 5448 sequences of KOW-like domains of proteins that have three domains typical for EF-P in a collection of 4736 complete bacterial genomes was obtained from the RefSeq database[25]. **(A)** Outer ring shows phylogenetic classification in bacterial phyla. Other rings show tip residues and length of β3Ωβ4 loop, as well as number of EF-P homologs and modification enzymes found in bacterial proteomes. Branch endings indicate affiliation to specific species. The green highlighting indicates the branch with protein sequences further annotated as EfpL proteins. **(B)** Corresponding bootstrap values.



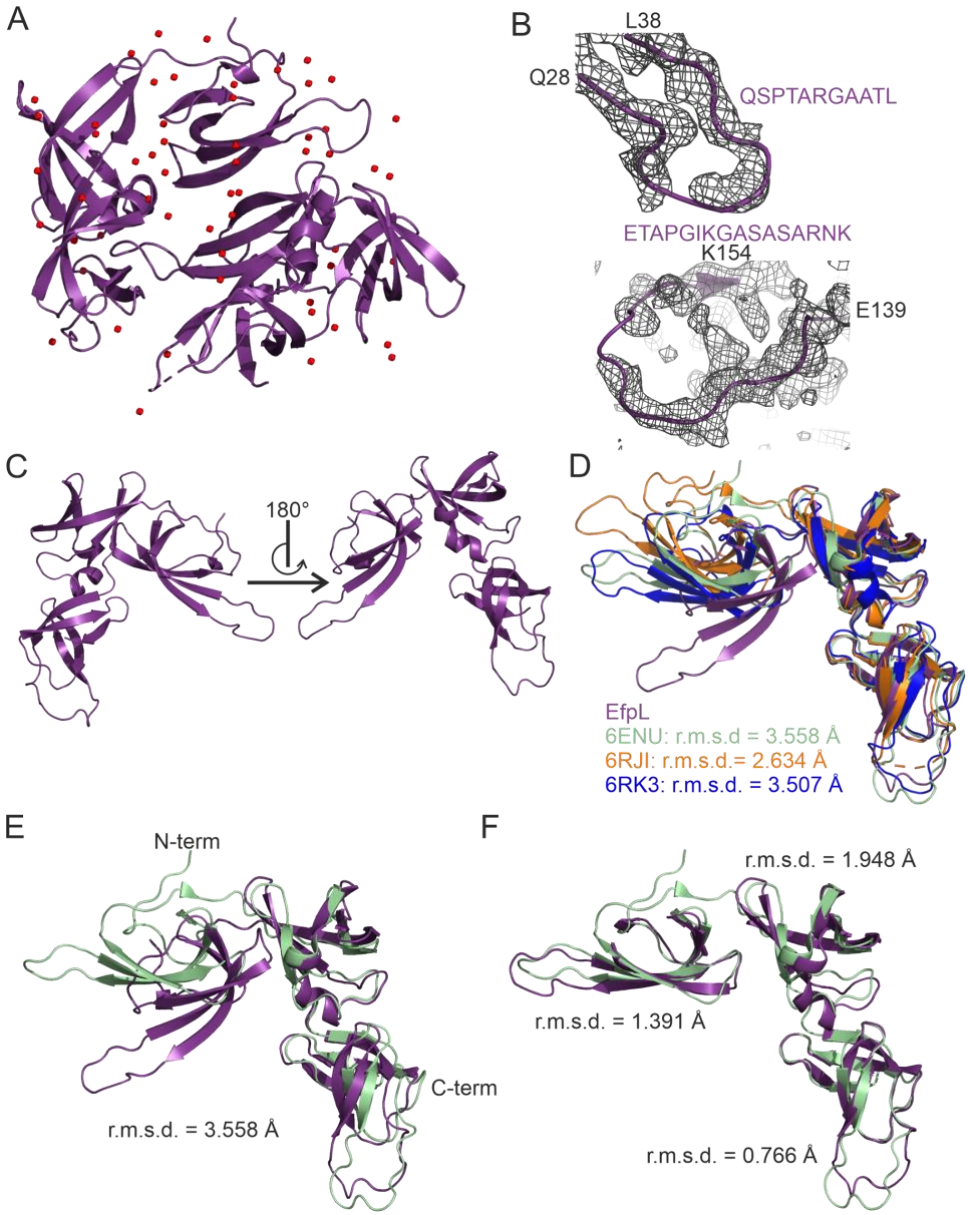

**Supplementary Fig. 2: Phylogenetic analysis of co-occurring EF-P and EfpL proteins**

Phylogenetic tree of EfpL (purple) and co-occurring EF-Ps (green). Colored lines indicate bacterial phyla. **(A)** EF-P homologs per proteome. **(B)** EF-P modification enzymes found per proteome. **(C)** β3Ωβ4 loop tip residue in EF-P or EfpL. **(D)** β3Ωβ4 loop length of EF-P or EfpL. **(E)** Corresponding bootstrap values.



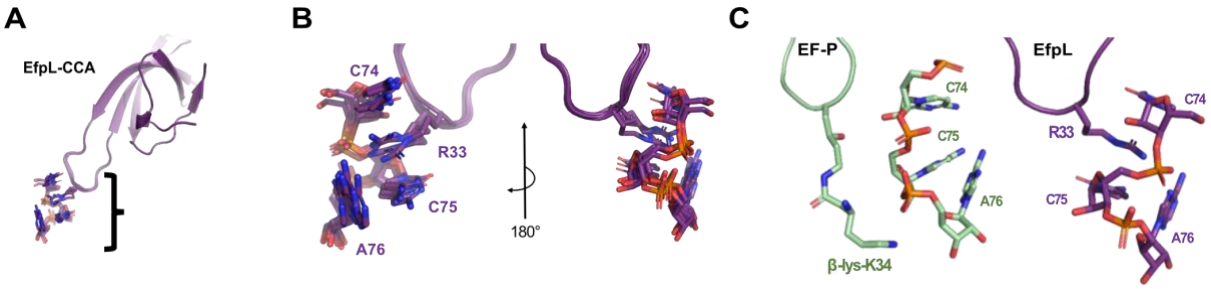

**Supplementary Fig. 3: Mass spectrometry (MS) for EfpL protein analysis**

MS spectra for **(A)** endogeneous *E. coli* EfpL (n=1, multiple spectra from same protein sample) and **(B)** recombinant produced *E. coli* EfpL (n=1, multiple spectra from same protein sample) to identify modification status. Left side: DDA raw files; right side: output mzML format. Mass ($m_{calc.}$) was calculated according the Uniprot database (identifier: B7UFI8 - EFPL_ECO27) [1]. 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.



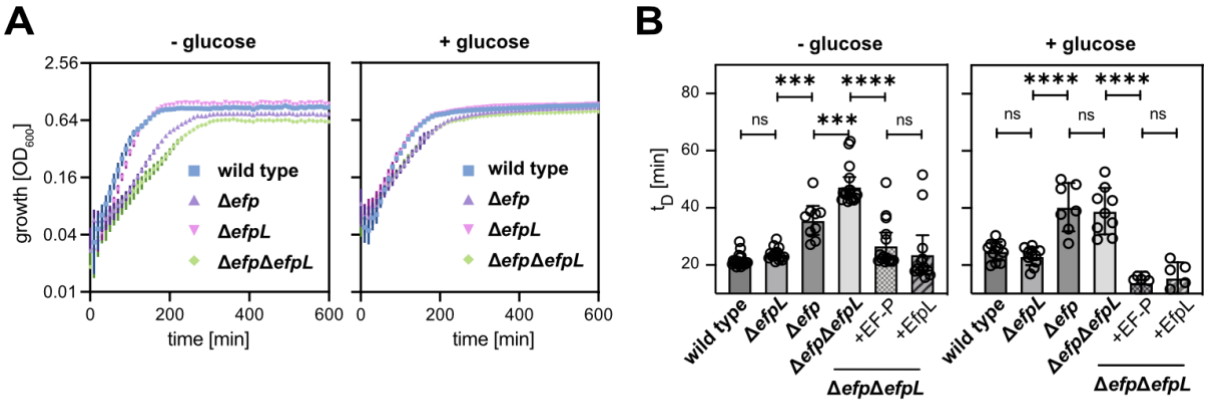

**Supplementary Fig. 4: Structural comparison of EfpL and EF-P**

**(A)** Structure of crystallographic dimer of EfpL determined by x-ray crystallography in this study. Waters and co-crystallized glycerol ligands are colored in red and blue respectively. **(B)** Electron density map of KOW β3Ωβ4 loop (upper panel) and OB d3 loop 1 (lower panel). The 2Fo-Fc electron density is contoured at 1.5 σ. **(C)** Two-sided view of fully build single chain from the EfpL x-ray structure in A. **(D)** Structural alignment of EfpL with EF-P from *E. coli* (cryo-EM structure, PDB entry 6ENU) and *S. aureus* (crystal structure, PDB entry 6RJI and NMR solution structure, PDB entry 6RK3). R.m.s.d. values in comparison to EfpL are shown. **(E)** Structural alignment of EfpL with EF-P from *E. coli* (cryo-EM structure, PDB entry 6ENU). Root mean square deviation (r.m.s.d) of the total alignment is shown. **(F)** The same as in E but with structured domains (residues 4-56, 68-128, 132-187) of EfpL separated and aligned individually to EF-P and are shown with respective r.m.s.d.



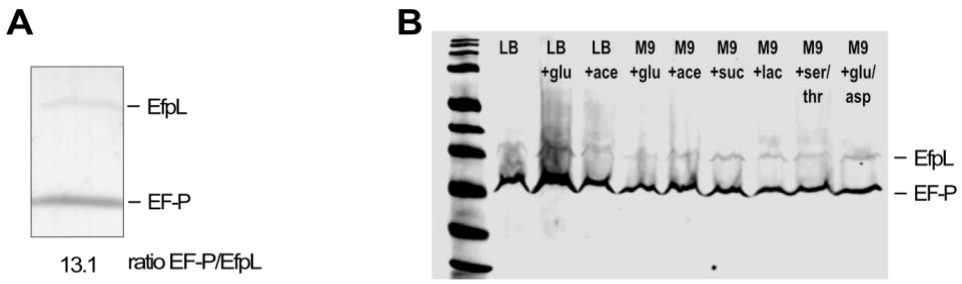

**Supplementary Fig. 5: *In silico* interaction analysis of EF-P and EfpL with the tRNA CCA trinucloetide with the RNA trinucleotide**

**(A)** Full-view superimposition of the four best solutions obtained from a HADDOCK run of EfpL together with $_{74}CCA_{76}$. **(B)** Zoom-in of panel A to the KOW domain β3Ωβ4 loop region in contact with the RNA trinucleotide. The r.m.s.d. is 0.12 ± 0.01 Å. **(C)** Excerpt of β3Ωβ4 loops from EF-P and EfpL (see panels A and B) in complex with the tRNA trinucleotide CCA. The central tip residue of EF-P is β-lysylated and the depicted complex based on PDB ID 6ENU[3].



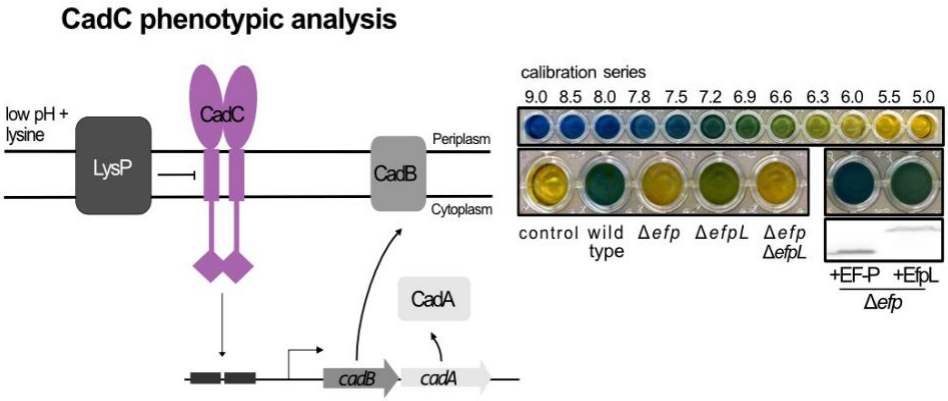

**Supplementary Fig. 6: Growth analysis of *E. coli* wild type and deletion strains.**

Growth analysis of *E. coli* BW25113 wild type and deletion strains in LB with and without addition of 20 mM glucose. **(A)** Growth curves shown for 600 minutes (n=10, biological replicates, mean with sd indicated as error bars). **(B)** For complementation *efp* (+EF-P) or *efpL* (+EfpL) were provided in trans. Doubling times ($t_D$) were calculated from exponentially grown cells in LB (n ≥ 6, biological replicates, mean with sd indicated as error bars) Statistically significant differences according to ordinary one-way ANOVA test with multiple comparison (*P value <0.0332, **P value <0.0021, ***P value <0.0002, ****P value <0.0001, ns not significant). **(A&B)** Source data are provided as a Source Data file



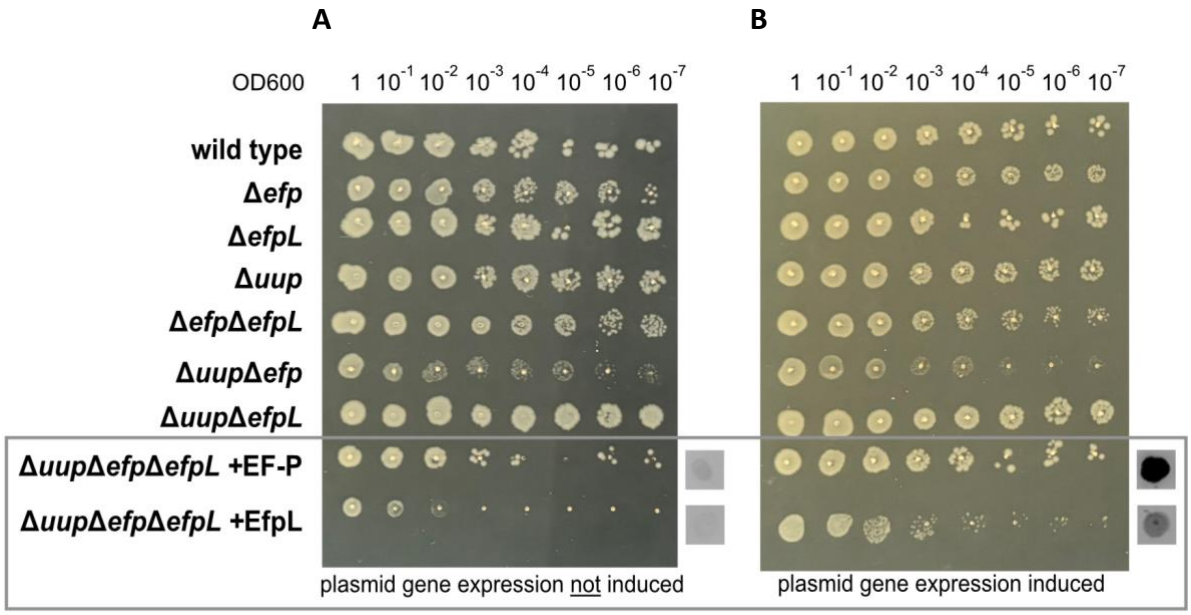

**Supplementary Fig. 7: Protein amount of EF-P and EfpL in *E. coli***

To quantify endogenous production of EF-P and EfpL a 6xHis encoding sequence was genomically integrated at the 3' end of the ORFs of *efp* and *efpL* in *E. coli* BW25113. **(A)** Production in LB was quantified via immunoblotting using Anti-His6 antibodies. Ratio determined using Fiji[2]. **(B)** Potein production was detected with different growth media, LB and LB supplemented with 40 mM glucose (LB+glu), and minimal media with different C-sources: M9 with 40 mM glucose (M9+glu), 40 mM acetate (M9+ace), 40 mM succinate (M9+suc), 40 mM lactose (M9+lac), 20mM serine together with 20 mM threonine (M9+ser/thr), and 20 mM glutamate together with 20 mM aspartate (M9+glu/asp). **(A&B)** Source data are provided as a Source Data file




**Supplementary Fig. 8: CadC phenotypic analysis**

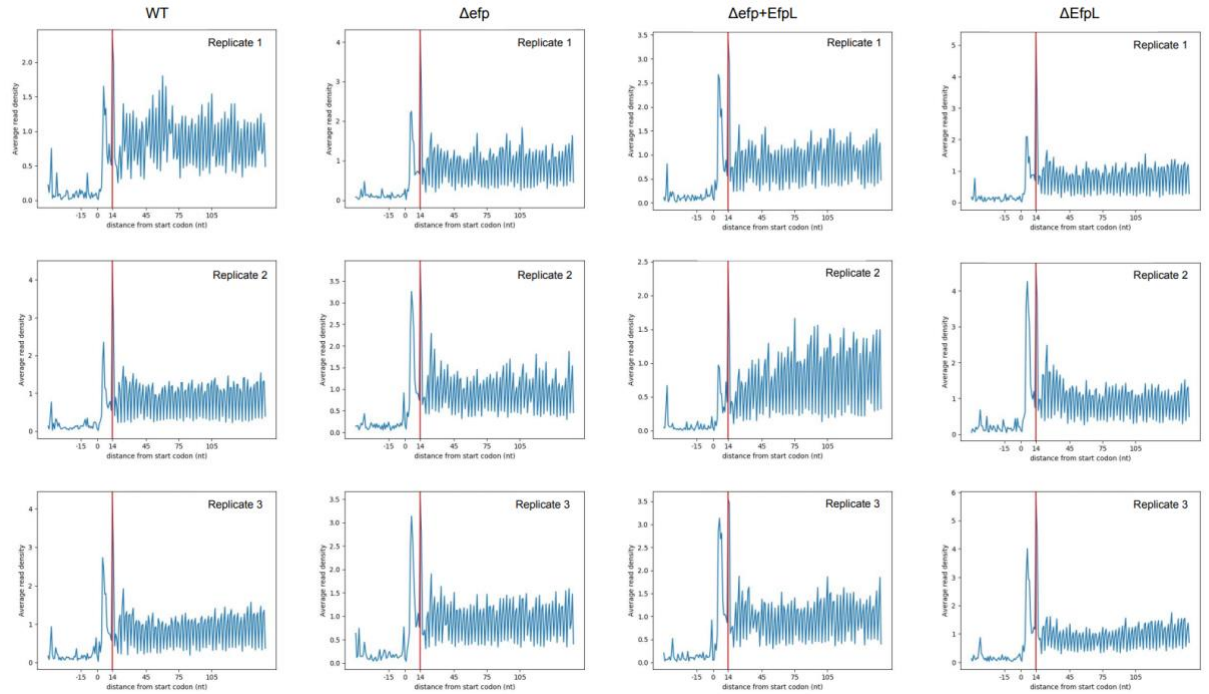
Scheme of CadC dependent pH regulation[3]: To visualize pH regulation, cells were cultivated in lysine decarboxylase indicator medium (indicator: bromothymol blue) and alkalization is depicted as a color change from yellow over green to blue (n=1). Production of EF-P and EfpL was confirmed by immunodetection of the C-terminally attached His6-tag using α-$His_6$ antibody. Source data are provided as a Source Data file.



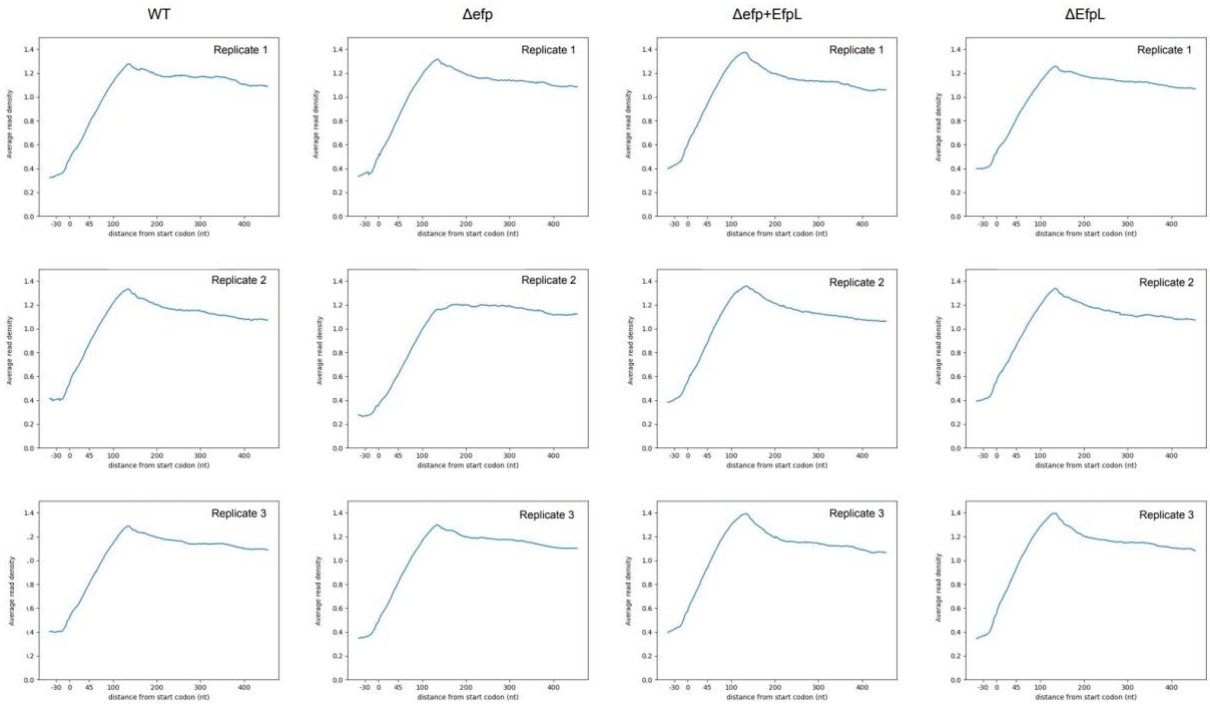

**Supplementary Fig. 9: Growth analysis of *E. coli* wild type and mutant strains on plate**

*E. coli* BW25113 wild type and mutants were grown over night in liquid media and spotted in different dilutions on LB plates containing **(A)** 40 mM glucose or **(B)** 40 mM arabinose, respectively. Expression status of complementation of Δ*uup*Δ*efp*Δ*efpL* with EF-P or EfpL on an arabinose inducible promoter was checked via α-$His_6$ antibody on a dot blot. **(A&B)** Source data are provided as a Source Data file

**A**



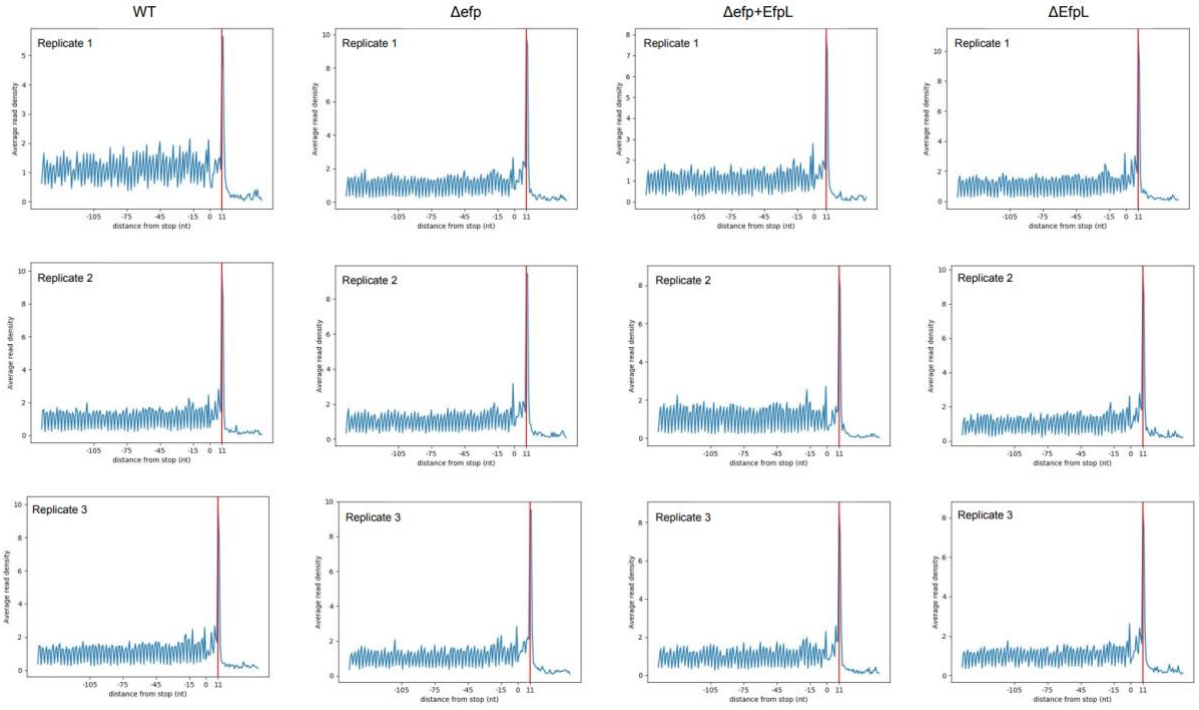

**B**



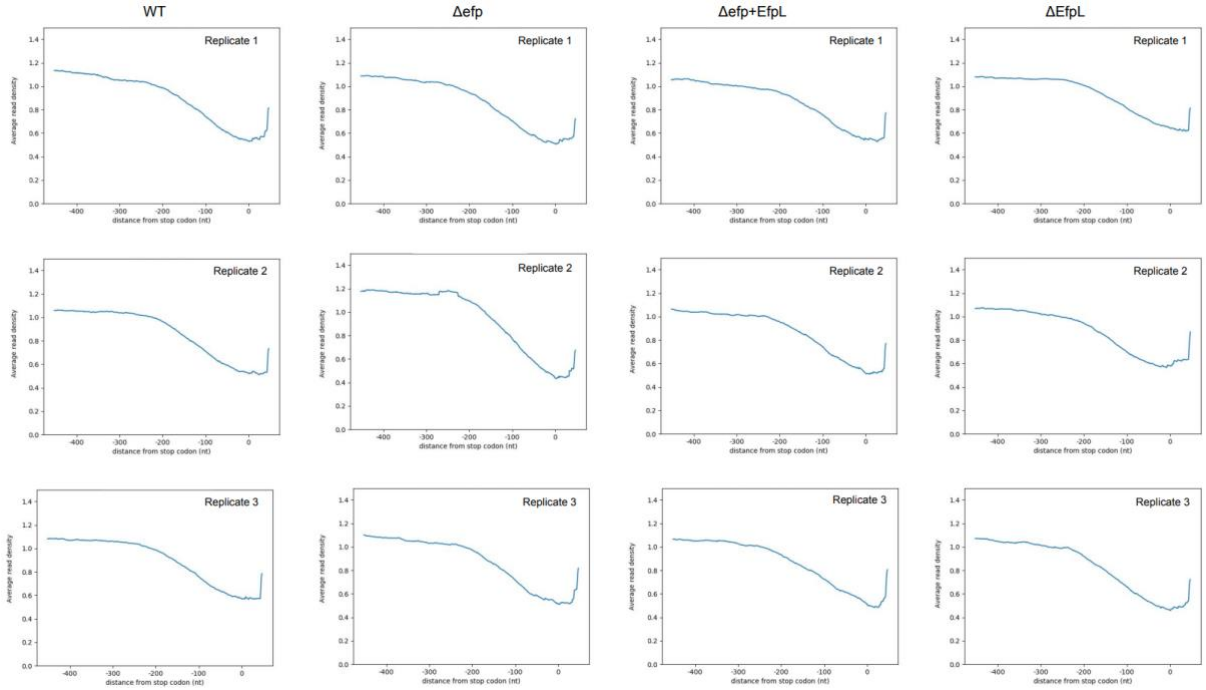

**Supplementary Fig. 10: Metagene plots in the region of initiation**

Metagene plots for Ribo-seq and RNA-seq samples of the wildtype (WT), *efp* deletion mutant (Δ*efp*), *efp* deletion mutant with overexpression of *efpL* (Δ*efp*+EfpL), and *efpL* deletion mutant (Δ*efpL*). The x-axis shows **(A)** the distance from the footprint or **(B)** RNA-seq read to the start or stop codons; the y-axis represents the average read density of the position. In the metagene plots for Ribo-seq samples (A), the 3'-end of the read was used to indicate the location of each footprint.

**A**

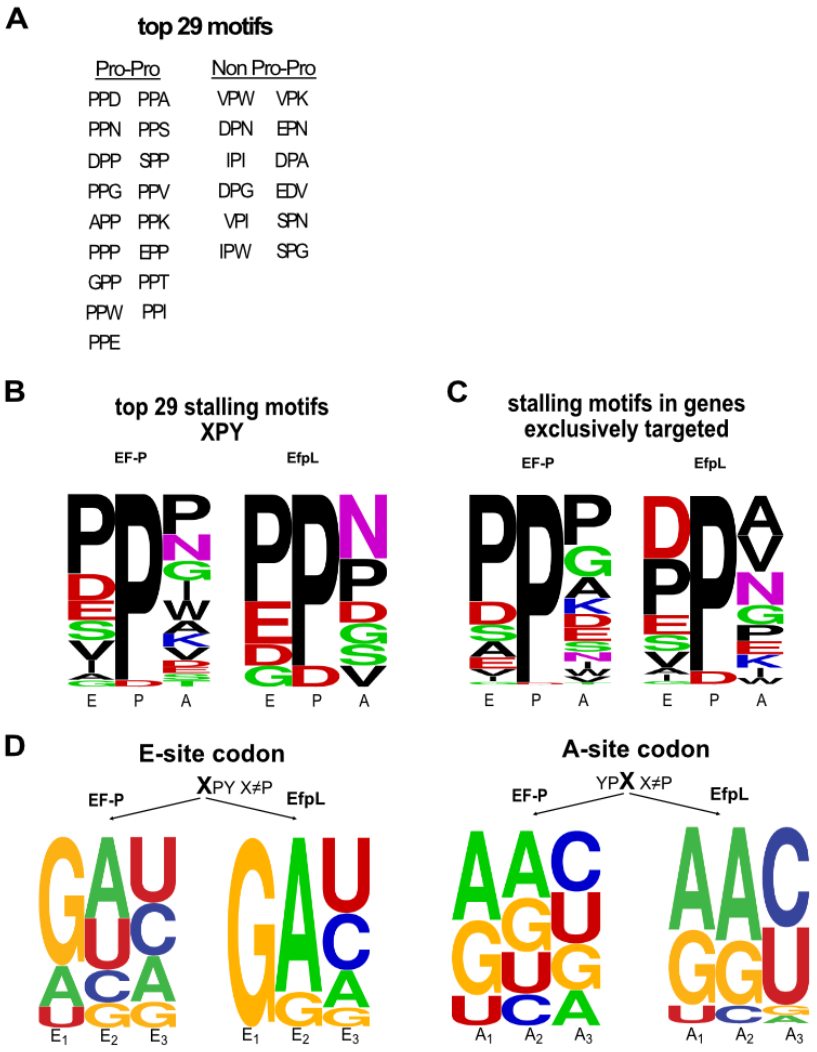



**B**



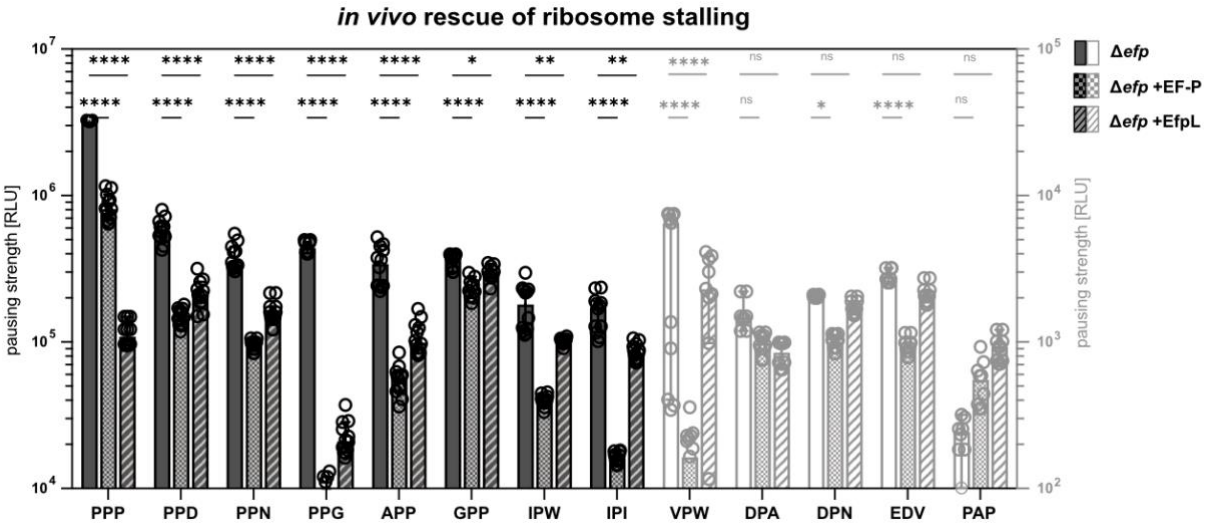

**Supplementary Fig. 11: Metagene plots in the region of termination**

Metagene plots for Ribo-seq and RNA-seq samples of the wildtype (WT), *efp* deletion mutant (Δ*efp*), *efp* deletion mutant with overexpression of *efpL* (Δ*efp*+EfpL), and *efpL* deletion mutant (Δ*efpL*). The x-axis shows **(A)** the distance from the footprint or **(B)** RNA-seq read to the start or stop codons; the y-axis represents the average read density of the position. In the metagene plots for Ribo-seq samples (A), the 3'-end of the read was used to indicate the location of each footprint.



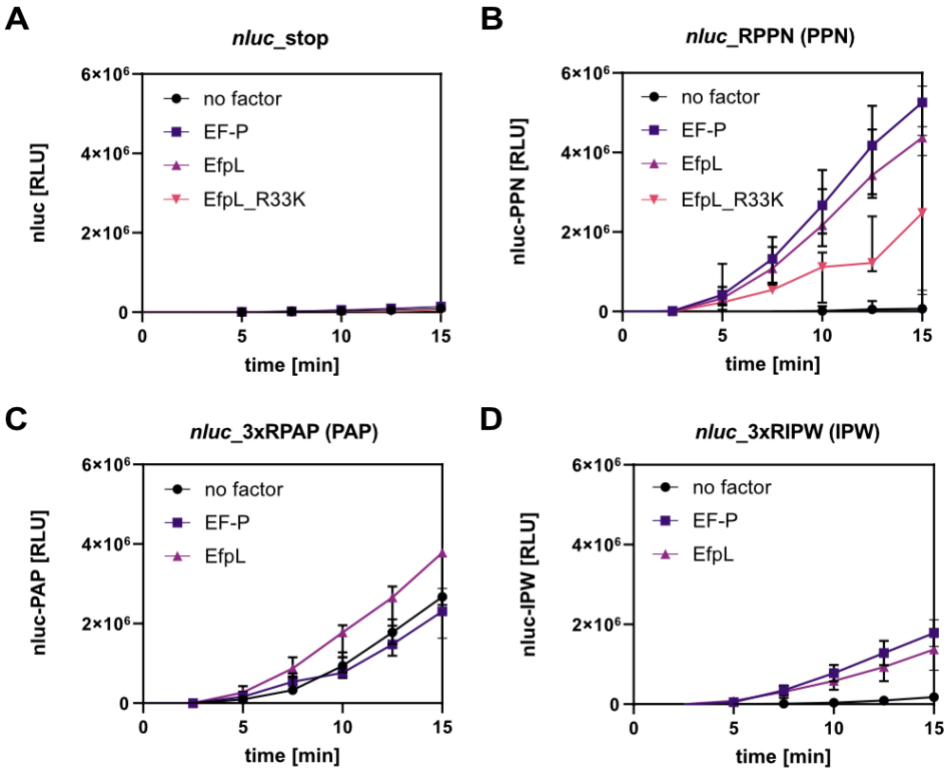

**Supplementary Fig. 12: Top 29 motifs in *E. coli* and comparison of codon bias of the E-site and A-site amino acid**

**(A)** Top 29 EF-P dependent arrest motifs associated with ribosome pausing in *E. coli* BW25113 determined by PausePred[36]. **(B)** Sequence logo[26] of the top 29 EF-P dependent arrest motifs. **(C)** Sequence logo[26] of the arrest motifs in genes targeted by exclusively EF-P or EfpL. **(D)** Sequence logos of the E- and A-site codons in XPY or YPX arrest motifs X≠P targeted by EF-P and EfpL, respectively.



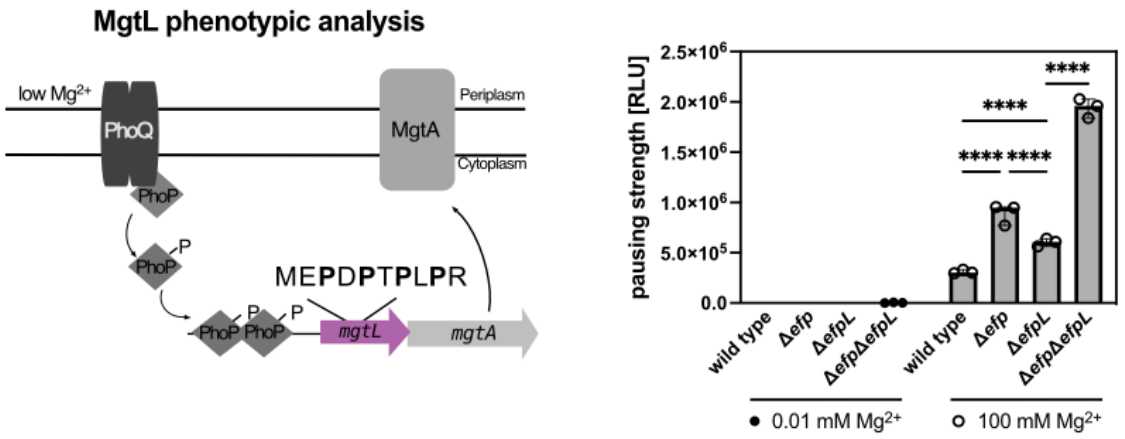

**Supplementary Fig. 13: *In vivo* detection of pausing strength at different motifs**

*In vivo* comparison of stalling strength of a set of stalling motifs and negative control PAP of *E. coli* Δ*efp* cells and respective trans complementation with EF-P (Δ*efp* +EF-P) and EfpL (Δ*efp* +EfpL). Pausing strength correlates with light emission and is given in relative light units (RLU) (n = 12, biological replicates, mean with sd indicated as error bars). Statistically significant differences according to 2-way ANOVA test with multiple comparison (*P value <0.0332, **P value <0.0021, ***P value <0.0002, ****P value <0.0001, ns not significant). Source data are provided as a Source Data file



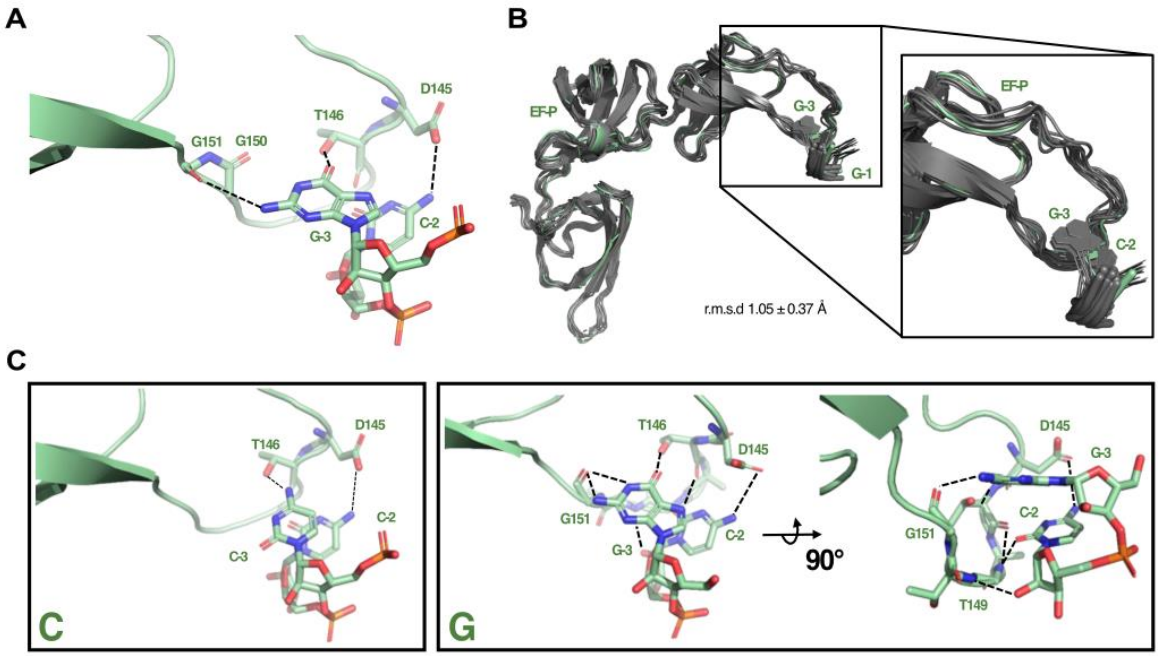

**Supplementary Fig. 14: Comparison of EF-P and EfpL of *E. coli* in translating different motifs**

*In vitro* transcription and translation of the *nLuc®* variants **(A)** nLuc_stop, **(B)** nLuc_RPPN (PPN), **(C)** nLuc_3xRIPW (IPW) or **(D)** nLuc_3xRPAP (PAP). The absence (no factor) or presence of the respective translation elongation factors of *E. coli* (EF-P, EfpL) is shown. Translational output was determined by measuring bioluminescence in a time course of 15 minutes (RLU) ($n \geq 3$, technical replicates, mean with sd indicated as error bars). **(A-D)** Source data are provided as a Source Data file



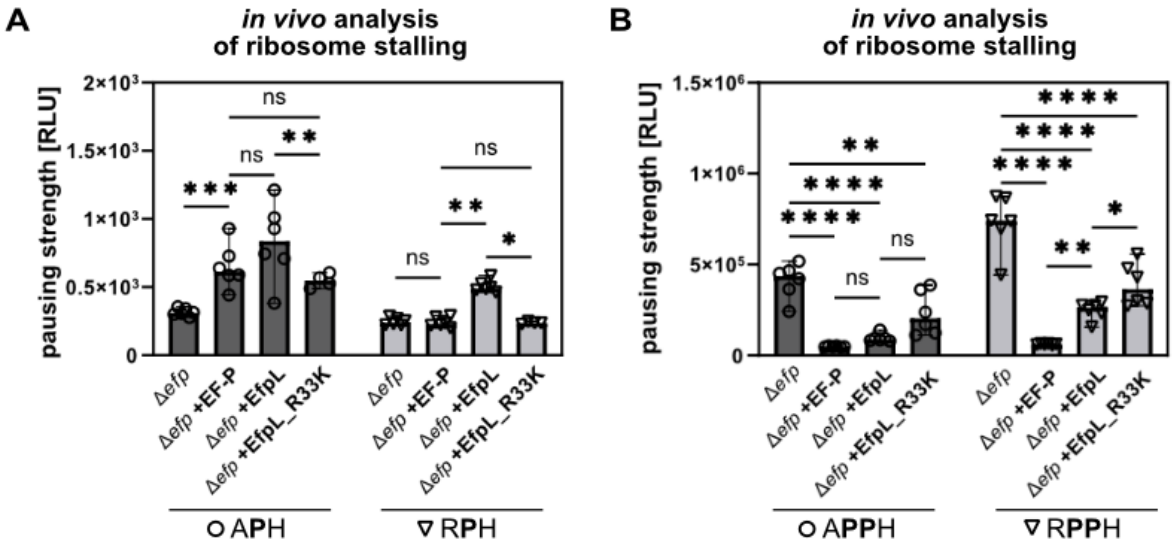

**Supplementary Fig. 15: MgtL phenotypic analysis**

Left: model illustrating the regulation mechanism of $Mg^{2+}$ uptake by MgtA [4-6]. *mgtL* consists of a proline-rich sequence and regulates *mgtA* expression. Right: reporter assay to detect pausing strength at the MgtL leader peptide with the sequence MEPDPTPLPR. Maximal luminescence emission under high (100 mM) and low (100 µM) $Mg^{2+}$ in *E. coli* BW23113 and corresponding mutant strains is depicted. Pausing strength correlates with light emission and is given in relative light units (RLU). (n = 3, biological replicates, mean with sd indicated as error bars). Statistically significant differences according to two-way ANOVA (*P value <0.0332, **P value <0.0021, ***P value <0.0002, ****P value <0.0001, ns not significant). Source data are provided as a Source Data file.



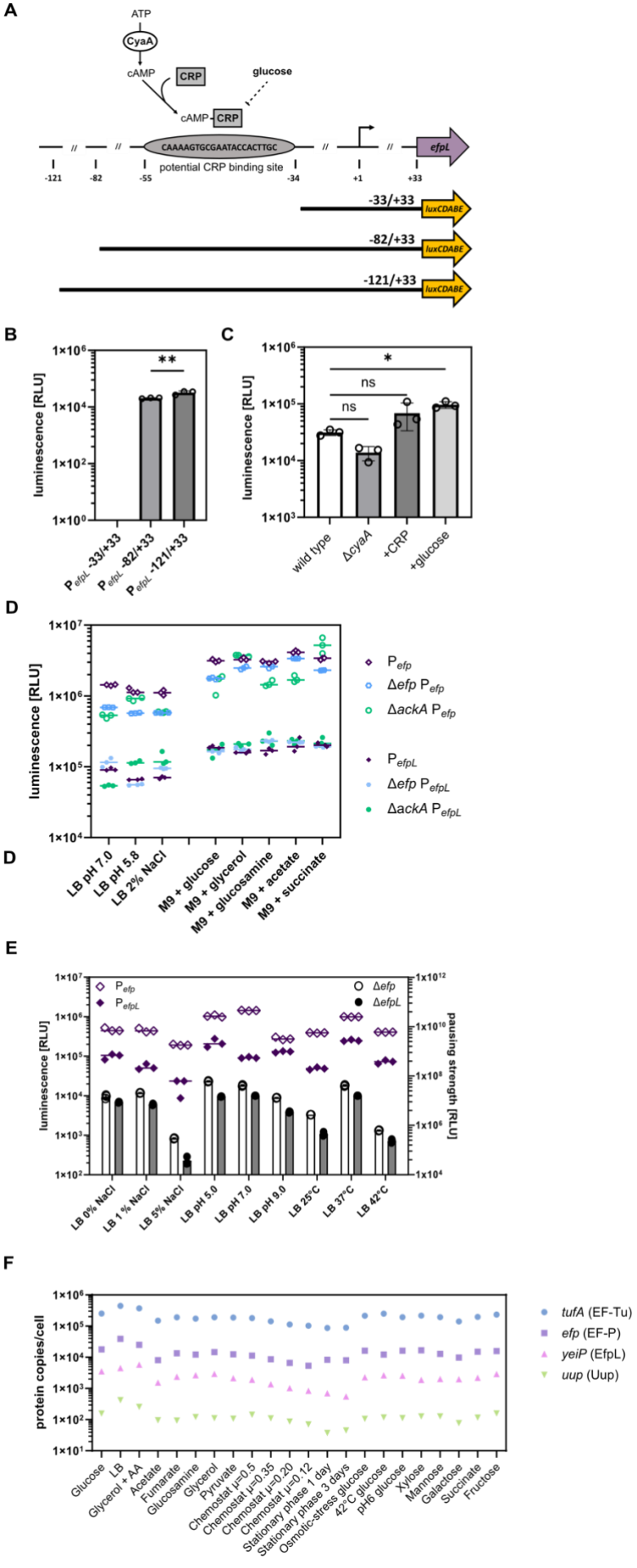

**Supplementary Fig. 16: EF-P interaction modelling with E-site codon**

**(A)** Close-up view of EF-P in contact with the E-site codon $_{-3}$GCG$_{-1}$, after *in silico* replacement of $_{-3}$C in the PDB entry 6ENU[7]. Polar contacts to EF-P OB domain 3 loop 1 residues are depicted with broken lines as obtained from the program PyMol (Delano Scientific). Note that only the first two nucleotides are shown for clarity. **(B)** Full-view superimposition of the 10 best solutions obtained from a HADDOCK run of EF-P together with $_{-3}$GCG$_{-1}$. The green model represents the non-docked and non-energy-minimized reference from panel A. The r.m.s.d. is given. The boxed view shows a zoom-in to the EF-P-RNA trinucleotide interface. **(C)** The same as shown in main text Fig. 3, but with an additional perspective depicted for EF-P in complex with GCG to highlight additional contacts.



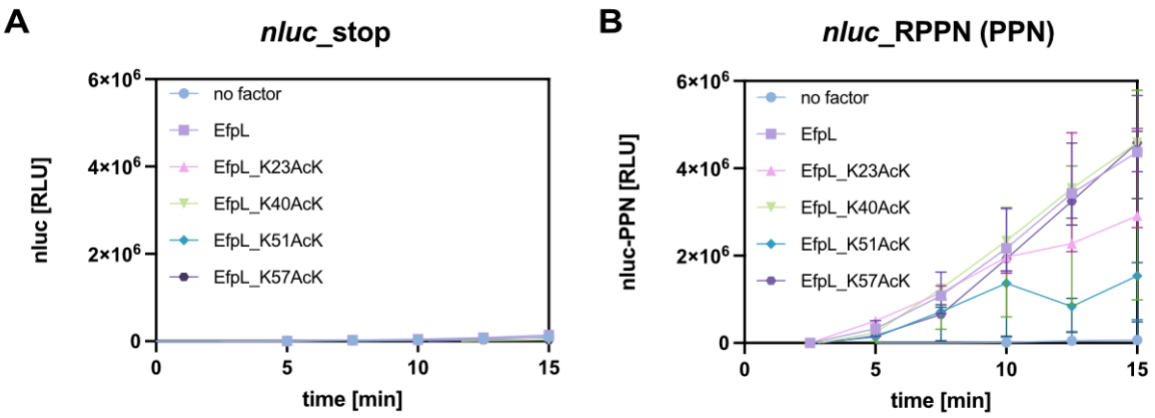

**Supplementary Fig. 17: EF-P and EfpL can induce ribosome stalling**

*In vivo* comparison of stalling strength of a **(A)** APH and RPH or **(B)** APPH and RPPH motif in Δ*efp* strains in the absence or presence of *efp* (+EF-P), or *efpL* (+EfpL). Pausing strength correlates with light emission and is given in relative light units (RLU). (n ≥ 4, biological replicates, mean with sd indicated as error bars). Statistically significant differences according to two-way ANOVA test with multiple comparisons (*P value <0.0332, **P value <0.0021, ***P value <0.0002, ****P value <0.0001, ns not significant). **(A&B)** Source data are provided as a Source Data file


A
ATP
CyaA
cAMP
CRP
glucose
cAMP-CRP
CAAAAGTGCGAATACCACTTGC
potential CRP binding site
efpL
-121
-82
-55
-34
+1
+33
-33/+33
-82/+33
-121/+33
luxCDABE
B
luminescence [RLU]
**
P efpL -33/+33
P efpL -82/+33
P efpL -121/+33
C
luminescence [RLU]
*
ns
ns
wild type
ΔcyaA
+CRP
+glucose
D
luminescence [RLU]
P efp
Δefp P efp
ΔackA P efp
P efpL
Δefp P efpL
ΔackA P efpL
LB pH 7.0
LB pH 5.8
LB 2% NaCl
M9 + glucose
M9 + glycerol
M9 + glucosamine
M9 + acetate
M9 + succinate
E
luminescence [RLU]
pausing strength [RLU]
P efp
P efpL
Δefp
ΔefpL
LB 0% NaCl
LB 1% NaCl
LB 5% NaCl
LB pH 5.0
LB pH 7.0
LB pH 9.0
LB 25°C
LB 37°C
LB 42°C
F
protein copies/cell
tufA (EF-Tu)
efp (EF-P)
yeiP (EfpL)
uup (Uup)
Glucose
LB
Glycerol + AA
Acetate
Fumarate
Glucosamine
Glycerol
Pyruvate
Chemostat μ=0.5
Chemostat μ=0.35
Chemostat μ=0.20
Chemostat μ=0.12
Stationary phase 1 day
Stationary phase 3 days
Osmotic-stress glucose
42°C glucose
pH6 glucose
Xylose
Mannose
Galactose
Succinate
Fructose

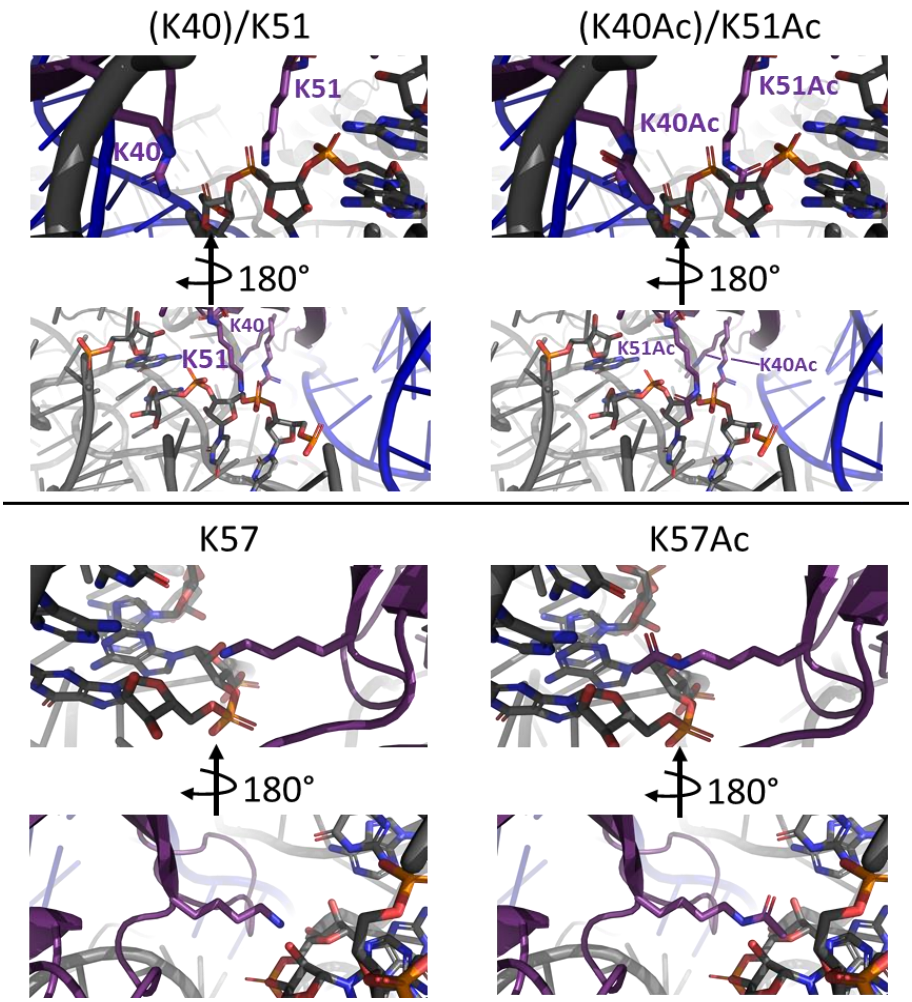
**Supplementary Fig. 18: *efp* and *efpL* gene expression analysis**

**(A)** Different lengths of the promoter region or *efpL* ($P_{efpL}$) were fused with the *luxCDABE* genes of *P. luminescens*. A schematic representation of the catabolite repression with cAMP-CRP shows inhibitory effects of glucose and activating conditions of CyaA for the potential CRP binding site in the promotor region of *efpL*. **(B)** *E. coli* BW25113 wild type strains were transformed with the different promoter reporter fusions. Maximal luminescence of a 16h time course in LB medium is given in relative light units (RLU) (n=3, biological replicates, mean with sd indicated as error bars). Statistically significant differences according to ordinary one-way ANOVA test (*P value <0.0332, **P value <0.0021, ***P value <0.0002, ****P value <0.0001, ns not significant). **(C)** *E. coli* BW25113 strains were transformed with the full-length promoter reporter fusion and tested in LB and LB supplemented with 20 mM CRP or 40 mM glucose. *E. coli* BW25113 Δ*cyaA* strain was transformed with the full-length promoter reporter fusion and tested in LB. Maximal luminescence of a 16h time course in LB medium is given in relative light units (RLU) (n=3, biological replicates, mean with sd indicated as error bars). Statistics as in (B). **(D)** *E. coli* wild type, Δ*efp* or Δ*ackA* strains were transformed with full-length promoter reporter fusion of *efpL* ($P_{efpL}$) or *efp* ($P_{efp}$). Maximal luminescence of a 16h time course in LB or M9 minimal medium supplemented with different carbon sources (20 mM) is given in relative light units (RLU) (n=3, biological replicates, line identifies mean value). **(E)** *E. coli* BW25113 strains were transformed with the full-length promoter reporter fusions of *efpL* ($P_{efpL}$) or *efp* ($P_{efp}$) and tested in LB and different conditions. Maximal luminescence of a 16h time course in LB medium is given in relative light units (RLU, left y-axis) (n=3, biological replicates, line identifies mean value, purple characters). *In vivo* comparison of stalling strength of Δ*efp* and Δ*efpL* strains for a PPN motif. Pausing strength correlates with light emission and is given in relative light units (RLU, right y-axis) (n=3, biological replicates, mean with sd indicated as error bars, bar chart). **(F)** Quantitative proteome analysis at different conditions for *tufA* (EF-Tu), *efp* (EF-P), *yeiP* (EfpL) and *uup* (Uup) analysed by Schmidt *et al.*[8]. **(A-F)** Source data are provided as a Source Data file



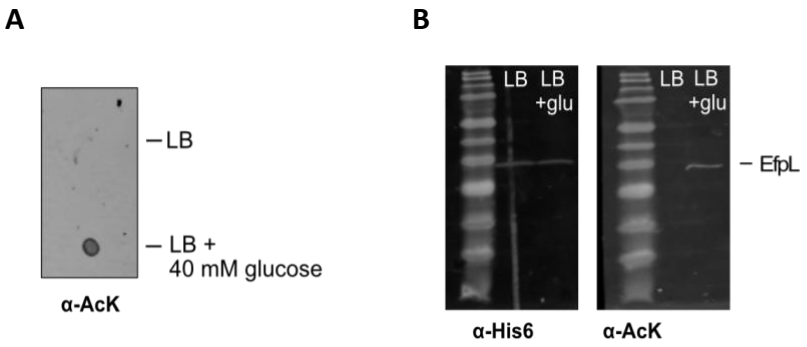

**Supplementary Fig. 19: EfpL function dependent on the acylation status**

*In vitro* transcription and translation of the *nLuc®* variants **(A)** nLuc_stop (no motif) or **(B)** nLuc_RPPN (PPN). The absence (no factor) or presence of the respective translation elongation factors of *E. coli* EfpL as well as the corresponding substitution variants EfpL_K23AcK, EfpL_K40AcK, EfpL_51AcK, EfpL_K57AcK is shown. Translational output was determined by measuring bioluminescence in a time course of 15 minutes and is given in relative light units (RLU) ($n \geq 3$, mean with sd indicated as error bars). **(A&B)** Source data are provided as a Source Data file




**Supplementary Fig. 20: Modelling of Acylation in EfpL**

Close-up views on EfpL lysines as shown in unmodified form (left panels) and when acetylated (right panels). Each view is shown from two different perspectives by 180° rotation as indicated. The N-terminal KOW domain (violet-purple) has been aligned to the PDB entry 6ENU[7] to enable monitoring of potential clashes and interactions with ribosomal components. Lysine sidechains are shown as sticks on an otherwise cartoon-typed presentation. For the K40/K51 region (upper panels), R42 is additionally shown to indicate the dense space, relevant in potential sidechain modifications. Relevant RNA regions in close vicinity of lysines are shown as sticks. Grey represents ribosomal RNA, blue indicates tRNA.

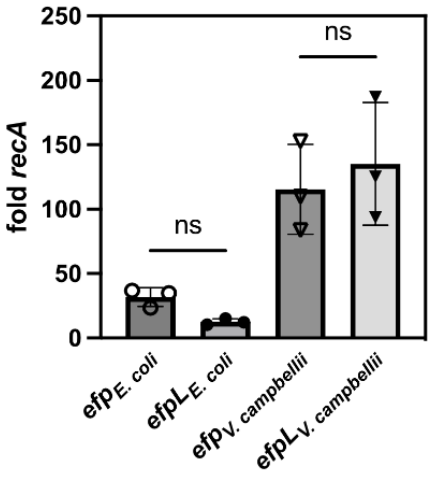



**Supplementary Fig. 21: *In vivo* acetylation of EfpL**

**(A)** Immunodetection of lysine acetylation status in BW25113 *efpL*_His grown in LB or LB supplemented with 40 mM glucose. **(B)** Immunodetection of EfpL purified from BW25113 *efpL*_His grown in LB or LB supplemented with 40 mM glucose. **(A&B)** Source data are provided as a Source Data file

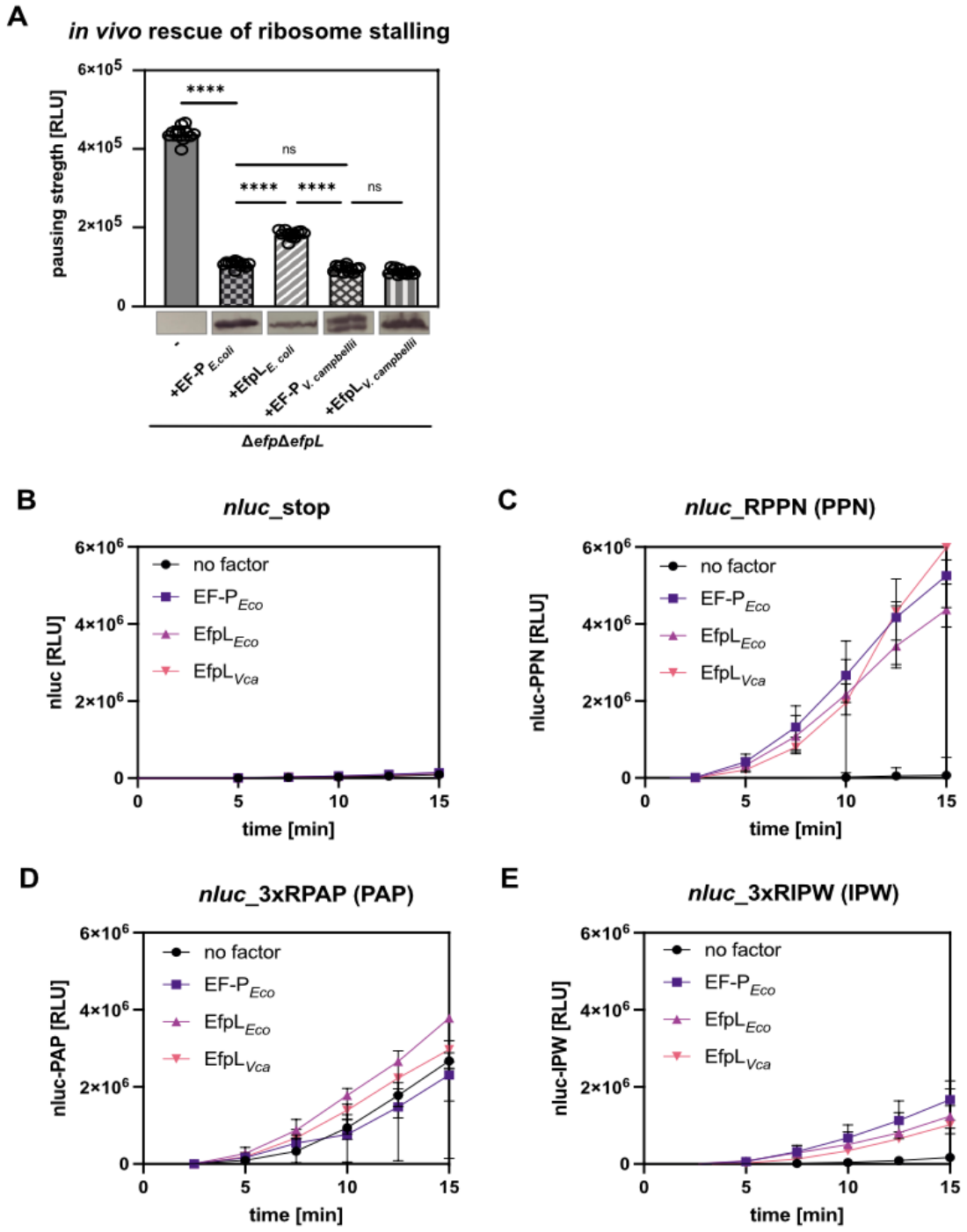

**Supplementary Fig. 22: Conservation status of acylation sites**

Sequence logos[26] for amino acids at positions 20-60 in all EfpLs, or EfpL from Enterobacteriales or Vibrionales.



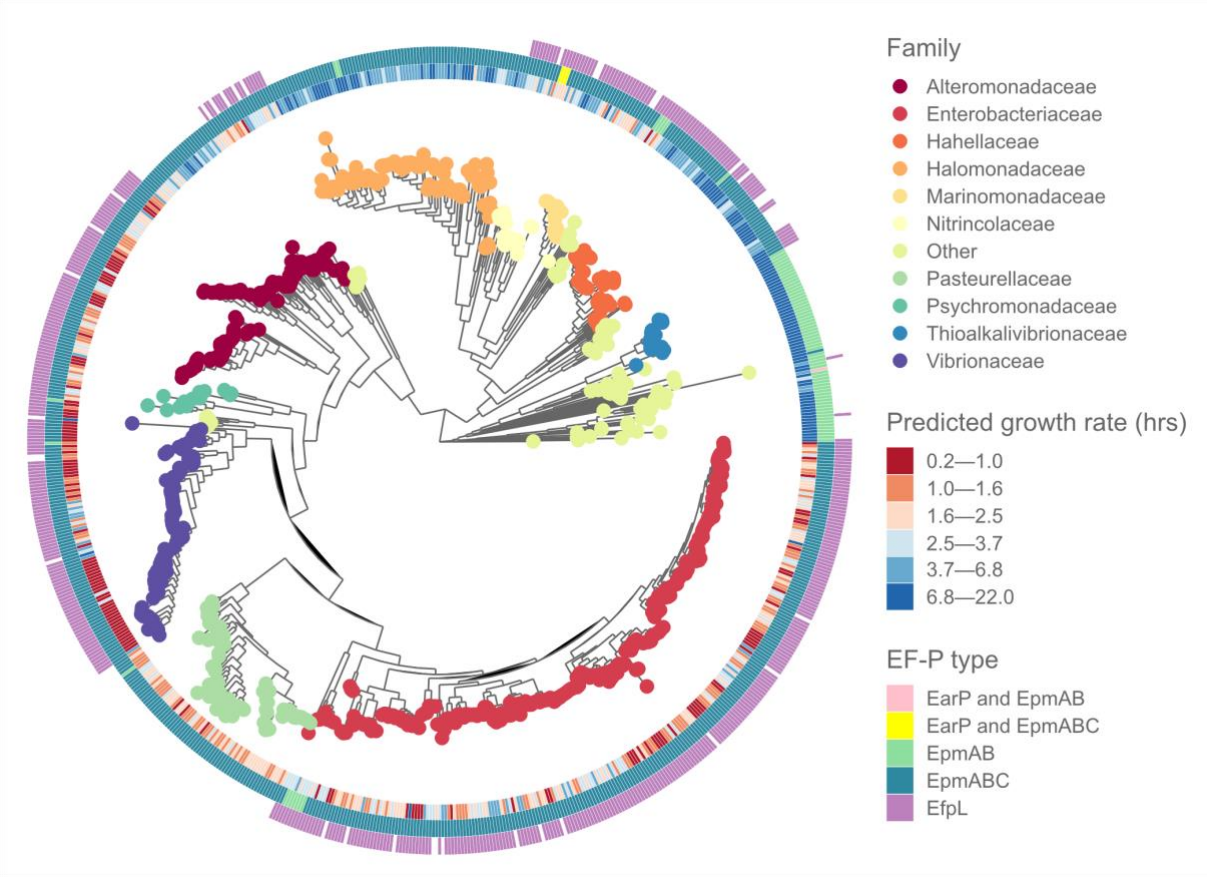

**Supplementary Fig. 23: Expression analysis of EF-P and *EfpL* from *E. coli* and *V. campbellii***

Quantitative real time PCR (qRT-PCRs) were performed to analyze expression of *efp* and *efpL* in *E. coli* or *V. campbellii.* (n=3, biological replicates, mean with sd indicated as error bars). Statistically significant differences according to ordinary one-way ANOVA test (*P value <0.0332, **P value <0.0021, ***P value <0.0002, ****P value <0.0001, ns not significant). Primer efficiency was as following: $recA_{E.\ coli}$ 1.987, $efp_{E.\ coli}$ 1.953, $efpL_{E.\ coli}$ 1.936, $recA_{V.\ campbellii}$ 2.084, $efp_{V.\ campbellii}$ 1.962, $efpL_{V.\ campbellii}$ 2.009. Normalization with reference gene *recA* for comparison of *efp* and *efpL* expression. Source data are provided as a Source Data file




**Supplementary Fig. 24: Functional comparison of EfpL from *E. coli* and *V. campbellii***

**(A)** *In vivo* comparison of stalling strength of a PPN motif of *E. coli* Δ*efp*Δ*efpL* cells and respective trans complementation with *E. coli* EF-P (+EF-$P_{Eco}$) and EfpL (+Efp$L_{Eco}$), as well as *V. campbellii* EF-P (+EF-$P_{Vca}$) and EfpL (+Efp$L_{Vca}$). Production of EF-P and EfpL was confirmed by immunodetection of the C-terminally attached His6-tag using α-$His_6$ antibody. Stalling strength correlates with light emission and is given in relative light units (RLU). (n = 12, biological replicates mean with sd indicated as error bars). Statistically significant differences according to ordinary one-way ANOVA test (*P value <0.0332, **P value <0.0021, ***P value <0.0002, ****P value <0.0001, ns not significant). **(B-E)** *In vitro* transcription and translation of the *nLuc®* variants **(B)** nLuc_stop (no motif), **(C)** nLuc_RPPN (PPN), **(D)** nLuc_3xRIPW (IPW) or **(E)** nLuc_3xRPAP (PAP). The absence (no factor) or presence of the respective translation elongation factors of *E. coli* (EF-$P_{Eco}$, Efp$L_{Eco}$) or *V. campbellii* (Efp$L_{Vca}$) is shown. Translational output was determined by measuring bioluminescence in a time course of 15 minutes ($n \geq 3$, technical replicates, mean with sd indicated as error bars). **(A-E)** Source data are provided as a Source Data file




**Supplementary Fig. 25: Phylogenetic analysis of predicted growth rates**

Set of 920 genomes from the class γ-proteobacteria from the Integrated Microbial Genomes (IMG) database[114]. Inner ring shows doubling times predicted from codon usage bias in ribosomal genes, middle two rings show EF-P types. Colors of tip ends depict phylogenetic family.

**Supplementary References**


1 Consortium, T. U. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523-D531 (2022). https://doi.org:10.1093/nar/gkac1052

2 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682 (2012). https://doi.org:10.1038/nmeth.2019

3 Ude, S. *et al.* Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* **339**, 82-85 (2013). https://doi.org:10.1126/science.1228985

4 Gall, A. R. *et al.* Mg2+ regulates transcription of mgtA in *Salmonella Typhimurium* via translation of proline codons during synthesis of the MgtL peptide. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 15096-15101 (2016). https://doi.org:10.1073/pnas.1612268113

5 Nam, D., Choi, E., Shin, D. & Lee, E. J. $tRNA^{Pro}$-mediated downregulation of elongation factor P is required for mgtCBR expression during *Salmonella* infection. *Mol. Microbiol.* **102**, 221-232 (2016). https://doi.org:10.1111/mmi.13454

6 Takada, H., Fujiwara, K., Atkinson, G. C., Chiba, S. & Hauryliuk, V. Resolution of ribosomal stalling by ABCF ATPases YfmR and YkpA/YbiT. *Nucleic Acids Res.* (2024). https://doi.org/10.1093/nar/gkae556

7 Huter, P. *et al.* Structural Basis for Polyproline-Mediated Ribosome Stalling and Rescue by the Translation Elongation Factor EF-P. *Mol. Cell* **68**, 515-527.e516 (2017). https://doi.org:10.1016/j.molcel.2017.10.014

8 Schmidt, A. *et al.* The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* **34**, 104-110 (2016). https://doi.org:10.1038/nbt.3418

# 4 Concluding discussion and outlook

The translation process is an intricate system involving various players that either facilitate or impede ribosomal progression. Proline pairs are notoriously challenging to translate, often causing ribosome stalling[50-52]. Despite the associated translational burden, their frequent occurrence in nature suggests a selective advantage that outweighs these challenges[17]. This advantage has even driven the evolution of specialized factors, which assist in the efficient translation of polyproline sequences. The unique structural and functional properties of polyprolines contribute to this evolutionary benefit, as they play critical roles in protein architecture and function[14,15,17,55,56,58]. The structural and biochemical data presented here contribute to a better understanding of ribosome stalling and its alleviation by distinct factors.

## 4.1 The nature of stalling motifs

The efficiency of proline incorporation into the nascent peptide chain is not solely determined by the presence of certain polyproline motifs. It is also influenced by the specific codons used to encode proline and the availability of cognate tRNAs **(Chapter 2)**. In *E. coli*, proline is encoded by four codons (CCA, CCC, CCG, and CCU) **(Fig. 2)**, each of which is recognized by distinct tRNAs with varying cellular abundances[128]. Bioinformatic analyses have revealed different codon usage between single and consecutive prolines **(Fig. 6A)**, namely a selective pressure against the use of certain proline codon pairs, such as CCC/U-CCC/U, which are prone to inducing frameshifting[129,130] and translational delays[50]. In contrast, codons such as CCG are overrepresented in polyproline motifs due to their association with faster translation rates and higher levels of the corresponding tRNAs[131]. Both overexpression and deletion of the tRNAs *proK*, *proL*, and *proM* affected translation efficiency at their respective codons **(Chapter 2)**. The most substantial effect was observed with *proL* on the CCC codon, likely due to its GGG anticodon and slow translation rates at CCC and CCU. Overexpression of *proL* improved decoding of CCC, suggesting potential benefits for heterologous expression of GC-rich genes. In contrast, *proM* recruitment appeared rate-limiting for CCA decoding, but its overexpression had minimal effects on other codons due to transcriptional inefficiencies and weak codon-anticodon interactions. This codon-tRNA interaction plays a pivotal role in regulating the speed of translation, with certain codons leading to slower translation rates and an increased likelihood of ribosomal stalling.

Synonymous codons for a given amino acid are not used equally, even though they are functionally equivalent in terms of protein structure[132] **(Fig. 2)**. The theory of codon bias suggests a correlation between codons and their corresponding isoaccepting tRNAs, where favorable codons correspond to the most abundant and efficient cognate tRNA[131,133,134]. While the tRNA abundance is relevant to proline **(Chapter 2)**, incorporation velocity appears to be

the more decisive factor. The CCC codon is rarely used in *E. coli*, because it exhibits the lowest efficiency in interacting with the prolyl-tRNA–EF-Tu–GTP ternary complex, resulting in slower recruitment during translation[50]. The selective codon usage for proline not only helps to optimize translation efficiency but also serves as a regulatory mechanism to modulate protein copy number in response to cellular demands **(Chapter 2)**. Proline-rich sequences are often strategically localized within proteins, frequently occupying key regulatory regions or domains where translational pausing provides functional advantages[17]. Notably, an accumulation of polyproline motifs can be found within the N-terminal 50 residues of the coding regions[17], where varying codon usage allows for fine-tuning of translation, to enhance precise regulation of protein synthesis. Ribosomes can initiate translation on mRNA at intervals of 1–3 seconds[135,136], and a low translational ramp at the start of the ORF may act as a delayed phase of translation initiation. This mechanism effectively reduces ribosomal traffic congestion and lowers the energy cost of protein biosynthesis[137-139]. Under optimal growth conditions, *E. coli* ribosomes translate at a rate of 42-51 nucleotides per second, while the leading RNA polymerase transcribes at a similar rate of 42-49 nucleotides per second[140]. Transcription and translation are tightly coupled in many bacteria, with the ribosome appearing to modulate the pace of the RNA polymerase[140]. However, a controlled uncoupling of transcription and translation can also be utilized for gene regulation, as seen for the biosynthesis of tryptophan[141] or histidine biosynthesis[51]. In these cases, stalled ribosomes and prolonged translation promote the formation of different mRNA secondary structures, which regulate the expression of the downstream genes[51,141].

This attenuation mechanism in the histidine biosynthesis operon has been further utilized to develop an *in vivo* system for quantifying translational pausing and assessing the effects of codon choice and tRNA abundance on translation efficiency **(Chapter 2)**. The system enables the measurement of ribosome stalling in living cells and represents a powerful approach to evaluating the translational challenges posed by specific sequences. Ribosome pausing induced by strong stallers was comparable in severity to a complete stop caused by a termination signal, highlighting the significant impact these sequences have on translation. This phenomenon was studied in a wild-type *E. coli* strain, where EF-P alleviates stalling at consecutive prolines, providing a native context to assess ribosome pausing[29]. These analyses demonstrated that the translational burden associated with PP motifs is not an anomaly but rather an inherent feature of the translation machinery. The difficulty in decoding these motifs imposes substantial selective pressure on the proteome, influencing codon usage and shaping protein composition. This intrinsic challenge explains why proline-rich motifs are rare in the genome and underscores the evolutionary adaptations, such as EF-P, that have evolved to mitigate these obstacles.

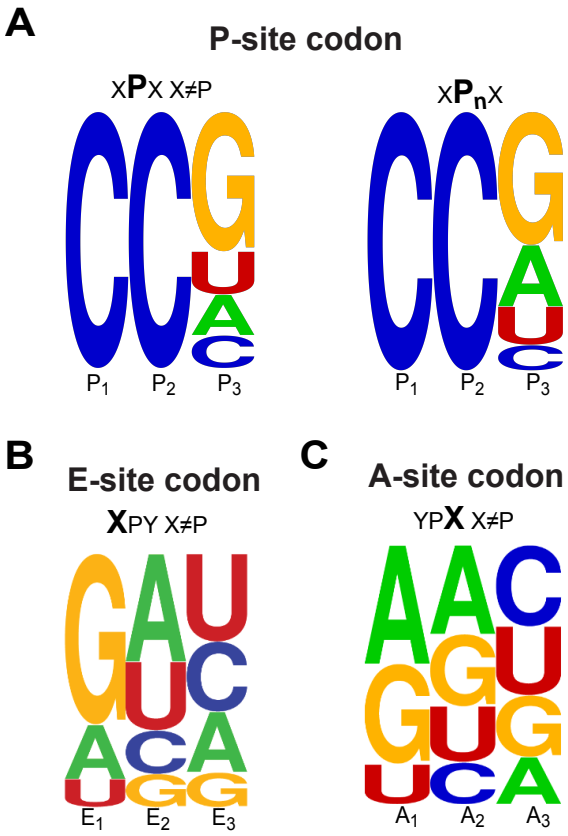



**Figure 6: Codon usage in proline-containing motifs**

Sequence logos[142] of the *E. coli* codon usage in proline-containing motifs. With proline in P-site, with X can be all amino acids, but proline, and with Y can be all amino acids, including proline. **(A)** P-site codon of non-consecutive proline in XPX (left) and P-site codon of polyproline in $XP_nX$ (right) (Chapter 2&3). **(B)** E-site codon in stalling motifs XPY targeted by EF-P, with guanidine overrepresentation at first position (Chapter 3). **(D)** A-site codon in YPX targeted by EF-P (Chapter 3).

### 4.1.1 EF-P dependent stalling motifs

EF-P promotes translation by interacting with several components of the stalled ribosome, including the 23S rRNA near the peptidyl transferase center[70], specific nucleotides of the 16S rRNA, and the anticodon stem of the P-site tRNA[70,71]. Notably, EF-P's retention time on the ribosome is modulated by the strength of the stalling motif and its dissociation rate[143]. EF-P rapidly scans for available ribosomal E-sites and can bind to any ribosome containing a P-site tRNA, irrespective of the ribosome's functional state[143]. This flexible binding enables EF-P to engage with a range of ribosomes, positioning itself to address potential translation stalls efficiently. EF-P was found to bind to the ribosome during many or most elongation cycles, which exceeds the number of polyproline motifs in *E. coli*[144]. As already shown for eIF5A[74], our ribosome profiling analysis in *E. coli* revealed an extended ribosome stalling arrest sequence spectrum beyond the canonical polyproline motifs that EF-P resolves **(Fig. 7A) (Chapter 3)**. Other than the polyproline motifs, which show considerable overlap with IF5A motifs, EF-P appears to resolve different non-consecutive proline arrest motifs **(Fig. 7)**. For *S. cerevisiae*, the top 29 IF5A-dependent tripeptide stalling motifs revealed a consensus sequence enriched

for proline or aspartate in the E- and P-sites and proline in the A-site, with glycine showing modest enrichment across all three sites. Interestingly, despite the prominent role of proline in the consensus sequence, 18 of these 29 motifs do not contain a PP combination[74] **(Fig. 7B)**. For *E. coli*, we successfully reproduced the findings from previous ribosome profiling analyses for EF-P dependent PP stalling motifs and additionally extended the list of EF-P targets by motifs without consecutive prolines, comprising 12 of the top 29 motifs. Like the IF5A consensus sequence, proline and aspartate were found to be enriched in the E- and P-site, while proline and asparagine were prominent in the A-site **(Fig. 7A)**.

The amino acid context influences the ability of EF-P to alleviate ribosome stalling[145] **(Chapter 2&3)**. Our data indicate that EF-P's rescue efficiency is highest for weak stalling motifs and diminishes for stronger stalls. This suggests a hierarchal mechanism of rescue efficiency, where EF-P operates most optimally under conditions of mild stalling. Interestingly, EF-P was also observed to induce stalling at certain non-stalling motifs, presumably by blocking tRNA translocation to the E-site **(Chapter 3)**. This supports the idea that EF-P binds to every free E-site ribosome irrespective of the sequence but remains bound only when interacting with a prolyl-tRNA in the P-site[143]. The dual functionality highlights EF-P's complex role in translation. While binding can be advantageous for resolving stalling motifs, it may be detrimental at non-stalling sequences, which could unnecessarily slow down translation. Notably, not all proline-containing motifs rely on EF-P, emphasizing that binding and alleviation are context-dependent. Whether EF-P-induced ribosome pausing at non-stalling motifs serves a regulatory function remains an open question, suggesting a potential additional layer of translational control.

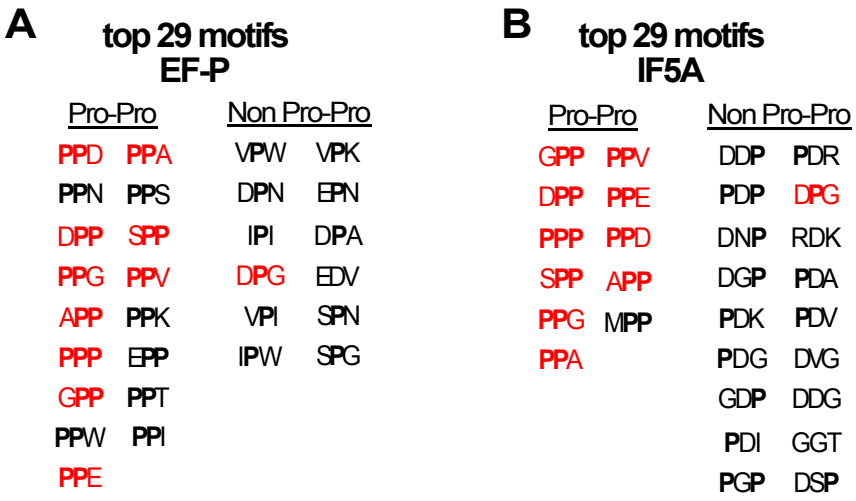



**Figure 7: Stalling motifs dependent on EF-P or IF5A**

Top 29 arrest motifs associated with ribosome pausing dependent on **(A)** EF-P in *E. coli* BW25113 determined by PausePred[146] (Chapter 3) or **(B)** eIF5A in *S. cerevisiae* cells with pause score greater than 10[74]. Red-marked motifs show EF-P and IF5A overlap.

### 4.1.2 The E-site codon in stalling motifs

Besides binding to the prolyl-tRNA and the ribosome, EF-P interacts with the mRNA, which varies according to the arrest motif[145] **(Chapter 3)**. When EF-P binds to the ribosome, it interacts with the first and second positions of the E-site codon via residues G144–G148 in the first loop of domain III, with D145 and T146 forming specific contacts[31]. The E-site codon has no effect on EF-P association but modulates its dissociation rate[143]. Our data suggest that codons with guanosine in first position (GNN) induce a particularly strong EF-P-dependent translational arrest in XP(P)X motifs, as we found predominantly aspartate (4x), glutamate (3x), valine (3x), or glycine (1x), if it is not proline (12x) in the first position of the strongest EF-P dependent stalling motifs **(Fig. 6&7)**. With guanosine, EF-P binding to the mRNA at the E-site codon could extend up to residue G151, potentially engaging the entire domain III loop **(Chapter 3)**. These additional interactions may affect EF-P dissociation from the ribosome and can therefore be a regulatory element of the ribosome pausing. The codon-specific interactions between EF-P and proline-rich sequences offer promising potential for synthetic biology, as it harbors the capability to predict and control ribosome stalling through codon choice. It was shown that translation can be tuned at the codon level, directly impacting translation efficiency and protein copy number in bacteria **(Chapter 2)**.

## 4.2 The novel translation factor EfpL

Most eukaryotes and some prokaryotes possess multiple isoforms of IF5A[30,147] or EF-P **(Chapter 3)**, highlighting the evolutionary importance of translational control and providing flexibility in managing ribosome stalling. In some bacteria, the EF-P paralog EfpL was found **(Chapter 3)**. Both factors share overlapping functions, particularly in resolving translation stalls at XPX motifs, ensuring the efficient synthesis of proteins under varying cellular conditions. This functional overlap likely provides a fail-safe mechanism, allowing one factor to compensate for the loss or reduced activity of the other. Like EF-P, EfpL contains a three-domain structure, with the C-terminal OB-domain playing a key role in ribosomal interactions and our high-resolution structure revealed that EfpL has a prolonged loop. EfpL was found predominantly in Proteobacteria of the γ-subdivision, but also in Acidobacteria, Thermodesulfobacteria, and the Planctomycetes-Verrucomicrobia-Chlamydiae-group (PVC-group), mainly together with the EpmA-type EF-P **(Fig. 8)**. Phylogenetic analysis linked EfpL most closely to the EF-P group activated by α-rhamnosylation, raising questions about the evolutionary origin of EfpL. We propose that an early form of EfpL emerged following gene duplication and sequence diversification, providing enhanced functionality for resolving XPX arrest motifs. Further evolutionary events, including the reduction of the loop and recruitment of the modifying enzyme EarP, refined the function of EF-P. However, our structural and functional analyses indicate that EfpL has diverged from EF-P in its specific function during translation. Through ribosome profiling, we have uncovered that EfpL alleviates stalling at a

subset of EF-P-dependent stalling motifs with an overrepresentation of DPA, PPV, and DPN, but excels in ribosome rescue at specific genes, including metabolite interconversion enzymes, transporters, metabolism proteins, transcriptional regulators and modifying proteins. In nutrient-rich conditions, EF-P appears critical for maintaining growth, while the role of EfpL becomes more apparent under specific nutrient contexts. When glucose was provided as the primary carbon source, population dynamics shifted, suggesting a nutrient-dependent function for EfpL. Translation serves as a dynamic process that enables cells to adapt to internal and external changes[148,149]. Comparative studies suggest that bacteria possessing both EF-P and EfpL tend to grow faster than those lacking EfpL **(Chapter 3)**. This suggests that EfpL provides a distinct selective advantage by enhancing translational efficiency and serves as a regulatory element in rapidly changing environments.

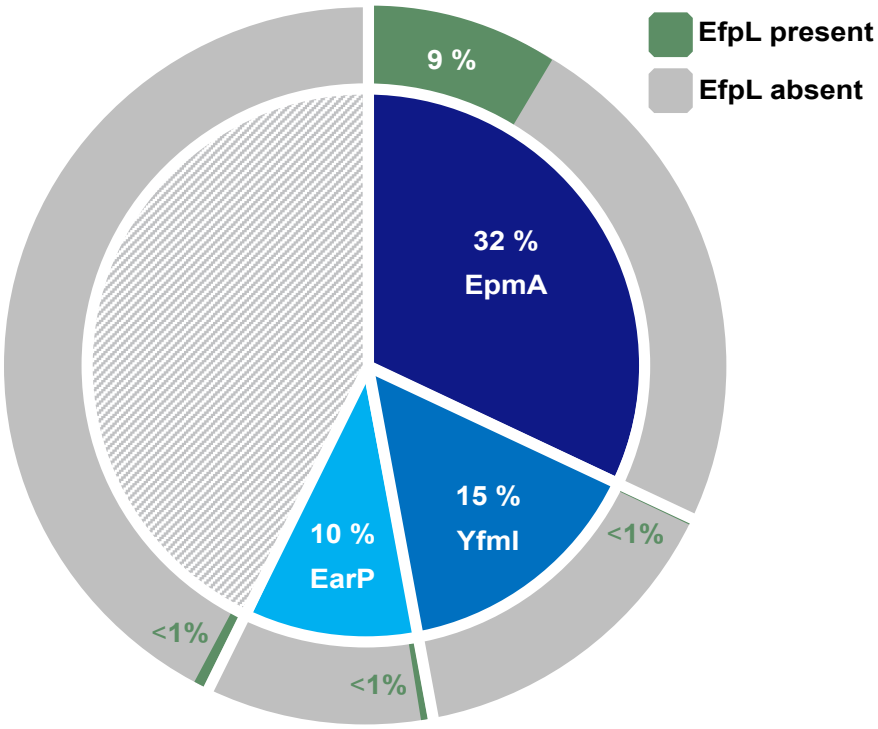



**Figure 8: Distribution of EfpL, EF-P, and modification enzymes**

Distribution of EF-P with corresponding modification enzymes and EfpL in bacteria of multiply aligned 5448 sequences of KOW-like domains of proteins with three domains typical for EF-P in a collection of 4736 complete bacterial genomes, obtained from the RefSeq database[25]. The inner chart represents EF-Ps with modification enzymes. Approximately 32 % of bacteria encoding for EpmA, indicating a modification of EF-P with (*R*)-β-lysine, while 15 % encoding YfmI, resulting in a modification of EF-P with 5-amino-pentanol. Another 10 % of bacteria have EarP that is capable to rhamnosylate EF-P. In around 42 % of sequenced genomes, none of the modification systems described were found. The outer ring represents the occurrence of EfpL in addition to EF-P with the respective modification enzymes. EfpL is mainly associated with the EpmA-type EF-P (Chapter 3).

### 4.2.1 The role of acylation in EfpL regulation

The activity of EfpL is modulated by post-translational lysine acylation, suggesting a potential role as a sensor of the metabolic state of the cell, linking translation efficiency with cellular metabolic conditions. Bacteria dynamically adapt their metabolism to changes in nutrient availability, growth conditions, and environmental stresses. Lysine acylation functions as a global regulatory mechanism, responding to the energy status of a cell[150] and targeting a wide array of metabolic enzymes to influence their activity[151-153]. In *E. coli*, the concentrations of acetyl-CoA levels vary between 200–600 µM and can be measured up to 610 µM in exponentially growing glucose-fed cultures[154,155]. The concentration of acetyl-CoA is closely tied to the metabolic state of the cells, peaking during the exponential growth phase and decreasing as the cells transition to the stationary phase[156].

With at least four lysine modification sites for acetylation, malonylation, and succinylation, EfpL is among the most heavily acylated proteins in *E. coli*[157-160] **(Fig. 9)**. Interestingly, eIF5A was found to undergo phosphorylation and acetylation, possibly influencing its subcellular localization[161-163]. Unlike eIF5A and EF-P, which are modified at the β3Ωβ4 loop within domain I to activate their function, EfpL lacks this modification. Instead, EfpL is negatively regulated via acylation. The high number of non-enzymatically modified lysine sites contributes to significant heterogeneity in EfpL populations. Our *in vivo* and *in vitro* analyses showed a different degree of deactivation when modifying different lysines **(Chapter 3)**. Furthermore, the regulation of EF-P and EfpL activity through post-translational modifications allows bacteria to fine-tune their translational machinery in response to metabolic changes. This dynamic regulation may enable bacteria to optimize protein synthesis under nutrient-limited conditions or during stress responses, further enhancing their adaptability and survival.

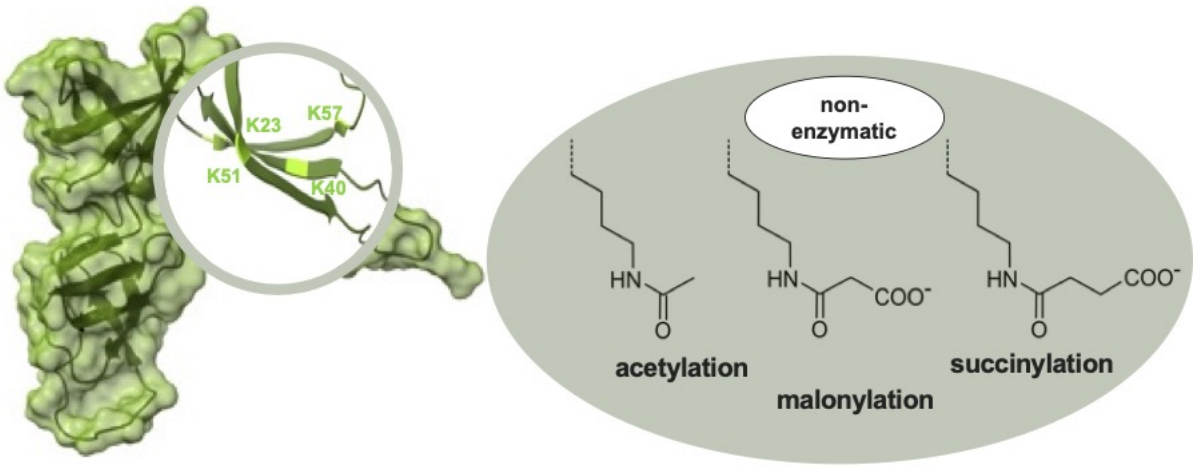



**Figure 9: EfpL acylation sites**

EfpL (*E. coli*, pdb: 8S8U) acylation sites at lysines K23, K40, K51 and K57 according to Weinert *et al.* (2013)[160], Weinert *et al.* (2013)[158], Kuhn *et al.* (2014)[157] and Qian *et al.* (2016)[159] and acylation types for non-enzymatic lysine modification at these positions are depicted.

### 4.2.2 EfpL in *Vibrio* species

In Enterobacteria, the four acylation sites identified in *E. coli* EfpL are largely conserved but exhibit less conservation in *Vibrio* species. Additionally, expression levels of EfpL in *Vibrio campbellii* are higher than in *E. coli*, and both *in vitro* as well as *in vivo* comparisons showed that *V. campbellii* EfpL has superior ribosomal rescue efficiency. Despite this enhanced efficiency, deleting *efp* or *efpL* alone in *V. campbellii* did not impact growth, but deleting both genes reduced growth severely, indicating that these factors can compensate for each other. This compensatory behavior, also observed in *Vibrio natriegens*, suggests a broader and potentially more critical role for EfpL in ribosomal rescue within *Vibrio* species, suggesting distinct regulation patterns in Vibrionales compared to Enterobacterales **(Chapter 3)**. For *Vibrio* species, including the human pathogens *Vibrio cholerae* and *Vibrio parahaemolyticus*, the aquatic animal pathogens *V. campbellii* and *Vibrio anguillarum*, as well as the fastest growing bacterium *V. natriegens*, efficient translation, and rapid adaption to changing metabolic environments are essential for survival and competitiveness. This may be related to the altered regulation and improved function of EfpL in *Vibrio* species. Analyzing EfpL across different bacteria may reveal additional ribosome stalling motifs, new mechanisms of EfpL regulation, and unique interactions between EF-P and EfpL.

## 4.3 Interplay of different translation factors

The *efp* gene can be deleted in most bacteria, but notable exceptions include *Neisseria meningitidis*[164], *Acinetobacter baumannii*[165], and *Mycobacterium tuberculosis*[166]. Interestingly, the proteome of *M. tuberculosis* contains over 420 proteins with at least three consecutive prolines, accounting for more than 10% of the proteome[98]. The proportion of proteins with three or more consecutive prolines exceeds 10% in humans, where eIF5A is essential[29,58]. The abundance of polyproline-containing proteins could be an explanation for the importance of EF-P for the survival of certain bacteria. Alternatively, the translation of species-specific essential proteins could depend on EF-P, making the factor obligatory rather than selecting against polyproline stretches in bacteria[98,167].

Our discovery and characterization of EfpL may provide another explanation: *N. meningitidis*, *A. baumannii*, and *M. tuberculosis* encode only for EF-P and not for EfpL. In the case of *E. coli* and *Vibrio* species, EfpL can partially or completely take over EF-Ps function, making EF-P dispensable. Moreover, *efp* and *efpL* can be both deleted in *E. coli*, *V. campbellii* and *V. natriegens*, resulting in a growth-deficient strain, which can be further explained by the existence of other proteins that alleviate ribosome stalling at proline-rich sequences **(Chapter 3)**. EF-P is an ancient translation factor that existed before the divergence of bacteria, eukaryotes, and archaea[103]. Proline, along with amino acids encoded by GNN codons **(Fig. 2)** can be synthesized under primordial conditions[168]. This suggests that EF-P/IF5A initially played a vital role in facilitating proline-containing peptide bond formation in the P-site, with the

ability to interpret E-site codons via a second OB-domain offering an evolutionary advantage. After the divergence, IF5A adopted a more generalized role in translation and underwent structural refinements[74,88,169]. Today, *eIF5A* ranks among the top 50 most highly expressed genes in *S. cerevisiae* or HeLa cells[170], with expression levels comparable to those of ribosomes[171]. EF-P's more specialized role in bacteria likely led to the evolution of other factors to support translation in diverse environments. These proteins share some structural similarity and comparable charge distribution with EF-P[143]. The ABCF ATPases consist of two nucleotide-binding domains (NBDs), the interdomain linker sequence, and C-terminal extension (CTE). The interdomain linker, in particular, can rearrange the architecture of the peptidyl transferase center (PTC) and influence the positioning of the tRNA within the P-site, thereby contributing to their functional versatility[114]. Furthermore, YebC is composed of three domains. Domains I and II are particularly noteworthy for their positively charged surface patches, associated with RNA-binding sites, underscoring their potential role in RNA interactions. In contrast, domain III exhibits a highly negatively charged surface, which may contribute to its specific functional or structural interactions within the cellular environment[115].

### 4.3.1 Regulation of translation factors

In *E. coli* four proteins are described to alleviate ribosome stalling at proline-rich sequences: EF-P[29], the ABCF ATPase Uup[114], and the YebC family proteins YebC and YeeN (YebC2)[115]. We could add and characterize another protein, the EF-P paralog EfpL **(Chapter 3)**. The ability of these factors to bind stalled ribosomes underscores the importance of their regulation and interplay. A single deletion of *uup*, *efp,* or *efpL,* as well as combined double deletions in *E. coli,* were compensated for by the existence of the other systems, but growth phenotypes of different strengths were observed. We were unable to construct a triple deletion mutant comprising *uup*, *efp,* and *efpL* unless one of the genes was reintroduced in *trans* **(Chapter 3)**. In *E. coli* EfpL, YebC, YeeN (YebC2), and Uup are low abundant, with 10-100 times less proteins per cell than EF-P[172] **(Fig. 10)**. This suggests that EF-P is the major ligand for free E-site ribosomes in *E. coli*. The overrepresentation of EF-P, however, could be altered by various environmental conditions, allowing other factors to take precedence. With varying ATP concentrations in the cell, the function of Uup is influenced[124]. ATP levels closely reflect changes in metabolic activity, physiology, and adaptive responses to diverse environmental conditions and stresses[173-175]. Together with the interdomain linker of Uup, the ATP hydrolysis coupled structural rearrangement is pivotal for handling polyproline motifs[114]. This dynamic interplay between ATP levels and Uup function underscores its importance in responding to cellular needs. Similarly, EfpL activity is closely tied to the metabolic status of the cell, reflecting its role as a responsive regulator within the translation machinery. Its activity is negatively influenced by non-enzymatic acylations, a form of post-translational modification that likely arises under specific metabolic or environmental conditions **(Chapter 3)**. The non-enzymatic

acylation of EfpL may serve as a protective strategy, temporarily downregulating its activity during periods of metabolic imbalance or environmental challenge. This flexibility allows EfpL to act as an important adaptive factor, enabling cells to respond effectively to fluctuations in their environment and ensuring the robustness of the translation system under diverse conditions. Conversely, the modification at the β3Ωβ4 loop in EF-P is, in most cases, important for the function and activity. These modifications exhibit great diversity, ranging from lysine-β-lysylation[83,93,97,176,177] to lysine-5-amino-pentanolylation[99,100] to arginine-rhamnosylation[80,106] **(Fig. 5)**. Each of these modifications is mediated by specialized enzymatic systems. Many EF-Ps remain still uncharacterized, and it is likely that not all modification systems have yet been discovered. For instance, a deoxyhypusine synthase gene of crenarchaeal origin was found to be horizontally transferred into some bacterial groups[178,179]. This raises the possibility that EF-P could be modified through a similar pathway seen in archaea. Furthermore, different EF-P subgroups feature extraordinary amino acids at the tip of the functionally significant β3Ωβ4 loop, like methionine, threonine, asparagine, and glutamine **(Chapter 3)**. These changes can contribute to variations in the structural flexibility of EF-P, thereby modulating the interaction of EF-P with ribosomal machinery and affecting its functional efficiency.

EF-P and EfpL were found to alleviate stalling at a similar motif spectrum, including not only polyproline sequences but also motifs likely containing an amino acid beginning with guanosine paired with a proline at the P-site **(Chapter 3)**. A *yebC* deletion mutant in *S. pyogenes* exhibited ribosome stalling at sequences containing PP, PXP, or DXP[118]. In *B. subtilis*, YebC2 was reported to alleviate stalling at five consecutive prolines[115], while YfmR was shown to resolve stalling at five consecutive prolines[125] and aspartate-proline-containing motifs[117]. Uup enhanced the translation of ten consecutive prolines, but less efficient than EF-P in *E. coli*[114]. Additionally, our novel finding that EF-P and EfpL themselves can induce ribosome stalling adds to the complexity **(Chapter 3)**. XPX motifs, like VPW and IPI, can be EF-P- and EfpL-dependent stalling motifs **(Fig. 7)**, but for example, APH showed an increase in pausing strength when overexpressing these factors. This suggests that EF-P and EfpL, by occasionally obstructing tRNA translocation to the E-site, may cause pauses at non-stalling XPX motifs. Additionally, ribosome profiling data revealed that EF-P might sometimes bind non-productively. These binding events became more frequent and severe when EfpL was overexpressed in an EF-P-deficient background, indicating that EF-P and EfpL structurally differ in aligning and stabilizing the P-site prolyl-tRNA **(Chapter 3)**. The presence of both EF-P and EfpL may thus allow gene-specific modulation of translation speed, as inducing ribosome pausing could allow time for processes such as protein folding or membrane insertion[15].

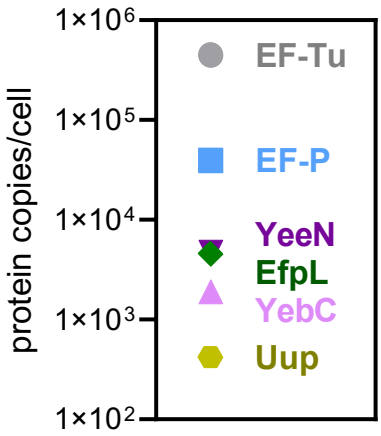



**Figure 10: Translation factors protein copies per cell**

Quantitative proteome analysis of *E. coli* grown in Lysogeny Broth for EF-Tu and the different proline-mediated ribosome stalling rescue factors: EF-P, YeeN (YebC2), EfpL, YebC and Uup analysed by Schmidt *et al.*[172].

### 4.3.2 The redundancy of ribosome rescue factors

Inefficient binding of one factor can block the ribosome, preventing other potentially more effective factors from alleviating stalling. Ribosome binding appears to occur randomly, primarily requiring an empty ribosomal E-site[143]. Differences among the factors, such as the motifs they target, their interaction efficiency with P-site tRNA, modification status, or cellular concentration, could affect their capacity to relieve the stalling. While the factors seem to target a similar range of stalling motifs[114,115,117,125], EF-P, and likely EfpL, also interact with the mRNA at the E-site codon[145] **(Chapter 3)**. Additionally, the different structures and interaction sites alter protein function and influence the binding and alleviation of stalling, furthermore, assisted or impeded by post-translational modifications, as shown for EF-P[80,83,95,101,103] or EfpL **(Chapter 3)**. The lysylation of EF-P was found to enhance the ribosome association rate and stabilize EF-P for the target support in peptide bond formation[143]. A simultaneous ribosomal association of two factors binding to the E-site of the ribosome is impossible[117], which is why the concentration of the factors in the cell plays an important role. Different bacteria exhibit varying compositions of proline-mediated ribosome rescue factors. *B. subtilis*, for instance, encode for 5-amino-pentanolylated EF-P[99], YebC2[115] and YfmR[125] but lacks EfpL The proteome of *V. campbellii* contains lysylated EF-P[180], YebC2[115,118], Uup[116], and EfpL, with *efpL* expression levels similar to that of *efp*, unlike in *E. coli* **(Chapter 3)**. This raises the question of whether the composition and concentration of these factors can be linked to the translation efficiency and, thus, the growth rate. We could already detect a positive correlation between EfpL being present and a higher growth speed in bacteria **(Chapter 3)**.

The presence of four or more factors capable of binding to a ribosome with an empty E-site may lead to competition among them **(Fig. 11)**. Different studies have shown that the factors work independently and can compensate for each other[115] **(Chapter 3)**. Another work

described a strong generic interaction between *efp* and *yfmR*, suggesting for supporting functions[117]. The authors assumed that EF-P could bind first, positions the prolyl-tRNA at the P-site and YfmR then resolves the stalling. They justify this by the fact that an unmodified EF-P variant, that cannot resolve stalling at polyprolines, is clearly able to support YfmR. Another hypothesis proposed for the cooperation of EF-P and YfmR is that EF-P could have an additional, yet unknown function, that does not require modification but involves an interaction with YfmR instead[117]. A broader role in translation elongation, specifically in ribosome-associated quality control, was already described for eIF5A[75].

Moreover, the translational initiation rate influences the effect of ribosome stalling, with highly expressed proteins being more dependent on EF-P[181]. Misregulation of translation has widespread effects on cellular processes. For instance, a reduced translation rate increases the likelihood of RNA polymerase backtracking, which in turn lowers the frequency of head-on collisions between RNA polymerase and the replication machinery[182]. Additionally, ribosome collisions would rapidly occur without factors that can facilitate the translation of hard-to-translate sequences. Upon ribosome collisions, the protein SmrB would be recruited and activated to cleave the mRNA sequences upstream of the stalled ribosomes, enabling the rescue factor tmRNA to release the stalled ribosome[183]. Since E-site binding ribosome rescue factors, like EF-P or EfpL, generally engage stalled ribosomes first, they help alleviate more transient stalls, especially those caused by specific sequences. When they fail to resolve the stall effectively, tmRNA or ArfA/ArfB step in as secondary rescue mechanisms[22]. Protein biosynthesis is the most energy-intensive process in cellular proliferation, with ribosomal translation alone estimated to consume approximately 50% of the energy in a rapidly growing bacterial cell[184]. This could explain why this variety of ribosome rescue factors exists. EF-P, EfpL, and other E-site binding factors handle sequence-specific stalls, while *trans*-translation and ArfA/ArfB resolve more terminal or unresolvable pauses[22]. The complementarity underscores an efficient and adaptable bacterial ribosome rescue system. By balancing the different rescue factors, bacteria could manage various stalling events efficiently, maintaining translation fidelity and allowing ribosomes to focus on productive protein synthesis.

Further research into how different bacterial species utilize ribosome rescue factors has the potential to provide profound insights into the evolution of translational machinery. Such studies could illuminate how organisms have fine-tuned their rescue systems to overcome specific environmental or metabolic challenges, offering understanding of the adaptive strategies employed across diverse ecological niches. The analyses of EF-P and its paralog, EfpL, revealed already the nuanced role of evolutionary pressures in shaping translational efficiency, particularly in managing ribosome stalling at proline-containing motifs **(Chapter 3)**. This regulatory mechanism underscores how subtle variations in translational dynamics can influence cellular fitness. Differential stalling strength observed based on codon choice and the

amino acid context not only highlights the complexity of translation but also demonstrates how natural selection acts at a molecular level to optimize translation for fitness, pathogenicity, and adaptability **(Chapter 2)**. Moreover, the diversity in ribosome rescue strategies across bacterial species underscores that the translational machinery is a highly adaptable and evolvable feature of the cell. Continued study of EF-P, EfpL, and other ribosome rescue factors will likely uncover additional layers of regulation, revealing how organisms balance the trade-offs between speed, accuracy, and energy efficiency in protein synthesis.

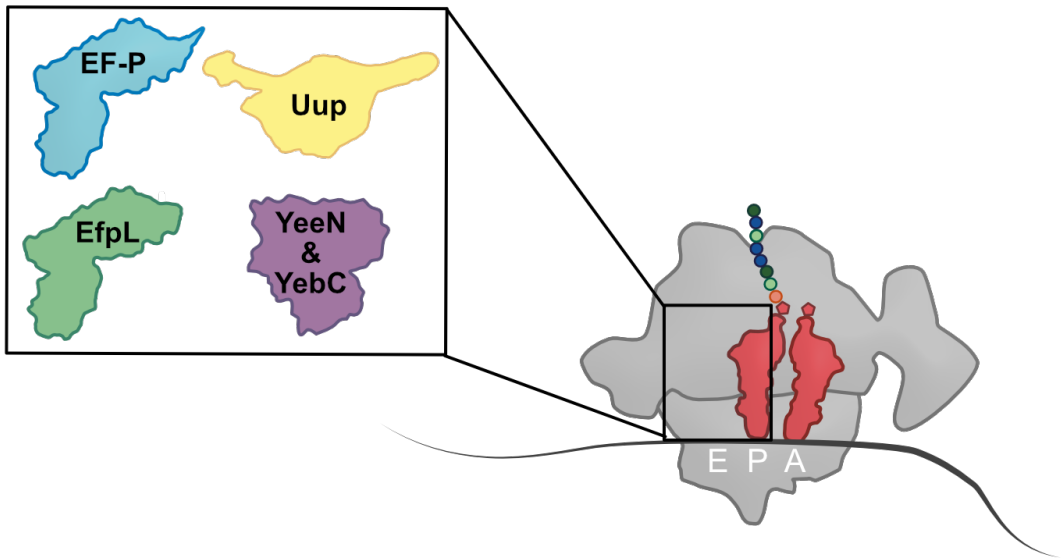



**Figure 11: Potential E-site ribosome binding proteins**

Depicted are EF-P (*E. coli*; pdb: 6ENU), EfpL (*E. coli*; pdb: 8S8U), ABCF ATPase Uup (*E. coli*; Alphafold prediction), and the RNA binding protein YeeN (YebC2) (*E. coli*; Alphafold prediction) as potential free E-site binding ribosome rescue factors, when translation stalls at proline-rich sequences.

### 4.3.3 The underlying potential of the translation factors

Beyond their biological roles, EF-P and EfpL hold promise for applications in biotechnology and medicine. The ability to control translation through codon selection and the codon interactions EF-P and EfpL mediate offers strategic opportunities for biotechnological innovation. By harnessing the precise control mechanisms of EF-P and EfpL, researchers could design systems to fine-tune protein expression, optimize metabolic pathways, and enhance the production of valuable biomolecules. Furthermore, targeting EF-P in pathogens represents a promising avenue for novel antibiotic development, offering a pathway to combat drug-resistant bacterial strains by disrupting their essential translational processes[185]. EF-P's ability to facilitate peptide bond formation with non-canonical amino acids (ncAAs) expands its utility into synthetic biology, where it could play a central role to extend the genetic code[57,186,187]. Intriguingly, EfpL's structural differences suggest, it could be particularly advantageous for incorporating unique ncAAs, potentially improving protein engineering applications.

Moreover, introducing EfpL proteins into engineered systems could alleviate translation bottlenecks when expressing proteins with complex motifs that typically challenge ribosomal processes. This capacity could not only improve yields in industrial-scale protein production but also enhance growth rates. EfpL offers the possibility of a functioning rescue system without modification enzymes and can, at the same time, be regulated via the metabolism of the cell. In summary, EF-P and EfpL exemplify the sophisticated translational control mechanisms that enable bacteria to survive in dynamic environments, enriching our understanding of microbial adaptation. With their broad applications, from biotechnology to antibiotic development, these factors are invaluable in unraveling bacterial translation regulation and developing innovative therapeutic and synthetic biology strategies.

## References for Chapters 1 and 4


1 Kloepfer, H. G. Struktur und Funktion von Ribosomen. *Chemie in unserer Zeit* **7**, 49-58 (1973). https://doi.org/https://doi.org/10.1002/ciuz.19730070204

2 Forchhammer, J. & Lindahl, L. Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of *Escherichia coli* 15. *J Mol Biol* **55**, 563-568 (1971). https://doi.org/10.1016/0022-2836(71)90337-8

3 Shine, J. & Dalgarno, L. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A* **71**, 1342-1346 (1974). https://doi.org/10.1073/pnas.71.4.1342

4 Gualerzi, C. O. & Pon, C. L. Initiation of mRNA translation in bacteria: structural and dynamic aspects. *Cell Mol Life Sci* **72**, 4341-4367 (2015). https://doi.org/10.1007/s00018-015-2010-3

5 Rodnina, M. V. & Wintermeyer, W. Protein Elongation, Co-translational Folding and Targeting. *J Mol Biol* **428**, 2165-2185 (2016). https://doi.org/10.1016/j.jmb.2016.03.022

6 Hirokawa, G., Demeshkina, N., Iwakura, N., Kaji, H. & Kaji, A. The ribosome-recycling step: consensus or controversy? *Trends Biochem Sci* **31**, 143-149 (2006). https://doi.org/10.1016/j.tibs.2006.01.007

7 Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25**, 117-124 (2007). https://doi.org/10.1038/nbt1270

8 Bremer, H. & Dennis, P. Feedback control of ribosome function in *Escherichia coli*. *Biochimie* **90**, 493-499 (2008). https://doi.org/10.1016/j.biochi.2007.10.008

9 Rodnina, M. V. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci* **25**, 1390-1406 (2016). https://doi.org/10.1002/pro.2950

10 Woolstenhulme, C. J., Guydosh, N. R., Green, R. & Buskirk, A. R. High-Precision Analysis of Translational Pausing by Ribosome Profiling in Bacteria Lacking EFP. *Cell Rep.* **11**, 13-21 (2015). https://doi.org/10.1016/j.celrep.2015.03.014

11 Wen, J. D. *et al.* Following translation by single ribosomes one codon at a time. *Nature* **452**, 598-603 (2008). https://doi.org/10.1038/nature06716

12 Mohammad, F., Green, R. & Buskirk, A. R. A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *Elife* **8** (2019). https://doi.org/10.7554/eLife.42591

13 Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S. & Grayhack, E. J. Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. *Cell* **166**, 679-690 (2016). https://doi.org/10.1016/j.cell.2016.05.070

14 Komar, A. A. A pause for thought along the co-translational folding pathway. *Trends Biochem Sci* **34**, 16-24 (2009). https://doi.org/10.1016/j.tibs.2008.10.002

15 Collart, M. A. & Weiss, B. Ribosome pausing, a dangerous necessity for co-translational events. *Nucleic Acids Res* **48**, 1043-1055 (2019). https://doi.org/10.1093/nar/gkz763

16 Yan, S., Wen, J.-D., Bustamante, C. & Tinoco, I., Jr. Ribosome Excursions during mRNA Translocation Mediate Broad Branching of Frameshift Pathways. *Cell* **160**, 870-881 (2015). https://doi.org/10.1016/j.cell.2015.02.003

17 Qi, F., Motz, M., Jung, K., Lassak, J. & Frishman, D. Evolutionary analysis of polyproline motifs in *Escherichia coli* reveals their regulatory role in translation. *PLoS Comput Biol* **14**, e1005987 (2018). https://doi.org/10.1371/journal.pcbi.1005987

18 Tenson, T. & Ehrenberg, M. Regulatory Nascent Peptides in the Ribosomal Tunnel. *Cell* **108**, 591-594 (2002). https://doi.org/https://doi.org/10.1016/S0092-8674(02)00669-4

19 Ramu, H. *et al.* Nascent Peptide in the Ribosome Exit Tunnel Affects Functional Properties of the A-Site of the Peptidyl Transferase Center. *Mol Cell* **41**, 321-330 (2011). https://doi.org/https://doi.org/10.1016/j.molcel.2010.12.031

20 Leedom, S. L. & Keiler, K. C. Ribosome collisions: New ways to initiate ribosome rescue. *Curr Biol* **32**, R469-r472 (2022). https://doi.org/10.1016/j.cub.2022.04.038
21 Keiler, K. C. Biology of *trans*-translation. *Annu Rev Microbiol* **62**, 133-151 (2008). https://doi.org/10.1146/annurev.micro.62.081307.162948
22 Keiler, K. C. Mechanisms of ribosome rescue in bacteria. *Nat Rev Microbiol* **13**, 285-297 (2015). https://doi.org/10.1038/nrmicro3438
23 Ivanova, N., Pavlov, M. Y., Felden, B. & Ehrenberg, M. Ribosome rescue by tmRNA requires truncated mRNAs. *J Mol Biol* **338**, 33-41 (2004). https://doi.org/10.1016/j.jmb.2004.02.043
24 Kurita, D., Chadani, Y., Muto, A., Abo, T. & Himeno, H. ArfA recognizes the lack of mRNA in the mRNA channel after RF2 binding for ribosome rescue. *Nucleic Acids Res* **42**, 13339-13352 (2014). https://doi.org/10.1093/nar/gku1069
25 Thomas, E. N., Kim, K. Q., McHugh, E. P., Marcinkiewicz, T. & Zaher, H. S. Alkylative damage of mRNA leads to ribosome stalling and rescue by *trans* translation in bacteria. *Elife* **9** (2020). https://doi.org/10.7554/eLife.61984
26 Abo, T., Ueda, K., Sunohara, T., Ogawa, K. & Aiba, H. SsrA-mediated protein tagging in the presence of miscoding drugs and its physiological role in *Escherichia coli*. *Genes Cells* **7**, 629-638 (2002). https://doi.org/10.1046/j.1365-2443.2002.00549.x
27 Zaher, H. S. & Green, R. Quality control by the ribosome following peptide bond formation. *Nature* **457**, 161-166 (2009). https://doi.org/10.1038/nature07582
28 Doerfel, L. K. *et al.* EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science* **339**, 85-88 (2013). https://doi.org/10.1126/science.1229017
29 Ude, S. *et al.* Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* **339**, 82-85 (2013). https://doi.org/10.1126/science.1228985
30 Lassak, J., Wilson, D. N. & Jung, K. Stall no more at polyproline stretches with the translation elongation factors EF-P and IF-5A. *Mol Microbiol* **99**, 219-235 (2016). https://doi.org/10.1111/mmi.13233
31 Huter, P. *et al.* Structural basis for polyproline-mediated ribosome stalling and rescue by the translation elongation factor EF-P. *Mol Cell* **68**, 515-527.e516 (2017). https://doi.org/10.1016/j.molcel.2017.10.014
32 Petropoulos, A. D., McDonald, M. E., Green, R. & Zaher, H. S. Distinct roles for release factor 1 and release factor 2 in translational quality control. *J Biol Chem* **289**, 17589-17596 (2014). https://doi.org/10.1074/jbc.M114.564989
33 T, N., Govindarajan, S. & Munavar, M. H. trans-translation system is important for maintaining genome integrity during DNA damage in bacteria. *Res Microbiol* **174**, 104136 (2023). https://doi.org/10.1016/j.resmic.2023.104136
34 Chadani, Y. *et al.* Ribosome rescue by *Escherichia coli* ArfA (YhdL) in the absence of *trans*-translation system. *Mol Microbiol* **78**, 796-808 (2010). https://doi.org/10.1111/j.1365-2958.2010.07375.x
35 Chan, K.-H. *et al.* Mechanism of ribosome rescue by alternative ribosome-rescue factor B. *Nat Commun* **11**, 4106 (2020). https://doi.org/10.1038/s41467-020-17853-7
36 Giudice, E. & Gillet, R. The task force that rescues stalled ribosomes in bacteria. *Trends Biochem Sci* **38**, 403-411 (2013). https://doi.org/10.1016/j.tibs.2013.06.002
37 Crowe-McAuliffe, C. *et al.* Structural Basis for Bacterial Ribosome-Associated Quality Control by RqcH and RqcP. *Mol Cell* **81**, 115-126.e117 (2021). https://doi.org/10.1016/j.molcel.2020.11.002
38 Filbeck, S. *et al.* Mimicry of Canonical Translation Elongation Underlies Alanine Tail Synthesis in RQC. *Mol Cell* **81**, 104-114.e106 (2021). https://doi.org/10.1016/j.molcel.2020.11.001
39 Lytvynenko, I. *et al.* Alanine Tails Signal Proteolysis in Bacterial Ribosome-Associated Quality Control. *Cell* **178**, 76-90.e22 (2019). https://doi.org/10.1016/j.cell.2019.05.002

40 Dinçbas, V., Heurgué-Hamard, V., Buckingham, R. H., Karimi, R. & Ehrenberg, M. Shutdown in protein synthesis due to the expression of mini-genes in bacteria. *J Mol Biol* **291**, 745-759 (1999). https://doi.org/10.1006/jmbi.1999.3028
41 Gonzalez de Valdivia, E. I. & Isaksson, L. A. Abortive translation caused by peptidyl-tRNA drop-off at NGG codons in the early coding region of mRNA. *Febs j* **272**, 5306-5316 (2005). https://doi.org/10.1111/j.1742-4658.2005.04926.x
42 Heurgué-Hamard, V., Dinçbas, V., Buckingham, R. H. & Ehrenberg, M. Origins of minigene-dependent growth inhibition in bacterial cells. *Embo j* **19**, 2701-2709 (2000). https://doi.org/10.1093/emboj/19.11.2701
43 Jacinto-Loeza, E., Vivanco-Domínguez, S., Guarneros, G. & Hernández-Sánchez, J. Minigene-like inhibition of protein synthesis mediated by hungry codons near the start codon. *Nucleic Acids Res* **36**, 4233-4241 (2008). https://doi.org/10.1093/nar/gkn395
44 Vivanco-Domínguez, S., Cruz-Vera, L. R. & Guarneros, G. Excess of charged $tRNA^{Lys}$ maintains low levels of peptidyl-tRNA hydrolase in *pth*(Ts) mutants at a non-permissive temperature. *Nucleic Acids Res* **34**, 1564-1570 (2006). https://doi.org/10.1093/nar/gkl046
45 Atherly, A. G. Peptidyl-transfer RNA hydrolase prevents inhibition of protein synthesis initiation. *Nature* **275**, 769 (1978). https://doi.org/10.1038/275769a0
46 Cuzin, F., Kretchmer, N., Greenberg, R. E., Hurwitz, R. & Chapeville, F. Enzymatic hydrolysis of *N*-substituted aminoacyl-tRNA. *Proc Natl Acad Sci U S A* **58**, 2079-2086 (1967). https://doi.org/10.1073/pnas.58.5.2079
47 Kössel, H. & RajBhandary, U. L. Studies on polynucleotides. LXXXVI. Enzymic hydrolysis of N-acylaminoacyl-transfer RNA. *J Mol Biol* **35**, 539-560 (1968). https://doi.org/10.1016/s0022-2836(68)80013-0
48 Menez, J., Heurgué-Hamard, V. & Buckingham, R. H. Sequestration of specific tRNA species cognate to the last sense codon of an overproduced gratuitous protein. *Nucleic Acids Res* **28**, 4725-4732 (2000). https://doi.org/10.1093/nar/28.23.4725
49 Menninger, J. R. Accumulation of peptidyl tRNA is lethal to *Escherichia coli*. *J Bacteriol* **137**, 694-696 (1979). https://doi.org/10.1128/jb.137.1.694-696.1979
50 Pavlov, M. Y. *et al.* Slow peptide bond formation by proline and other *N*-alkylamino acids in translation. *Proc Natl Acad Sci U S A* **106**, 50-54 (2009). https://doi.org/10.1073/pnas.0809211106
51 Chevance, F. F., Le Guyon, S. & Hughes, K. T. The effects of codon context on i*n vivo* translation speed. *PLoS Genet* **10**, e1004392 (2014). https://doi.org/10.1371/journal.pgen.1004392
52 Johansson, M. *et al.* pH-sensitivity of the ribosomal peptidyl transfer reaction dependent on the identity of the A-site aminoacyl-tRNA. *Proc Natl Acad Sci U S A* **108**, 79-84 (2011). https://doi.org/10.1073/pnas.1012612107
53 Lu, K. P., Finn, G., Lee, T. H. & Nicholson, L. K. Prolyl *cis-trans* isomerization as a molecular timer. *Nat Chem Biol* **3**, 619-629 (2007). https://doi.org/10.1038/nchembio.2007.35
54 Macinga, D. R., Parojcic, M. M. & Rather, P. N. Identification and analysis of *aarP*, a transcriptional activator of the 2'-*N*-acetyltransferase in *Providencia stuartii*. *J Bacteriol* **177**, 3407-3413 (1995). https://doi.org/10.1128/jb.177.12.3407-3413.1995
55 Adzhubei, A. A., Sternberg, M. J. & Makarov, A. A. Polyproline-II helix in proteins: structure and function. *J Mol Biol* **425**, 2100-2132 (2013). https://doi.org/10.1016/j.jmb.2013.03.018
56 Jha, A. K. *et al.* Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry* **44**, 9691-9702 (2005). https://doi.org/10.1021/bi0474822
57 Doerfel, L. K. *et al.* Entropic Contribution of Elongation Factor P to Proline Positioning at the Catalytic Center of the Ribosome. *J Am Chem Soc* **137**, 12997-13006 (2015). https://doi.org/10.1021/jacs.5b07427
58 Morgan, A. A. & Rubenstein, E. Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human

Proteome. *PLoS ONE* **8**, e53785 (2013). https://doi.org/10.1371/journal.pone.0053785
59 Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* **28**, 292 (2000). https://doi.org/10.1093/nar/28.1.292
60 Mandal, A., Mandal, S. & Park, M. H. Genome-Wide Analyses and Functional Classification of Proline Repeat-Rich Proteins: Potential Role of eIF5A in Eukaryotic Evolution. *PLoS ONE* **9**, e111800 (2014). https://doi.org/10.1371/journal.pone.0111800
61 Gutierrez, E. *et al.* eIF5A Promotes Translation of Polyproline Motifs. *Mol. Cell* **51**, 1-11 (2013).
62 Kemper, W. M., Berry, K. W. & Merrick, W. C. Purification and properties of rabbit reticulocyte protein synthesis initiation factors M2B*alpha* and M2B*beta*. *J Biol Chem* **251**, 5551-5557 (1976).
63 Schreier, M. H., Erni, B. & Staehelin, T. Initiation of mammalian protein synthesis. I. Purification and characterization of seven initiation factors. *J Mol Biol* **116**, 727-753 (1977). https://doi.org/10.1016/0022-2836(77)90268-6
64 Benne, R., Brown-Luedi, M. L. & Hershey, J. W. Purification and characterization of protein synthesis initiation factors eIF-1, eIF-4C, eIF-4D, and eIF-5 from rabbit reticulocytes. *J Biol Chem* **253**, 3070-3077 (1978).
65 Dever, T. E., Gutierrez, E. & Shin, B. S. The hypusine-containing translation factor eIF5A. *Crit Rev Biochem Mol Biol* **49**, 413-425 (2014). https://doi.org/10.3109/10409238.2014.939608
66 Glick, B. R. & Ganoza, M. C. Identification of a soluble protein that stimulates peptide bond synthesis. *Proc Natl Acad Sci U S A* **72**, 4257-4260 (1975). https://doi.org/10.1073/pnas.72.11.4257
67 Katinka, M. D. *et al.* Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450-453 (2001). https://doi.org/10.1038/35106579
68 Derelle, E. *et al.* Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc Natl Acad Sci U S A* **103**, 11647-11652 (2006). https://doi.org/10.1073/pnas.0604795103
69 Ganoza, M. C. & Aoki, H. Peptide bond synthesis: function of the *efp* gene product. *Biol Chem* **381**, 553-559 (2000). https://doi.org/10.1515/bc.2000.071
70 Blaha, G., Stanley, R. E. & Steitz, T. A. Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. *Science* **325**, 966-970 (2009). https://doi.org/325/5943/966 10.1126/science.1175800
71 Hanawa-Suetsugu, K. *et al.* Crystal structure of elongation factor P from *Thermus thermophilus* HB8. *Proc Natl Acad Sci U S A* **101**, 9595-9600 (2004). https://doi.org/10.1073/pnas.0308667101 0308667101
72 Bailly, M. & de Crecy-Lagard, V. Predicting the pathway involved in post-translational modification of elongation factor P in a subset of bacterial species. *Biol Direct* **5**, 3 (2010). https://doi.org/1745-6150-5-3 10.1186/1745-6150-5-3
73 Cheng, Y. *et al.* Eukaryotic translation factor eIF5A contributes to acetic acid tolerance in *Saccharomyces cerevisiae* via transcriptional factor Ume6p. *Biotechnol Biofuels* **14**, 38 (2021). https://doi.org/10.1186/s13068-021-01885-2
74 Schuller, A. P., Wu, C. C., Dever, T. E., Buskirk, A. R. & Green, R. eIF5A Functions Globally in Translation Elongation and Termination. *Mol Cell* **66**, 194-205.e195 (2017). https://doi.org/10.1016/j.molcel.2017.03.003
75 Tesina, P. *et al.* Molecular basis of eIF5A-dependent CAT tailing in eukaryotic ribosome-associated quality control. *Mol Cell* **83**, 607-621.e604 (2023). https://doi.org/10.1016/j.molcel.2023.01.020
76 Kim, K. K., Hung, L. W., Yokota, H., Kim, R. & Kim, S. H. Crystal structures of eukaryotic translation initiation factor 5A from *Methanococcus jannaschi*i at 1.8 A resolution. *Proc Natl Acad Sci U S A* **95**, 10419-10424 (1998). https://doi.org/10.1073/pnas.95.18.10419

77 Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C. & Ferrin, T. E. Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* **7**, 339 (2006). https://doi.org/10.1186/1471-2105-7-339
78 Yao, M., Ohsawa, A., Kikukawa, S., Tanaka, I. & Kimura, M. Crystal structure of hyperthermophilic archaeal initiation factor 5A: a homologue of eukaryotic initiation factor 5A (eIF-5A). *J Biochem* **133**, 75-81 (2003). https://doi.org/10.1093/jb/mvg011
79 Peat, T. S., Newman, J., Waldo, G. S., Berendzen, J. & Terwilliger, T. C. Structure of translation initiation factor 5A from *Pyrobaculum aerophilum* at 1.75 A resolution. *Structure* **6**, 1207-1214 (1998). https://doi.org/10.1016/s0969-2126(98)00120-8
80 Lassak, J. *et al.* Arginine-rhamnosylation as new strategy to activate translation elongation factor P. *Nat Chem Biol* **11**, 266-270 (2015). https://doi.org/10.1038/nchembio.1751
81 Lassak, J., Sieber, A. & Hellwig, M. Exceptionally versatile take II: post-translational modifications of lysine and their impact on bacterial physiology. *Biol Chem* (2022). https://doi.org/10.1515/hsz-2021-0382
82 Park, J. H. *et al.* Molecular cloning, expression, and structural prediction of deoxyhypusine hydroxylase: a HEAT-repeat-containing metalloenzyme. *Proc Natl Acad Sci U S A* **103**, 51-56 (2006). https://doi.org/10.1073/pnas.0509348102
83 Park, J. H. *et al.* Post-translational modification by β-lysylation is required for activity of *Escherichia coli* elongation factor P (EF-P). *J Biol Chem* **287**, 2579-2590 (2012). https://doi.org/10.1074/jbc.M111.309633
84 Duncan, R. F. & Hershey, J. W. Changes in eIF-4D hypusine modification or abundance are not correlated with translational repression in HeLa cells. *J Biol Chem* **261**, 12903-12906 (1986).
85 Murphey, R. J. & Gerner, E. W. Hypusine formation in protein by a two-step process in cell lysates. *J Biol Chem* **262**, 15033-15036 (1987).
86 Park, M. H. Regulation of biosynthesis of hypusine in Chinese hamster ovary cells. Evidence for eIF-4D precursor polypeptides. *J Biol Chem* **262**, 12730-12734 (1987).
87 Gordon, E. D., Mora, R., Meredith, S. C. & Lindquist, S. L. Hypusine formation in eukaryotic initiation factor 4D is not reversed when rates or specificity of protein synthesis is altered. *J Biol Chem* **262**, 16590-16595 (1987).
88 Park, M. H., Nishimura, K., Zanelli, C. F. & Valentini, S. R. Functional significance of eIF5A and its hypusine modification in eukaryotes. *Amino Acids* **38**, 491-500 (2010). https://doi.org/10.1007/s00726-009-0408-7
89 Wolff, E. C., Park, M. H. & Folk, J. E. Cleavage of spermidine as the first step in deoxyhypusine synthesis. The role of NAD. *J Biol Chem* **265**, 4793-4799 (1990).
90 Wolff, E. C., Folk, J. E. & Park, M. H. Enzyme-substrate intermediate formation at lysine 329 of human deoxyhypusine synthase. *J Biol Chem* **272**, 15865-15871 (1997). https://doi.org/10.1074/jbc.272.25.15865
91 Wolff, E. C., Wolff, J. & Park, M. H. Deoxyhypusine synthase generates and uses bound NADH in a transient hydride transfer mechanism. *J Biol Chem* **275**, 9170-9177 (2000). https://doi.org/10.1074/jbc.275.13.9170
92 Thompson, G. M., Cano, V. S. & Valentini, S. R. Mapping eIF5A binding sites for Dys1 and Lia1: *in vivo* evidence for regulation of eIF5A hypusination. *FEBS Lett* **555**, 464-468 (2003). https://doi.org/10.1016/s0014-5793(03)01305-x
93 Yanagisawa, T., Sumida, T., Ishii, R., Takemoto, C. & Yokoyama, S. A paralog of lysyl-tRNA synthetase aminoacylates a conserved lysine residue in translation elongation factor P. *Nat Struct Mol Biol* **17**, 1136-1143 (2010). https://doi.org/nsmb.1889 10.1038/nsmb.1889
94 Sumida, T., Yanagisawa, T., Ishii, R. & Yokoyama, S. Crystallization and preliminary X-ray crystallographic study of GenX, a lysyl-tRNA synthetase paralogue from *Escherichia coli*, in complex with translation elongation factor P. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**, 1115-1118 (2010). https://doi.org/10.1107/s1744309110032008

95 Navarre, W. W. *et al.* PoxA, YjeK, and elongation factor P coordinately modulate virulence and drug resistance in *Salmonella enterica*. *Mol Cell* **39**, 209-221 (2010). https://doi.org/S1097-2765(10)00462-4 10.1016/j.molcel.2010.06.021
96 Behshad, E. *et al.* Enantiomeric free radicals and enzymatic control of stereochemistry in a radical mechanism: the case of lysine 2,3-aminomutases. *Biochemistry* **45**, 12639-12646 (2006). https://doi.org/10.1021/bi061328t
97 Peil, L. *et al.* Lys34 of translation elongation factor EF-P is hydroxylated by YfcM. *Nat Chem Biol* **8**, 695-697 (2012). https://doi.org/10.1038/nchembio.1001
98 Starosta, A. L. *et al.* A conserved proline triplet in Val-tRNA synthetase and the origin of elongation factor P. *Cell Rep* **9**, 476-483 (2014). https://doi.org/10.1016/j.celrep.2014.09.008
99 Rajkovic, A. *et al.* Translation Control of Swarming Proficiency in *Bacillus subtilis* by 5-Amino-pentanolylated Elongation Factor P. *J Biol Chem* **291**, 10976-10985 (2016). https://doi.org/10.1074/jbc.M115.712091
100 Hummels, K. R. *et al.* Carbonyl reduction by YmfI in *Bacillus subtilis* prevents accumulation of an inhibitory EF-P modification state. *Mol Microbiol* **106**, 236-251 (2017). https://doi.org/https://doi.org/10.1111/mmi.13760
101 Witzky, A. *et al.* EF-P Posttranslational Modification Has Variable Impact on Polyproline Translation in *Bacillus subtilis*. *mBio* **9** (2018). https://doi.org/10.1128/mBio.00306-18
102 Pinheiro, B. *et al.* Structure and Function of an Elongation Factor P Subfamily in Actinobacteria. *Cell Reports* **30**, 4332-4342.e4335 (2020). https://doi.org/https://doi.org/10.1016/j.celrep.2020.03.009
103 Volkwein, W. *et al.* Switching the post-translational modification of translation elongation factor EF-P. *Front Microbiol* **10**, 1148 (2019). https://doi.org/10.3389/fmicb.2019.01148
104 Li, X. *et al.* Resolving the α-glycosidic linkage of arginine-rhamnosylated translation elongation factor P triggers generation of the first $Arg^{Rha}$ specific antibody. *Chem Sci* **7**, 6995-7001 (2016). https://doi.org/10.1039/c6sc02889f
105 Rajkovic, A. & Ibba, M. Elongation Factor P and the Control of Translation Elongation. *Annu Rev Microbiol* **71**, 117-131 (2017). https://doi.org/10.1146/annurev-micro-090816-093629
106 Krafczyk, R. *et al.* Structural Basis for EarP-Mediated Arginine Glycosylation of Translation Elongation Factor EF-P. *mBio* **8** (2017). https://doi.org/10.1128/mBio.01412-17
107 Gäbel, K., Schmitt, J., Schulz, S., Näther, D. J. & Soppa, J. A comprehensive analysis of the importance of translation initiation factors for *Haloferax volcanii* applying deletion and conditional depletion mutants. *PLoS One* **8**, e77188 (2013). https://doi.org/10.1371/journal.pone.0077188
108 Ohashi, Y. *et al.* Expression profiling of translation-associated genes in sporulating *Bacillus subtilis* and consequence of sporulation by gene inactivation. *Biosci Biotechnol Biochem* **67**, 2245-2253 (2003). https://doi.org/10.1271/bbb.67.2245
109 Kearns, D. B., Chu, F., Rudner, R. & Losick, R. Genes governing swarming in *Bacillus subtilis* and evidence for a phase variation mechanism controlling surface motility. *Mol Microbiol* **52**, 357-369 (2004). https://doi.org/10.1111/j.1365-2958.2004.03996.x
110 Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008 (2006). https://doi.org/10.1038/msb4100050
111 Bearson, S. M., Bearson, B. L. & Rasmussen, M. A. Identification of *Salmonella enterica* serovar Typhimurium genes important for survival in the swine gastric environment. *Appl Environ Microbiol* **72**, 2829-2836 (2006). https://doi.org/10.1128/aem.72.4.2829-2836.2006
112 Iannino, F., Ugalde, J. E. & Iñón de Iannino, N. *Brucella abortus efp* gene is required for an efficient internalization in HeLa cells. *Microb Pathog* **52**, 31-40 (2012). https://doi.org/10.1016/j.micpath.2011.09.008

113 Peng, W. T., Banta, L. M., Charles, T. C. & Nester, E. W. The *chvH* locus of *Agrobacterium* encodes a homologue of an elongation factor involved in protein synthesis. *J Bacteriol* **183**, 36-45 (2001). https://doi.org/10.1128/jb.183.1.36-45.2001
114 Chadani, Y. *et al.* The ABCF proteins in *Escherichia coli* individually cope with 'hard-to-translate' nascent peptide sequences. *Nucleic Acids Res* **52**, 5825-5840 (2024). https://doi.org/10.1093/nar/gkae309
115 Hong, H.-R., Prince, C. R., Wu, L., Lin, I. & Feaga, H. YebC2 resolves ribosome stalling at polyprolines independent of EF-P and the ABCF ATPase YfmR. *bioRxiv*, 2024.2010.2018.618948 (2024). https://doi.org/10.1101/2024.10.18.618948
116 Murina, V. *et al.* ABCF ATPases Involved in Protein Synthesis, Ribosome Assembly and Antibiotic Resistance: Structural and Functional Diversification across the Tree of Life. *J Mol Biol* **431**, 3568-3590 (2019). https://doi.org/10.1016/j.jmb.2018.12.013
117 Takada, H., Fujiwara, K., Atkinson, G. C., Chiba, S. & Hauryliuk, V. Resolution of ribosomal stalling by EF-P and ABCF ATPases YfmR and YkpA/YbiT. *Nucleic Acids Res* **52**, 9854–9866 (2024). https://doi.org/https://doi.org/10.1093/nar/gkae556
118 Ignatov, D. *et al.* Novel RNA-binding protein YebC enhances translation of proline-rich amino acid stretches in bacteria. *bioRxiv*, 2024.2008.2026.607280 (2024). https://doi.org/10.1101/2024.08.26.607280
119 Boël, G. *et al.* The ABC-F protein EttA gates ribosome entry into the translation elongation cycle. *Nat Struct Mol Biol* **21**, 143-151 (2014). https://doi.org/10.1038/nsmb.2740
120 Chen, B. *et al.* EttA regulates translation by binding the ribosomal E site and restricting ribosome-tRNA dynamics. *Nat Struct Mol Biol* **21**, 152-159 (2014). https://doi.org/10.1038/nsmb.2741
121 Sharkey, L. K., Edwards, T. A. & O'Neill, A. J. ABC-F Proteins Mediate Antibiotic Resistance through Ribosomal Protection. *mBio* **7**, e01975 (2016). https://doi.org/10.1128/mBio.01975-15
122 Rappas, M., Niwa, H. & Zhang, X. Mechanisms of ATPases -a multi-disciplinary approach. *Curr Protein Pept Sci* **5**, 89-105 (2004). https://doi.org/10.2174/1389203043486874
123 Cui, Z. *et al.* Interplay between an ATP-binding cassette F protein and the ribosome from *Mycobacterium tuberculosis*. *Nat Commun* **13**, 432 (2022). https://doi.org/10.1038/s41467-022-28078-1
124 Murat, D., Bance, P., Callebaut, I. & Dassa, E. ATP hydrolysis is essential for the function of the Uup ATP-binding cassette ATPase in precise excision of transposons. *J Biol Chem* **281**, 6850-6859 (2006). https://doi.org/10.1074/jbc.M509926200
125 Hong, H.-R., Prince, C. R., Tetreault, D. D., Wu, L. & Feaga, H. A. YfmR is a translation factor that prevents ribosome stalling and cell death in the absence of EF-P. *Proc Natl Acad Sci U S A* **121**, e2314437121 (2023). https://doi.org/https://doi.org/10.1073/pnas.2314437121
126 Zhang, Y., Lin, J. & Gao, Y. In silico identification of a multi-functional regulatory protein involved in Holliday junction resolution in bacteria. *BMC Syst Biol* **6 Suppl 1**, S20 (2012). https://doi.org/10.1186/1752-0509-6-s1-s20
127 Brischigliaro, M. *et al.* The human mitochondrial translation factor TACO1 alleviates mitoribosome stalling at polyproline stretches. *Nucleic Acids Res* **52**, 9710-9726 (2024). https://doi.org/10.1093/nar/gkae645
128 Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol* **260**, 649-663 (1996). https://doi.org/10.1006/jmbi.1996.0428
129 Gamper, H. B., Masuda, I., Frenkel-Morgenstern, M. & Hou, Y. M. Maintenance of protein synthesis reading frame by EF-P and $m^1G37$-tRNA. *Nat Commun* **6**, 7226 (2015). https://doi.org/10.1038/ncomms8226
130 de Smit, M. H., van Duin, J., van Knippenberg, P. H. & van Eijk, H. G. CCC.UGA: a new site of ribosomal frameshifting in *Escherichia coli*. *Gene* **143**, 43-47 (1994). https://doi.org/10.1016/0378-1119(94)90602-5

131 Ikemura, T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol* **151**, 389-409 (1981). https://doi.org/10.1016/0022-2836(81)90003-6
132 Grantham, R., Gautier, C., Gouy, M., Mercier, R. & Pavé, A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res* **8**, r49-r62 (1980). https://doi.org/10.1093/nar/8.1.197-c
133 Andersson, S. G. & Kurland, C. G. Codon preferences in free-living microorganisms. *Microbiol Rev* **54**, 198-210 (1990). https://doi.org/10.1128/mr.54.2.198-210.1990
134 Ikemura, T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* **2**, 13-34 (1985). https://doi.org/10.1093/oxfordjournals.molbev.a040335
135 Samatova, E., Daberger, J., Liutkute, M. & Rodnina, M. V. Translational Control by Ribosome Pausing in Bacteria: How a Non-uniform Pace of Translation Affects Protein Production and Folding. *Front Microbiol* **11**, 619430 (2020). https://doi.org/10.3389/fmicb.2020.619430
136 Mitarai, N., Sneppen, K. & Pedersen, S. Ribosome Collisions and Translation Efficiency: Optimization by Codon Usage and mRNA Destabilization. *J Mol Biol* **382**, 236-245 (2008). https://doi.org/https://doi.org/10.1016/j.jmb.2008.06.068
137 Tuller, T. *et al.* An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344-354 (2010). https://doi.org/10.1016/j.cell.2010.03.031
138 Tuller, T. *et al.* Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol* **12**, R110 (2011). https://doi.org/10.1186/gb-2011-12-11-r110
139 Tuller, T. & Zur, H. Multiple roles of the coding sequence 5' end in gene expression regulation. *Nucleic Acids Res* **43**, 13-28 (2015). https://doi.org/10.1093/nar/gku1313
140 Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504-508 (2010). https://doi.org/10.1126/science.1184939
141 Yanofsky, C. Attenuation in the control of expression of bacterial operons. *Nature* **289**, 751-758 (1981). https://doi.org/10.1038/289751a0
142 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004). https://doi.org/10.1101/gr.849004
143 Mudryi, V. *et al.* Kinetic mechanism and determinants of EF-P recruitment to translating ribosomes. *Nucleic Acids Res* **52**, 11870-11883 (2024). https://doi.org/10.1093/nar/gkae815
144 Mohapatra, S., Choi, H., Ge, X., Sanyal, S. & Weisshaar, J. C. Spatial distribution and ribosome-binding dynamics of EF-P in live *Escherichia coli*. *mBio* **8** (2017). https://doi.org/10.1128/mBio.00300-17
145 Starosta, A. L. *et al.* Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site. *Nucleic Acids Res* **42**, 10711-10719 (2014). https://doi.org/10.1093/nar/gku768
146 Kumari, R., Michel, A. M. & Baranov, P. V. PausePred and Rfeet: webtools for inferring ribosome pauses and visualizing footprint density from ribosome profiling data. *Rna* **24**, 1297-1304 (2018). https://doi.org/10.1261/rna.065235.117
147 Jenkins, Z. A., Hååg, P. G. & Johansson, H. E. Human eIF5A2 on chromosome 3q25-q27 is a phylogenetically conserved vertebrate variant of eukaryotic translation initiation factor 5A with tissue-specific expression. *Genomics* **71**, 101-109 (2001). https://doi.org/10.1006/geno.2000.6418
148 Morgan, G. J., Burkhardt, D. H., Kelly, J. W. & Powers, E. T. Translation efficiency is maintained at elevated temperature in *Escherichia coli*. *J Biol Chem* **293**, 777-793 (2018). https://doi.org/https://doi.org/10.1074/jbc.RA117.000284
149 Dai, X. *et al.* Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. *Nat Microbiol* **2**, 16231 (2016). https://doi.org/10.1038/nmicrobiol.2016.231

150 Schilling, B. *et al.* Global Lysine Acetylation in *Escherichia coli* Results from Growth Conditions That Favor Acetate Fermentation. *J Bacteriol* **201** (2019). https://doi.org/10.1128/jb.00768-18
151 Butler, C. A., Veith, P. D., Nieto, M. F., Dashper, S. G. & Reynolds, E. C. Lysine acetylation is a common post-translational modification of key metabolic pathway enzymes of the anaerobe *Porphyromonas gingivalis*. *J Proteomics* **128**, 352-364 (2015). https://doi.org/10.1016/j.jprot.2015.08.015
152 Nakayasu, E. S. *et al.* Ancient Regulatory Role of Lysine Acetylation in Central Metabolism. *mBio* **8** (2017). https://doi.org/10.1128/mBio.01894-17
153 Wolfe, A. J. Bacterial protein acetylation: new discoveries unanswered questions. *Curr Genet* **62**, 335-341 (2016). https://doi.org/10.1007/s00294-015-0552-4
154 Chohnan, S., Furukawa, H., Fujio, T., Nishihara, H. & Takamura, Y. Changes in the size and composition of intracellular pools of nonesterified coenzyme A and coenzyme A thioesters in aerobic and facultatively anaerobic bacteria. *Appl Environ Microbiol* **63**, 553-560 (1997). https://doi.org/10.1128/aem.63.2.553-560.1997
155 Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* **5**, 593-599 (2009). https://doi.org/10.1038/nchembio.186
156 Lammers, M. Post-translational Lysine Ac(et)ylation in Bacteria: A Biochemical, Structural, and Synthetic Biological Perspective. *Front Microbiol* **12**, 757179 (2021). https://doi.org/10.3389/fmicb.2021.757179
157 Kuhn, M. L. *et al.* Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. *PLoS One* **9**, e94816 (2014). https://doi.org/10.1371/journal.pone.0094816
158 Weinert, B. T. *et al.* Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. *Cell Rep* **4**, 842-851 (2013). https://doi.org/10.1016/j.celrep.2013.07.024
159 Qian, L. *et al.* Global Profiling of Protein Lysine Malonylation in *Escherichia coli* Reveals Its Role in Energy Metabolism. *J Proteome Res* **15**, 2060-2071 (2016). https://doi.org/10.1021/acs.jproteome.6b00264
160 Weinert, B. T. *et al.* Acetyl-phosphate is a critical determinant of lysine acetylation in *E. coli*. *Mol Cell* **51**, 265-272 (2013). https://doi.org/10.1016/j.molcel.2013.06.003
161 Klier, H., Wöhl, T., Eckerskorn, C., Magdolen, V. & Lottspeich, F. Determination and mutational analysis of the phosphorylation site in the hypusine-containing protein Hyp2p. *FEBS Lett* **334**, 360-364 (1993). https://doi.org/10.1016/0014-5793(93)80712-4
162 Kang, K. R., Kim, Y. S., Wolff, E. C. & Park, M. H. Specificity of the deoxyhypusine hydroxylase-eukaryotic translation initiation factor (eIF5A) interaction: identification of amino acid residues of the enzyme required for binding of its substrate, deoxyhypusine-containing eIF5A. *J Biol Chem* **282**, 8300-8308 (2007). https://doi.org/10.1074/jbc.M607495200
163 Ishfaq, M. *et al.* Acetylation regulates subcellular localization of eukaryotic translation initiation factor 5A (eIF5A). *FEBS Lett* **586**, 3236-3241 (2012). https://doi.org/10.1016/j.febslet.2012.06.042
164 Yanagisawa, T. *et al. Neisseria meningitidis* Translation Elongation Factor P and Its Active-Site Arginine Residue Are Essential for Cell Viability. *PLoS One* **11**, e0147907 (2016). https://doi.org/10.1371/journal.pone.0147907
165 Guo, Q. *et al.* Elongation factor P modulates *Acinetobacter baumannii* physiology and virulence as a cyclic dimeric guanosine monophosphate effector. *Proc Natl Acad Sci U S A* **119**, e2209838119 (2022). https://doi.org/10.1073/pnas.2209838119
166 Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**, 77-84 (2003). https://doi.org/10.1046/j.1365-2958.2003.03425.x
167 Pinheiro, B. *et al.* Elongation factor P is required for $EII^{Glc}$ translation in *Corynebacterium glutamicum* due to an essential polyproline motif. *Mol Microbiol* **115**, 320-331 (2021). https://doi.org/10.1111/mmi.14618

168 Novozhilov, A. S. & Koonin, E. V. Exceptional error minimization in putative primordial genetic codes. *Biol Direct* **4**, 44 (2009). https://doi.org/10.1186/1745-6150-4-44
169 Pelechano, V. & Alepuz, P. eIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic Acids Res* **45**, 7326-7338 (2017). https://doi.org/10.1093/nar/gkx479
170 Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **11**, 319-324 (2014). https://doi.org/10.1038/nmeth.2834
171 von der Haar, T. A quantitative estimation of the global translational activity in logarithmically growing yeast cells. *BMC Syst Biol* **2**, 87 (2008). https://doi.org/10.1186/1752-0509-2-87
172 Schmidt, A. *et al.* The quantitative and condition-dependent *Escherichia coli* proteome. *Nature Biotechnology* **34**, 104-110 (2016). https://doi.org/10.1038/nbt.3418
173 Papadimitriou, K. *et al.* Stress Physiology of Lactic Acid Bacteria. *Microbiol Mol Biol Rev* **80**, 837-890 (2016). https://doi.org/10.1128/mmbr.00076-15
174 Mookerjee, S. A., Gerencser, A. A., Nicholls, D. G. & Brand, M. D. Quantifying intracellular rates of glycolytic and oxidative ATP production and consumption using extracellular flux measurements. *J Biol Chem* **292**, 7189-7207 (2017). https://doi.org/10.1074/jbc.M116.774471
175 Pu, Y. *et al.* ATP-Dependent Dynamic Protein Aggregation Regulates Bacterial Dormancy Depth Critical for Antibiotic Tolerance. *Mol Cell* **73**, 143-156.e144 (2019). https://doi.org/10.1016/j.molcel.2018.10.022
176 Peil, L. *et al.* Distinct X/PP/X sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc Natl Acad Sci U S A* **110**, 15265-15270 (2013). https://doi.org/10.1073/pnas.1310642110
177 Roy, H. *et al.* The tRNA synthetase paralog PoxA modifies elongation factor-P with (*R*)-β-lysine. *Nat Chem Biol* **7**, 667-669 (2011). https://doi.org/10.1038/nchembio.632
178 Brochier, C., López-García, P. & Moreira, D. Horizontal gene transfer and archaeal origin of deoxyhypusine synthase homologous genes in bacteria. *Gene* **330**, 169-176 (2004). https://doi.org/10.1016/j.gene.2004.01.018
179 López-García, P., Brochier, C., Moreira, D. & Rodríguez-Valera, F. Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ Microbiol* **6**, 19-34 (2004). https://doi.org/10.1046/j.1462-2920.2003.00533.x
180 Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279-285 (2016). https://doi.org/10.1093/nar/gkv1344
181 Hersch, S. J., Elgamal, S., Katz, A., Ibba, M. & Navarre, W. W. Translation initiation rate determines the impact of ribosome stalling on bacterial protein synthesis. *J Biol Chem* **289**, 28160-28171 (2014). https://doi.org/10.1074/jbc.M114.593277
182 Myka, K. K. *et al.* Inhibiting translation elongation can aid genome duplication in *Escherichia coli*. *Nucleic Acids Res* **45**, 2571-2584 (2017). https://doi.org/10.1093/nar/gkw1254
183 Saito, K. *et al.* Ribosome collisions induce mRNA cleavage and ribosome rescue in bacteria. *Nature* **603**, 503-508 (2022). https://doi.org/10.1038/s41586-022-04416-7
184 Russell, J. B. & Cook, G. M. Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol Rev* **59**, 48-62 (1995). https://doi.org/10.1128/mr.59.1.48-62.1995
185 Mudryi, V. *Elongation factor P: mechanism of action and opportunities for drug design*, Georg-August Universität Göttingen, (2023).
186 Katoh, T. & Suga, H. Ribosomal incorporation of consecutive β-amino acids. *J Am Chem Soc* **140**, 12159-12167 (2018). https://doi.org/10.1021/jacs.8b07247
187 Katoh, T., Tajima, K. & Suga, H. Consecutive elongation of D-amino acids in translation. *Cell Chem Biol* **24**, 46-54 (2017). https://doi.org/10.1016/j.chembiol.2016.11.012

## Danksagung

Abschließend möchte ich mich bei Allen bedanken, die zur Entstehung dieser Arbeit beigetragen und mich auf diesem Weg unterstützt haben.

An erster Stelle – und das vollkommen zurecht – möchte ich PD Dr. Jürgen Lassak für die Betreuung danken. Seit meiner Bachelorarbeit, über die Masterarbeit, bis hin zu meiner Doktorarbeit wurde ich immer mit wertvollen Ideen, lehrreichen Beiträgen und unermüdlicher Begeisterung unterstützt und gefördert. Ich hatte eine großartige Zeit in der Gruppe und habe wertvolle Erfahrungen sammeln können.

Ein besonderer Dank gilt allen Mitgliedern meiner Prüfungskommission, insbesondere Prof. Dr. Dmitrij Frishman als Zweitgutachter und für die Zusammenarbeit in verschiedenen Projekten. Ebenso möchte ich mich bei unseren weiteren Kollaborationspartnern Dr. Marina Parr, Dr. Fei Qi, Prof. Dr. Andreas Schlundt und Dr. Pavel Kielkowski bedanken. Die unkomplizierte und effiziente Art der Zusammenarbeit war eine große Bereicherung und ein Vorbild, an dem ich mich auch in Zukunft orientieren möchte.

Ein großes Dankeschön geht an Grazyna für die gesamte Organisation abseits des Labors und ihren Einsatz für alle Anliegen und Probleme, ohne dabei den Überblick zu verlieren. Tatjana, Slavica, Korinna, Sabine, Agata, Michelle und Ingrid möchte ich für die Hilfe im Labor danken, die meinen Alltag enorm erleichtert haben.

Ich danke allen Mitgliedern (und Ehemaligen) unserer kleinen Arbeitsgruppe für die stetige Unterstützung und Zusammenarbeit. Giovanni, mit seiner unvergleichlichen Hilfsbereitschaft, Erica, mit ihrer bemerkenswerten Genauigkeit, und Tess, mit ihrer beeindruckenden Gelassenheit – hier habe ich unglaublich viel gelernt. Auch meinen Studentinnen Anja, Nina, Bhavna, Niran und Eli danke ich für ihre Neugier und Motivation, die unsere Zusammenarbeit so bereichernd gemacht hat. Ein besonderer Dank geht an Ralph, sowie Judith, Franzi und Wolfram, die mich bei meinem Einstieg in die Mikrobiologie unterstützt und begleitet haben. Die letzten Jahre sind wie im Flug vergangen, nicht zuletzt dank meiner wundervollen Bürokolleginnen Nicola und Elena, die sowohl wissenschaftliche als auch weniger wissenschaftliche Gespräche stets zu einem Highlight des Tages gemacht haben. Für die kleinen, aber so wichtigen Pausen und guten Gespräche zwischendurch danke ich Steffi, Urte, Kilian, Sophie, Bibakhya and Leo. Besonders möchte ich mich auch bei Sophia bedanken, mit der das Lachen nie zu kurz kam und die immer für mich da war. Ich bin sehr dankbar, dass ich die letzten Jahre mit so großartigen Kollegen aus den AGs K. Jung, H. Jung und Heilbronner zusammenarbeiten durfte und dadurch so wundervolle Freunde finden konnte.

Ganz besonders möchte ich meiner Familie und Freunden danken, die mich von Anfang an bedingungslos unterstützt und den nötigen Ausgleich geschaffen haben. Ich bin meinen Eltern unendlich dankbar – sie haben mir immer den Rücken freigehalten und mir alle Möglichkeiten geschaffen, sodass ich heute an dieser Stelle stehen kann. Auch bei Tobias, meinem Bruder, möchte ich mich bedanken. Er ist für mich ein verlässlicher Rückhalt, auf den ich mich jederzeit verlassen kann. Zuletzt möchte ich Phil danken, der mich mit unfassbarer Geduld, Ruhe und Motivation begleitet und immer an mich geglaubt hat. Danke für diese Stärke!

Ohne diese Unterstützung hätte ich das Alles nicht geschafft!

## Curriculum Vitae

### PERSONAL INFORMATION

| | |
|---|---|
| Name | Alina Maria Sieber |
| languages | german – first language<br>english – very good knowledge |

### HIGHER EDUCATION

| | |
|---|---|
| 2021-2025 | **Doctoral Studies "Biology"**<br>Ludwig-Maximilians-Universität München, Germany |
| Project for Doctoral Thesis: | The problem with proline – Regulation of ribosome stalling by translation factors and codon choice |
| 2018-2021 | **Master of Science "Molecular and Cellular Biology"**<br>Ludwig-Maximilians-Universität München, Germany |
| Project for Master Thesis: | Artificial evolution of lysyl-tRNA-synthetase paralogs towards non-cognate target recognition |
| 2015-2018 | **Bachelor of Science** "**Biology"**<br>Ludwig-Maximilians-Universität München, Germany |
| Project for Bachelor Thesis: | Comparative analyses of phylogenetically distinct EF-P species at ribosome stalling on consecutive prolines and non-polyproline motifs |

### CONFERENCES

| | |
|---|---|
| 2024 (Rouen, FRA) | **5th International Conference on Post Translational Modifications in Bacteria (PTMBact)**<br>• Short talk: "The Role of EF-P Post-translational Modifications in Bacterial Translation Stress Response" |
| 2024 (Michelsberg, GER) | **Symposium on Regulation in Microbes (RiM2024)**<br>• Short talk: "Regulation of Translational Stress Response by Elongation Factor P and EF-P Like Proteins" |
| 2023 (Göttingen, GER) | **Annual Conference 2023 of the Association for General and Applied Microbiology (VAAM)**<br>• Poster presentation: "Lost in Translation? - The Role of EfpL in Efficient Protein Synthesis in Bacteria"<br>• VAAM Poster Award 2023 |
| 2023 (Heidelberg, GER) | **EMBL Conference: Protein Synthesis and Translational Control**<br>• Poster presentation: "From Stalling to Success: Discovering the Role of EfpL in Efficient Protein Synthesis in Bacteria" |
| 2022 (virtual) | **Annual Conference 2022 of the Association for General and Applied Microbiology (VAAM)**<br>• Poster presentation: "Physiological Significance of the EF-P Paralog EfpL" |