## ARTICLE

Check for updates

# 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.

## 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 CCC/**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
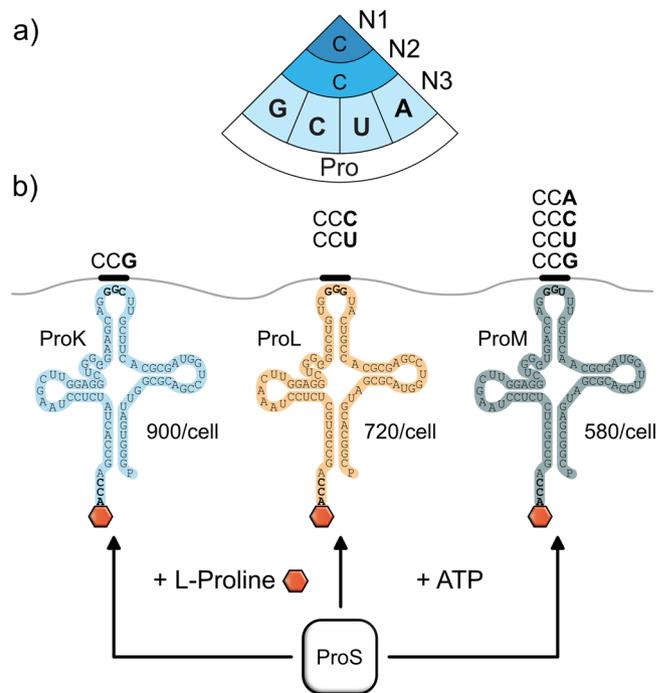


**Fig. 1 Diversity of proline codons and their corresponding tRNAs. a** The genetic code contains four codons for proline: CCG, CCC, CCU, and CCA. **b** The three tRNAs ProK, ProL, and ProM recognize distinct sets of proline codons and exhibit different levels of abundance within the cell[34]. All three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

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
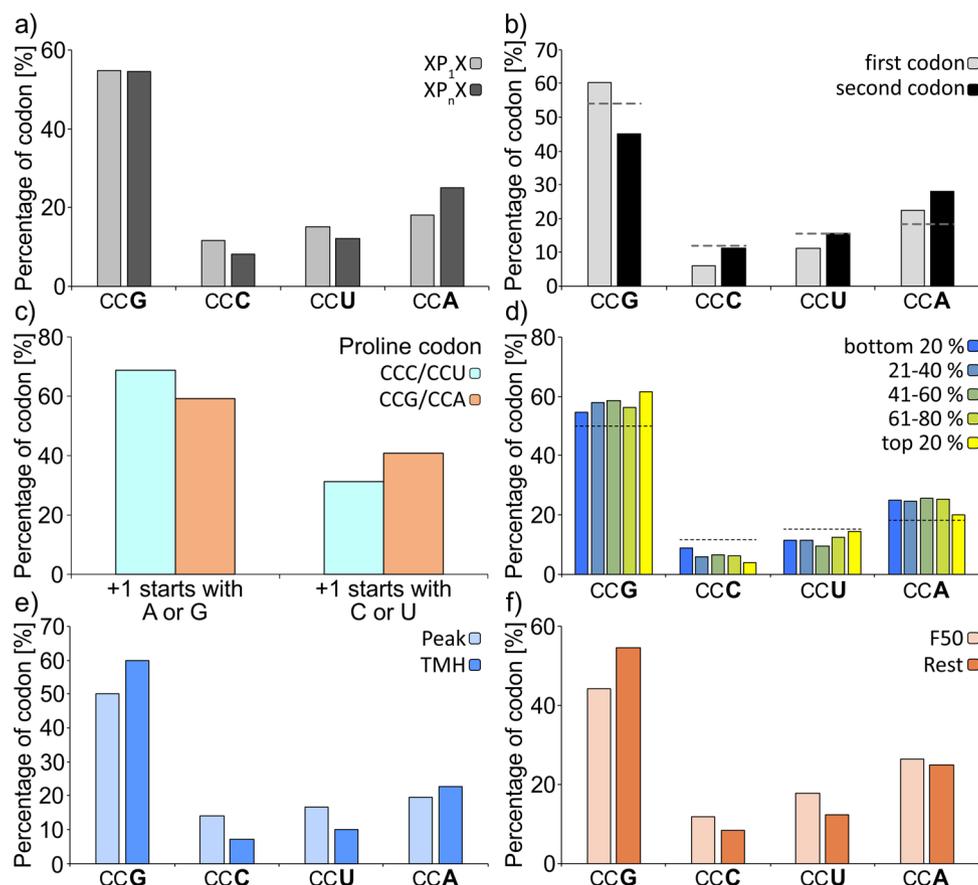
**Fig. 2 Bioinformatic analysis of proline codon bias in *E. coli*. a** Codon usage of either single (XP$_1$X) or consecutive (XP$_n$X) prolines (with X being any amino acid except proline and n > 1). p value = 1.7e−30, chi-squared test. **b** Codon usage of the first and second proline in PP-motifs. 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 CCG, TPPP CCG, FPPP CCG, and RPPP CCG 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
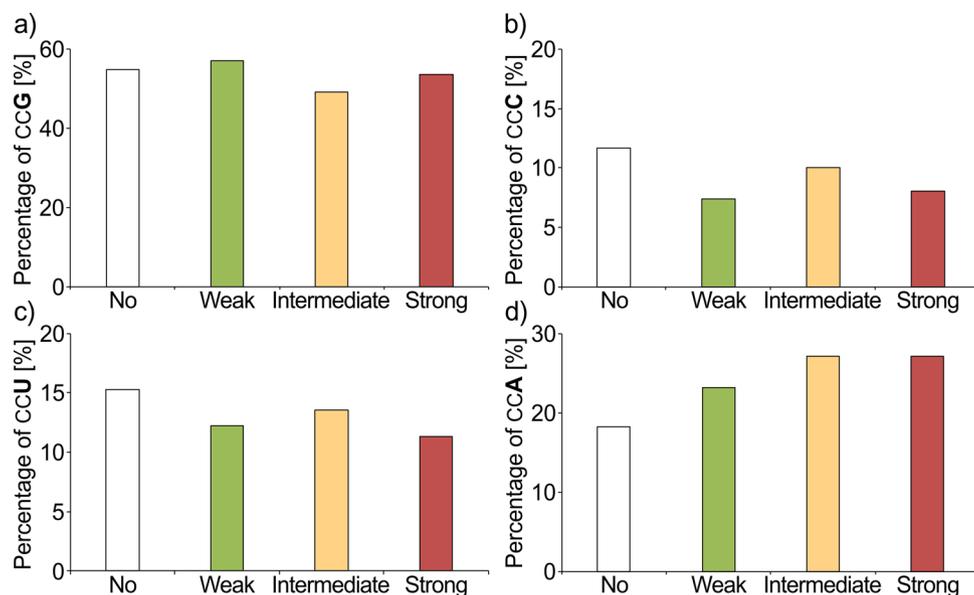
**Fig. 3 Codon usage in PP-motifs of different pausing strength.** Pausing strength of PP-motifs depends on the upstream amino acid context[11,20,22] resulting in weak, intermediate, and strong pausers. The pausing strength resulting from amino acid context is indicated by colored bars (no pausing—white; 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.
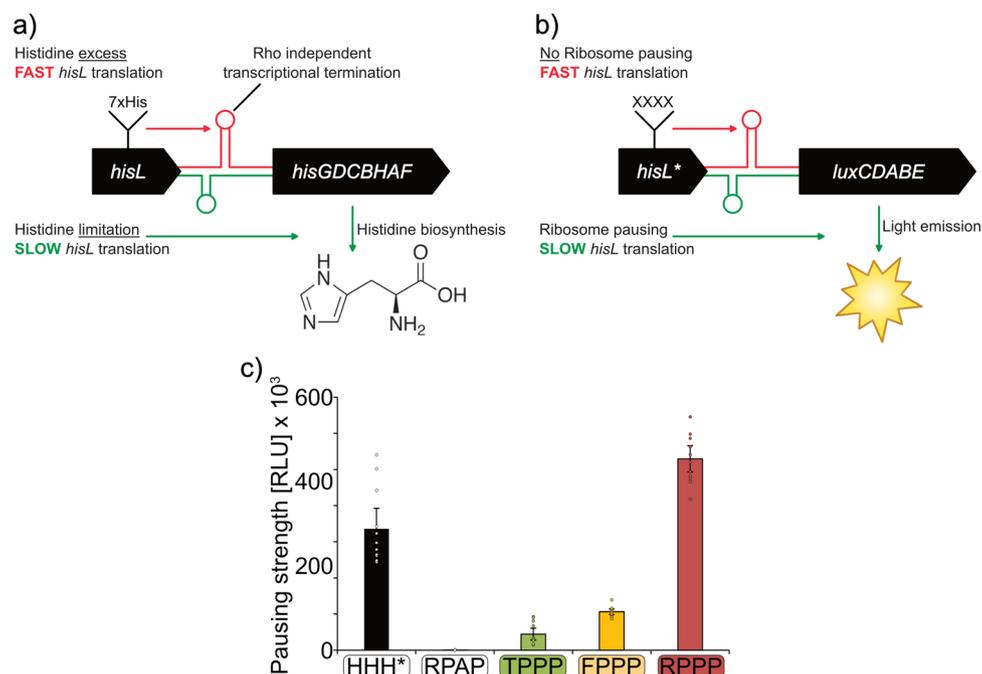


**Fig. 4 The His-pausing system for in vivo measurement of pausing strength. a** Architecture of the histidine biosynthesis operon in *E. coli*. In its native state, the histidine biosynthesis gene cluster (*hisGDCBHAF*) is regulated by the His-leader peptide (*hisL*). This peptide contains seven consecutive 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.
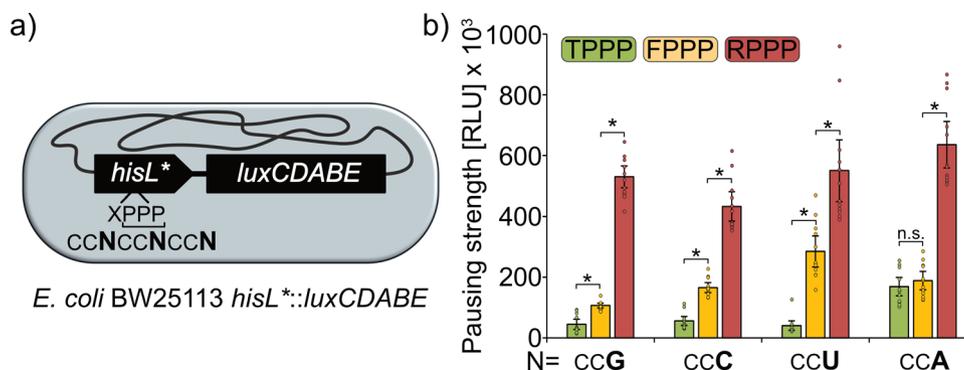
**Fig. 5 Codon-dependent pausing strength at weak, intermediate, and strong PP-motifs. a** Genomic organization of the HisL*_Lux reporter. Synthetic His-Leader peptides (HisL*) preceding the *lux* genes (*luxCDABE*) were genomically integrated at the *his*-locus. In *hisL**, His1 one was replaced by 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 $^{XPPP}$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 ($^{TPPP}$CCG, $^{TPPP}$CCC, $^{TPPP}$CCU, $^{TPPP}$CCA) (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
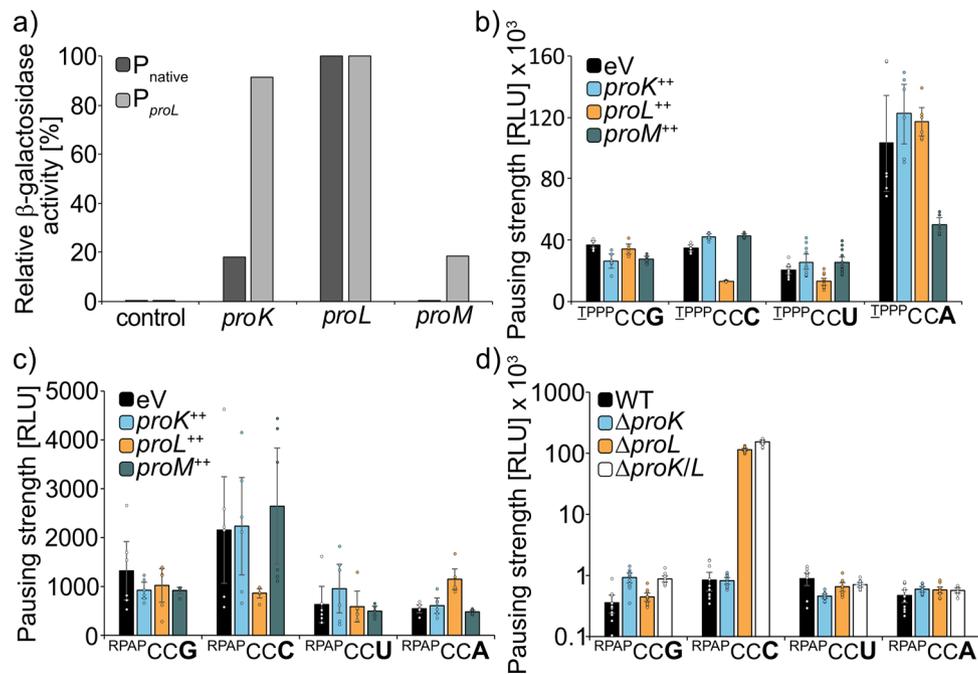
**Fig. 6 Influence of prolyl-tRNA copy number on the codon-dependent pausing strength at PP-motifs. a** Approximation of *E. coli* BW25113 cells carrying the weak HisL*_Lux operon (TPPP) with different proline codon usage were transformed with pBBR1 MCS4-*lacZ* plasmids encoding ProK, ProL, or ProM under 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.
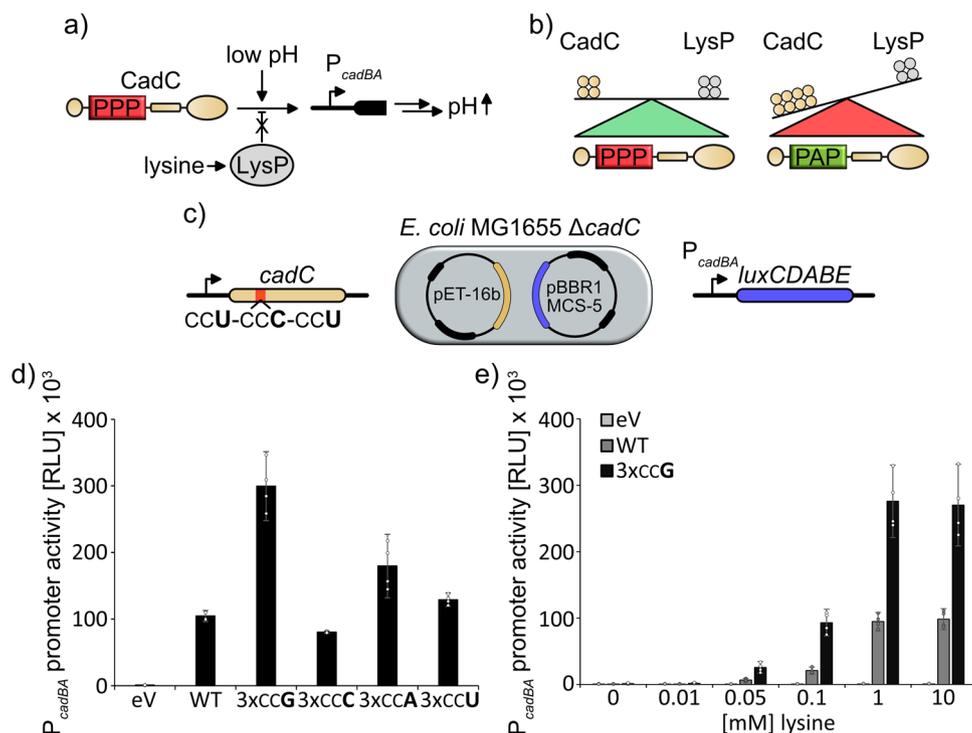
**Fig. 7 Codon choice modulates protein expression and ensures physiological protein stoichiometry of the Cad system. a** The Cad system. CadC is a pH sensor that induces expression of its target genes at low pH by binding to the *cadBA* promoter ($P_{cadBA}$). Expression of the corresponding gene products 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 $3 \times CCG$ 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 SphI and NcoI. 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 Δefp 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 ATTAACC<u>ATGGGGNNN*TAG***G**ACTAAAAAAATTTCATTC</u> (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$_2$HPO$_4$, 0.04 M NaH$_2$PO$_4$, 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$_2$CO$_3$ 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 $20,000 \times g$ 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 = $1000 \times \text{Abs}_{420} \times t^{-1} \times V^{-1} \times \text{Abs}_{600}^{-1}$,[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.

## 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. Peptidyl-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).

## 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.

## 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.