

# Codon usage and protein length-dependent feedback from translation elongation regulates translation initiation and elongation speed

Xueliang Lyu<sup>1,2</sup>, Qian Yang<sup>1</sup>, Fangzhou Zhao<sup>1</sup> and Yi Liu<sup>1,\*</sup>

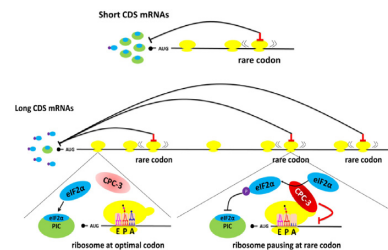
<sup>1</sup>Department of Physiology, The University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390, USA and <sup>2</sup>State Key Laboratory of Agricultural Microbiology, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, Hubei 430070, China

Received April 16, 2021; Revised July 26, 2021; Editorial Decision August 09, 2021; Accepted August 17, 2021

## ABSTRACT

Essential cellular functions require efficient production of many large proteins but synthesis of large proteins encounters many obstacles in cells. Translational control is mostly known to be regulated at the initiation step. Whether translation elongation process can feedback to regulate initiation efficiency is unclear. Codon usage bias, a universal feature of all genomes, plays an important role in determining gene expression levels. Here, we discovered that there is a conserved but codon usage-dependent genome-wide negative correlation between protein abundance and CDS length. The codon usage effects on protein expression and ribosome flux on mRNAs are influenced by CDS length; optimal codon usage preferentially promotes production of large proteins. Translation of mRNAs with long CDS and non-optimal codon usage preferentially induces phosphorylation of initiation factor eIF2 $\alpha$ , which inhibits translation initiation efficiency. Deletion of the eIF2 $\alpha$  kinase CPC-3 (GCN2 homolog) in *Neurospora* preferentially up-regulates large proteins encoded by non-optimal codons. Surprisingly, CPC-3 also inhibits translation elongation rate in a codon usage and CDS length-dependent manner, resulting in slow elongation rates for long CDS mRNAs. Together, these results revealed a codon usage and CDS length-dependent feedback mechanism from translation elongation to regulate both translation initiation and elongation kinetics.

## GRAPHICAL ABSTRACT



## INTRODUCTION

Degeneracy of genetic codons allows 18 of 20 standard amino acids to be encoded by two to six synonymous codons. The preference for specific synonymous codons is a universal feature of all genomes (1–3). Although synonymous codons encode the same amino acid and synonymous mutations are often referred to as ‘silent’, genetic, biochemical, molecular, and bioinformatics studies have demonstrated that codon usage can influence gene expression levels and protein structures (3–9). Preferred codons (also called optimal codons) are enriched in genes that encode highly expressed proteins, and codon optimization can strongly increase endogenous and heterologous gene expression in diverse eukaryotes and prokaryotes (10–20). Codon usage regulates the speed of translation elongation during mRNA translation: Preferred codons speed up elongation, whereas rare codons slow down elongation and can cause ribosome pausing (7,21–24). The role of codon usage in translation kinetics regulates translation efficiency in part because rare codons can slow elongation and potentially cause premature translation termination (23). Codon usage can also impact mRNA stability and translation fidelity (5,25–28). In addition, the codon usage-dependent translation kinetics also regulates co-translational protein folding and protein functions in prokaryotic and eukaryotic organisms (6,16,21,29–32). Furthermore, there are translation-independent effects of codon usage due to the roles of

\*To whom correspondence should be addressed. Tel: +1 214 645 6033; Fax: +1 214 645 6049; Email: Yi.Liu@UTsouthwestern.edu

nucleotide sequence in transcription, chromatin structure, splicing, mRNA structure, transcriptional termination and mRNA localization (7,13,28,33,34).

Synthesis of large proteins encounters many obstacles in cells. Errors in transcriptional, posttranscriptional and translational processes are expected to increase as coding sequence (CDS) length increases. Long mRNAs and large proteins are more likely to be degraded or damaged than shorter ones (35). In addition, large proteins have more protein domains, resulting in increased complexity in the protein folding process and greater likelihood of misfolding than small proteins. Previous studies have shown that CDS length negatively correlates with protein abundance, translation initiation rate, and ribosome density (22,36–40), suggesting the existence of mechanisms that preferentially inhibit translation of large proteins. Modeling studies suggest that the negative influence of CDS length on translation is likely due to less efficient ribosome recycling on mRNAs with longer compared to shorter CDS regions (39,41). Despite these issues, many large proteins are critical for cellular functions in diverse biological processes. The mechanisms that allow efficient production and proper folding of large proteins are not clear.

The best characterized mechanism that regulates translation efficiency is the initiation process, which largely determines the number of protein molecules that can be made from an individual mRNA transcript (42–44). Ribosome queuing near AUG start codon caused by ribosome stalling or collision impacts translation initiation efficiency (45,46). A previous study using a reporter gene in *Saccharomyces cerevisiae* suggested that rare codons near the start codon could inhibit translation initiation probably due to ribosome queuing near the start codon, whereas optimal codons near the start codon presumably result in rapid liberation of the start codon and therefore high translation initiation rates (47). However, other studies showed that codon usage near the start codon appears to influence translation initiation rate due to its effects on mRNA structures rather than translation elongation (48,49). Thus, the mechanism underlying the coordination between translation initiation and elongation under nutrient replete growth conditions and the role of codon usage in this process are still unclear.

In eukaryotes, translation initiation begins with the binding of the ternary complex (the aminoacylated initiator methionyl-tRNA (Met-tRNA<sub>i</sub>), GTP and the initiation factor 2 (eIF2)) to the 40S ribosome to form the pre-initiation complex (42–44,50). Phosphorylation of eIF2 $\alpha$ , a subunit of eIF2, at serine 51 is an important regulator of translation initiation and is known to be induced by many types of stress conditions to result in global inhibition of translation initiation. The phosphorylation of eIF2 $\alpha$ -GDP inhibits the guanine nucleotide exchange activity of eIF2B and blocks the recycling of unphosphorylated eIF2 $\alpha$ -GDP into the translationally active form eIF2 $\alpha$ -GTP. In higher eukaryotes, GCN2 is one of the several kinases responsible for the phosphorylation of eIF2 $\alpha$  at serine 51 after its activation from an autoinhibited state (42,50,51). In fungi, however, the GCN2 homolog is the major and the only known eIF2 $\alpha$  kinase responsible for eIF2 $\alpha$  phosphorylation at serine 51 (52). Upon nutrient limitation or amino acid starvation or other stress conditions, GCN2 is activated and phos-

phorylates eIF2 $\alpha$  to initiate the adaptive pro-survival integrated stress response, resulting in temporary translation repression of most mRNAs and activation of amino acid biosynthesis (53–56). In addition, mutation or depletion of enzymes required for tRNA modification can also trigger eIF2 $\alpha$  phosphorylation (57,58). Although GCN2 can be activated by interacting with uncharged tRNA caused by amino acid starvation, recent evidence also suggests that GCN2 can also be activated by other mechanisms such as by interacting with the ribosomal P-stalk (59–61). Although the stress-induced eIF2 $\alpha$  phosphorylation is expected to cause global translation repression, recent studies showed that this is not the case for low levels of eIF2 $\alpha$  phosphorylation, suggesting that, under certain conditions, eIF2 $\alpha$  phosphorylation has specific rather than a broad inhibitory effect on general translation (50,62).

The filamentous fungus *Neurospora crassa* exhibits a strong codon usage bias for C/G at wobble positions and has been an important experimental model system for studying the functions of codon usage biases (6,7,16). We have previously shown that codon usage plays an important role in regulating elongation speed and the co-translational protein folding process in *Neurospora* (9,16,21,63). Use of preferred codons speeds up the local rate of translation elongation while rare codons slow down translation elongation and potentially result in ribosome pausing and premature termination, a mechanism that can affect translation efficiency (21,23). We also showed that codon usage could influence gene expression levels by affecting transcription efficiency (13,33,64). These results led us to propose that codon usage represents a code within the genetic codons that regulates both gene expression level and protein structure.

In this study, we used *N. crassa* as a model system to understand the CDS length-dependent mechanism that regulates translation. We discovered that there is a conserved codon usage-dependent genome-wide negative correlation between protein abundance and length, suggesting that optimal codon usage is a mechanism that allows for efficient production of large proteins critical for cell functions. We found that translation of mRNAs with non-optimal codon usage preferentially induced eIF2 $\alpha$  phosphorylation and reduced protein levels in a CDS length-dependent manner, indicating a feedback mechanism from translation elongation to control translation initiation. Furthermore, we showed that the GCN2 homolog, CPC-3, which is the only known kinase responsible for eIF2 $\alpha$  phosphorylation in *Neurospora* (52), also regulates translation elongation rate in both codon usage and CDS length-dependent manners, resulting in slow elongation rates for mRNAs with long CDS regions. Together, these results revealed a codon usage- and CDS length-based feedback mechanism from translation elongation to regulate both translation initiation and elongation kinetics.

## MATERIALS AND METHODS

### *Neurospora* strains and growth conditions

The *N. crassa* wild-type (WT) strain FGSC 4200 (a) and the  $\Delta cpc-3$  strain (65) were used in this study. Strains were cultured on slants containing 1  $\times$  Vogel's, 3% sucrose, and 1.5%

agar. Liquid cultures were grown in 2% glucose medium (1 × Vogel's, 2% glucose). Specifically, fresh conidia (7–10 days post inoculation on slants) of the WT or  $\Delta cpc-3$  strains were cultured in 50 ml 2% glucose medium in petri dishes at room temperature for 2 days. The cultures were cut into small discs with a diameter of 1 cm, and then the discs were transferred into flasks with the same liquid medium and were grown with orbital shaking (200 rpm) for 12 h before various experiments. Race tube medium contained 1 × Vogel's, 0.1% glucose, 0.17% arginine, and 1.5% agar. All the strains were cultured under constant light at room temperature unless otherwise specified. For chemical treated experiments, various indicated drugs were added into 2% glucose medium for the WT strain for 20 min before harvesting. 3-aminotriazole (3-AT, Sigma-Aldrich, Cat. No. 8056), puromycin (Puro, Sigma-Aldrich, Cat. No. 540411) and tigecycline (TIG, Sigma-Aldrich, Cat. No. PZ0021) were used at the final concentration of 50 mM, 0.06 mg/mL and 0.5 mg/mL, respectively. Anisomycin (ANS, Sigma-Aldrich, Cat. No. A9789) and cycloheximide (CHX, Sigma-Aldrich, Cat. No. C1988) were used at the indicated concentration.

#### Plasmid construction, transformation and $\Delta cpc-3$ complementation

For gene expression at the *csr-1* locus in *N. crassa*, a basta-resistance (*bar*) gene was inserted downstream of the *ccg-1* promoter (Pccg-1) of a parental plasmid, Pcsr1, to create the Pcsr1-*bar* plasmid. Pcsr1-*bar* is a *csr-1*-targeting expression vector with an expression cassette in which Pccg-1 and *bar* flank the gene of interest (*edp* or *cpc-3* in this study), and this cassette is flanked by *csr-1* downstream and upstream fragments which serve as the recombination sites for double homologous recombination (66). Afterwards, the resulting plasmid was transformed into *N. crassa* strains by electroporation, transformants were screened for resistance to both glufosinate-ammonium (0.25 mg/ml, Sigma-Aldrich, Cat. No. 45520) and to cyclosporin A (5 μg/ml, Sigma-Aldrich, Cat. No. 30024), which resulted in >90% positive transformants. Homokaryotic strains were obtained by microconidia purification. To generate the  $\Delta cpc-3$  complementation strains, a construct expressing the WT *cpc-3* (cloned from genomic DNA) with 3× Flag epitope tag under the control of the *ccg-1* promoter was introduced at the *csr-1* locus in the  $\Delta cpc-3$  strain. The expression of the Flag-tagged CPC-3 and the rescue of eIF2α phosphorylation in the complementation strains were confirmed by immunoblotting (Supplementary Figure S1).

#### RNA isolation and quantitative reverse transcription PCR (qRT-PCR)

The culture conditions were the same as described above. RNA extraction and qRT-PCR were performed as previously described (58).  $\beta$ -tubulin transcript (*NCU04054*) was quantified as an internal control. The primer pairs 5'-ACAACCCCTCACATCAACCAA-3', 5'-CCGCCCTTGTCATCGTCATCC-3' and 5'-GCGTATCGGCGAGCAGTT-3', 5'-CCTCACCAGTGTACCAATGCA-3' were used to amplify the

reporter gene *edp* and  $\beta$ -tubulin gene, respectively. The primers for different versions of *edp* were designed to amplify the 5' end region of the transcripts (5'UTR, 3× Flag and 8xGly linker), which is common to all the transgenes to ensure the same amplification efficiency.

#### *In vitro* translation and protein analyses

*In vitro* translation assay was performed as previously described (21). Specifically, the *N. crassa* cell free lysate was obtained as previously described (21,67), except that the protease inhibitor cocktail from MedChemExpress company (Cat. No. HY-K0010) was used. Equal moles (0.65 pmol for each reaction) of different versions of mRNAs were individually added into *N. crassa* cell free lysate to translate for 15 min at 26°C unless otherwise specified. SDS-PAGE loading buffer was added into the samples immediately and followed by heating the samples at 98°C for 10 min.

Protein extraction was performed as previously described (58). For western blot analysis of the Flag-tagged EDP, the anti-Flag (Sigma-Aldrich, Cat. No. F3165) antibody was used. Densitometric analyses of the western blot results were performed using Image J.

For phosphatase treatment, total proteins from the WT and  $\Delta cpc-3$  strains were obtained by using protein lysate buffer with or without phosphatase inhibitors (PP inhibitors: 25 mM NaF, 10 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub>·10H<sub>2</sub>O, 2 mM Na<sub>3</sub>VO<sub>4</sub> and 1 mM EDTA). The protein extracts from lysate buffer without PP inhibitors were further treated with Lambda protein phosphatase (Lambda PP, NEB, Cat NO.: P0753S) according to its protocol.

#### Cell-free translation assay to determine TFAs

To prepare the mRNA templates for *in vitro* translation, *in vitro* transcription of different mRNAs was performed as previously described (21). *In vitro* translation assays to determine TFAs were performed as previously described (21) and the luminescence of luciferase signal was recorded continuously in 20-s intervals.

#### Ribosome profiling and mRNA-seq

The WT and  $\Delta cpc-3$  strains expressing 1× WT-EDP, 3× WT-EDP, 1× OPT-EDP and 3× OPT-EDP, respectively, were used for ribosome profiling and accompanying mRNA-seq experiments under nutrient replete condition at room temperature (1 × Vogel's, 2% glucose). At least three biological replicates for the ribosome profiling experiment for each strain were used. Ribosome profiling experiments were performed as previously described (58). CHX was not added into cultures before sample collection and was only added into the lysate buffer at final concentration of 0.1 mg/ml. The ribosome profiling and mRNA-seq methods were described in the protocol for ARTseq Ribosome Profiling Kit (Yeast) (Illumina, Cat NO.: RPYSC12116). The adaptor in the kit was replaced with synthesized 5'-/5rApp/NNNNAGATCGGAAGAGCACACGTCT/3ddC/ to avoid potential adaptor ligation biases. The RT-primer in the kit was replaced with synthesized 5'-/5Phos/RNCGTTCGGACTGTAGAACTCTG/i



Sp18/AGACGTGTGCTCTTCCGATCT to avoid potential ligation biases during the circularization process. The resulting libraries were sequenced by the BGI DNBseq platform.

### Bioinformatics analyses of ribosome profiling results

The workflow for ribosome profiling experiment data analysis is the same as previously described (58), with the following modifications: Both of the ribo-seq and RNA-seq reads were mapped to CDS regions of genes. Both ribosome protected fragment (RPF) and mRNA level for each gene were measured by Transcripts Per Kilobase Million (TPM). Ribosome density was measured by TPM of RPFs normalized by TPM of mRNAs. The calculation of the relative codon decoding time (RCDT) is the same as previously described (21,23). The raw reads of biological replicates of each strain were merged when we analyzed the ribosome profiling data. Ribo-seq and RNA-seq data from WT and  $\Delta cpc-3$  strains expressing 1× WT-EDP, 3× WT-EDP, 1× OPT-EDP and 3× OPT-EDP, respectively, were regarded as four independent biological replicates when we performed genome-wide analyses for the WT and  $\Delta cpc-3$  strains.

### Metabolic labeling

Fresh *Neurospora* conidia of the WT and  $\Delta cpc-3$  strains were cultured separately in 50 ml 2% sucrose medium (1 × Vogel's, 2% sucrose) in flasks with orbital shaking (200 rpm). After culturing at 30°C for 8 h, EasyTag L-[<sup>35</sup>S]-methionine (PerkinElmer) was added into the medium for 45 min before sample collection. The same amounts of protein extracts (100 μg) from each sample were used to determine the levels of <sup>35</sup>S incorporation as previously described (68).

### Polysome profiling

The culture condition for polysome profiling is the same as the metabolic labeling experiment. Cultures of the WT and  $\Delta cpc-3$  strains were frozen in liquid nitrogen immediately after collection. Tissue samples were grounded into powder in liquid nitrogen and equal volume of the tissue powder of each sample was added the same volume of lysis buffer (1 × polysome buffer in ARTseq Ribosome Profiling Kit (Illumina, Cat. No. RPYSC12116), 1% Triton X-100, 0.1 mg/ml CHX, 1× protease inhibitor cocktail (EDTA-free, MedChemExpress, Cat. No. HY-K0010), 0.2 U/μl SUPERase•In (Invitrogen, Cat. No. AM2694) and 2 mM DTT). The lysates were then centrifugated at 15 000 rpm for 10 min before the  $A_{260}$  of the supernatant was measured by NanoDrop Microvolume Spectrophotometer. The  $A_{260}$ /ml of the lysate was calculated according to the protocol for the ARTseq Ribosome Profiling Kit. The same OD amount (20 OD<sub>260nm</sub>) of the lysate for each sample was loaded onto 10–50% sucrose gradient buffer containing 20 mM HEPES (pH 7.6), 0.1 M KCl, 5 mM MgCl<sub>2</sub>, 10 μg/ml CHX, the 1× protease inhibitor cocktail and 10 units/ml SUPERase•In. The sucrose gradients were then centrifuged at 35 000 rpm for 2 h at 4°C using a SW41Ti rotor in a Beckman Coulter

(Optima L-80 ultra) centrifuge. Sucrose gradients were analyzed using a BioLogic LP chromatography System (BioRad, Cat. No. 731-8350).

### Mass spectrometry analyses

Cell culturing was performed as described in the metabolic labeling experiment above. For mass spectrometry (MS) analysis to compare the relative amounts of different proteins within the same sample, ~100 μg proteins for each sample from the WT strain were run 1 cm into 7.5% SDS-PAGE gel. Gel slices were cut into small pieces for quantitative MS analysis. MaxQuant was used to analysis the MS data (69), and an intensity-based absolute quantification (iBAQ) value was used as a measure of protein abundance (70–72). For the quantitative MS analysis to compare the protein levels between the  $\Delta cpc-3$  strain and its complementation strain, the cultures were grounded into powder in liquid nitrogen before adding equal volume of lysis buffer (50 mM triethylammonium bicarbonate (TEAB) and 5% SDS) to equal volume of the powder for each sample. The protein extracts were centrifuged at 12 000 rpm for 10 min and the supernatants were used for the subsequent TMT Mass Tagging and MS analyses. The protein concentration for each sample was measured and equal amounts of total proteins were used during the experiment. There are four repeats for each strain. The abundance data of the quantitative MS (TMT labeling) analysis were not normalized by molecular weights. The protein levels were determined by the abundance data normalized by the sum of all the raw abundance data in each sample. The *P*-values for the four replicates were determined by Student's two-tailed *t*-test, which were further adjusted by False Discovery Rate (FDR). Those with FDR values <0.05 were identified as differentially expressed proteins, either significantly up-regulated or down-regulated. All the MS analyses were performed by the UT Southwestern Proteomics facility.

### Codon manipulation, indices calculation and data collection from databases

The codon usage of luciferase and *edp* genes were optimized based on the *N. crassa* codon usage frequency from the Codon Usage Database (<https://www.kazusa.or.jp/codon/>). The WT and optimized (OPT) versions of luciferase genes are the same as previously described (21). The sequences of WT/OPT-*Luc*, WT/OPT-*edp*, and WT/OPT-*GFP* were shown in Supplementary Figure S2. The CBIs and tAIs were calculated using CodonW (<http://codonw.sourceforge.net/>) and stAICalc (73), respectively. The tRNA copy number-related data used for calculating tAIs was collected from GtRNAdb database (<http://gtrnadb.ucsc.edu/GtRNAdb2/>). For the protein abundance data, the ePAI data of *N. crassa* was obtained previously in our lab (13) and the iBAQ data is obtained from this study. The publicly available protein abundance data of *S. cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans* and *Mus musculus* were used in the analyses (74–77).

## Gene functional enrichment analysis

The functional category (including Gene Ontology (GO), Interpro, and KEGG terms) enrichment analyses were performed with the functional annotation tool of the DAVID bioinformatics web server (<http://david.abcc.ncifcrf.gov/>), and the whole genome annotation was used as background. The genes of each enriched functional category, the enrichment fold change, and the various statistical parameters of the enrichment analysis including *P*-values, Bonferroni-corrected *P*-values, Benjamini-corrected *P*-values, and FDR values were determined.

## RESULTS

### Optimal codon usage has a conserved role in promoting the production of large proteins

To determine the potential role of codon usage in determining the levels of proteins of different sizes, we performed MS analysis of the WT *N. crassa* whole-cell extracts and determined the relative abundances of ~3000 proteins (Supplementary Table S1). As expected, a negative correlation (Pearson's correlation coefficient  $R = -0.37$ ) was observed between protein abundance and protein length proteome-wide (Figure 1A, left). However, when the analysis was only limited to genes with strong codon usage biases (codon bias index (CBI) > 0.5, genes with a strong preference for optimal codons) (63,78), the negative correlation between protein abundance and protein length was mostly abolished ( $R = -0.08$ ) (Figure 1A, center). When genes were limited to those with low CBI values (<0.2, genes with weak codon usage biases), the negative correlation between protein abundance and protein length became stronger ( $R = -0.47$ ) (Figure 1A, right). We also quantified gene codon usage bias using the tRNA adaptation index (tAI), a measure that takes into account tRNA concentrations and the efficiencies of codon-anticodon pairing (79). As gene tAI values increase (i.e. codon usage becomes more optimal), the negative correlation between protein abundance and protein length in a scanning window progressively weakened. The same observation was seen for the MS results obtained in this study using the iBAQ method or our previous result using the emPAI method (13) (Figure 1B).

To determine whether the codon usage effect observed above in *Neurospora* is conserved in other eukaryotes, we determined the correlations between protein abundance and protein length as a function of gene tAI values in *S. cerevisiae*, *D. melanogaster*, *C. elegans* and different mouse tissues using previously reported proteomic MS data (74–77). As in *N. crassa*, there are negative correlations between protein abundances and protein lengths in all these eukaryotic organisms, and the negative correlations progressively weaken as codon usage becomes more optimal (Figure 1B), regardless of the protein quantification methods used. These results suggest that optimal codon usage can counter the negative impact of CDS length on protein expression to allow large proteins to be efficiently expressed.

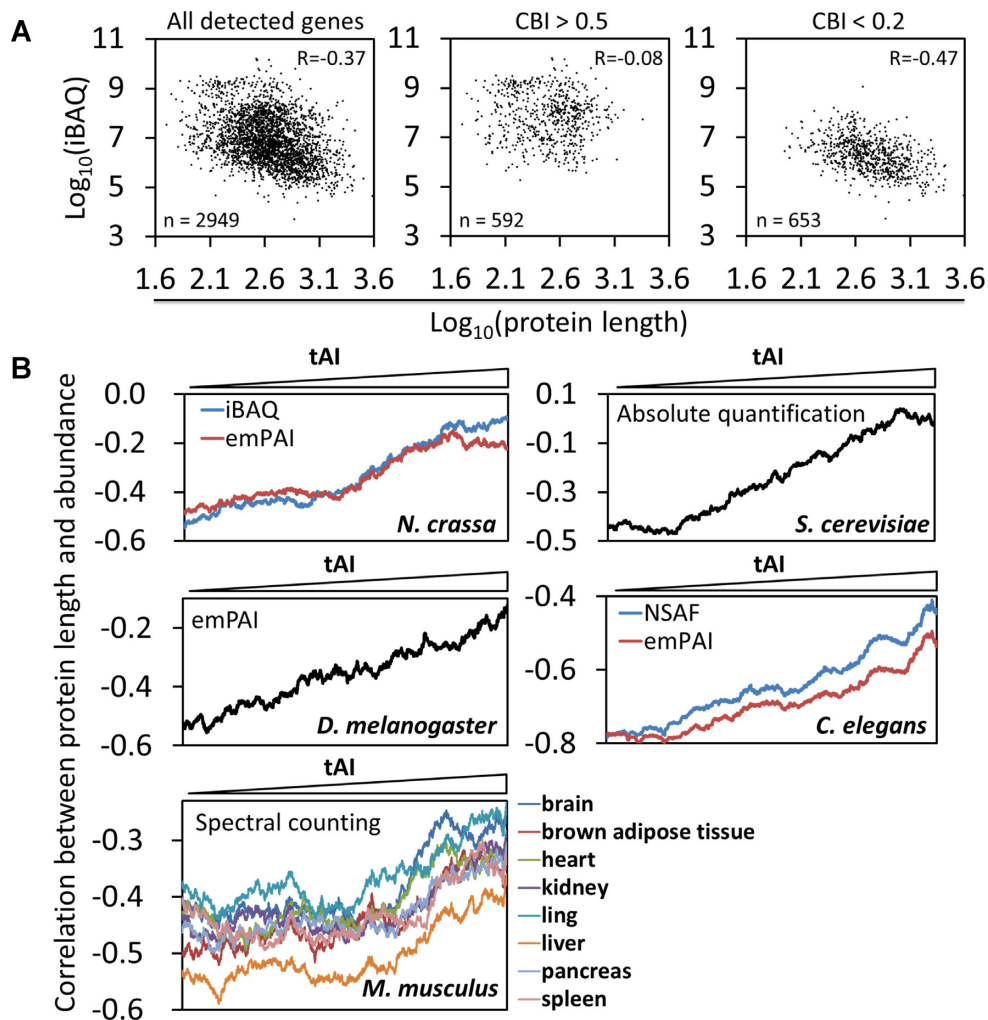
To determine whether the codon usage and length-dependent effect on protein levels is due to the regulation at mRNA level, we determined the correlations between mRNA levels and CDS lengths as a function of gene codon

usage using our RNA-seq results from the *Neurospora* WT strain (see below). As shown in Supplementary Figure S3, codon usage does not appear to affect the negative correlation between mRNA level and CDS, suggesting that the codon usage effect on the correlation between protein abundance and length is likely regulated at the translational level.

### The codon usage effects on mRNA translation and ribosome density are CDS length-dependent

To confirm the codon usage effect on protein abundance in a CDS length-dependent manner, we created four N-terminally Flag-tagged reporter constructs with different codon usage biases and CDS lengths for expression (Figure 2A). The CDS regions of the reporters correspond to the sequence of *NCU05784*, which encodes a small (125 aa), hypothetical protein, which we named elongation-dependent phosphorylated protein (EDP, see below). To determine the effect of CDS length on protein expression independent of codon usage, we created the 1 × EDP and 3 × EDP (3 tandem EDP repeats) constructs (Figure 2A). The EDP open reading frames are composed of either the WT codons or OPT codons. These expression constructs were targeted to the *csr-1* locus in the *N. crassa* genome. Homokaryotic transformants containing each reporter construct were obtained. For the WT codon usage constructs, 1 × EDP was produced in a considerably higher level than 3 × EDP (Figure 2B and C). Codon optimization resulted in higher protein levels for both 1 × EDP and 3 × EDP, but the codon optimization effect on protein up-regulation was much more robust for 3 × EDP than for 1 × EDP such that their abundances were comparable after codon optimization (Figure 2B, C and Supplementary Figure S4A). Thus, consistent with the bioinformatics analysis results, codon usage has differential effects on protein expression in a CDS length-dependent manner: Optimal codon usage preferentially allows large proteins to be efficiently expressed, and non-optimal codon usage has a more potent inhibitory effect on the expression of larger proteins than smaller ones.

To determine whether codon usage affects translation efficiency in a CDS length-dependent manner, we performed ribosome profiling using the WT strains expressing the different reporter proteins. Ribosome profiling is a powerful approach for studying mRNA translation dynamics *in vivo* as it provides codon-level resolution of ribosome locations and ribosome occupancy on mRNAs (21,80,81). Ribosome density on a given mRNA can be determined by the number of its RPFs normalized by its RNA level within the CDS region, and can reflect the ribosome flux of that mRNA (80). The relative ribosome density on the WT 3 × EDP mRNA was significantly lower than that on the WT 1 × EDP mRNA (Figure 2D and Supplementary Figure S4B). Note that the CDS regions of these two mRNAs have the identical codon usage profile. In contrast, the relative ribosome densities were comparable for the optimized (1 × OPT and 3 × OPT) mRNAs (Figure 2D and Supplementary Figure S4B). These results suggest that non-optimal codon usage preferentially inhibits translation of mRNAs with longer CDS regions. In addition, despite of the higher EDP protein levels for the OPT reporters, their ribosome densities were actually lower than that for the



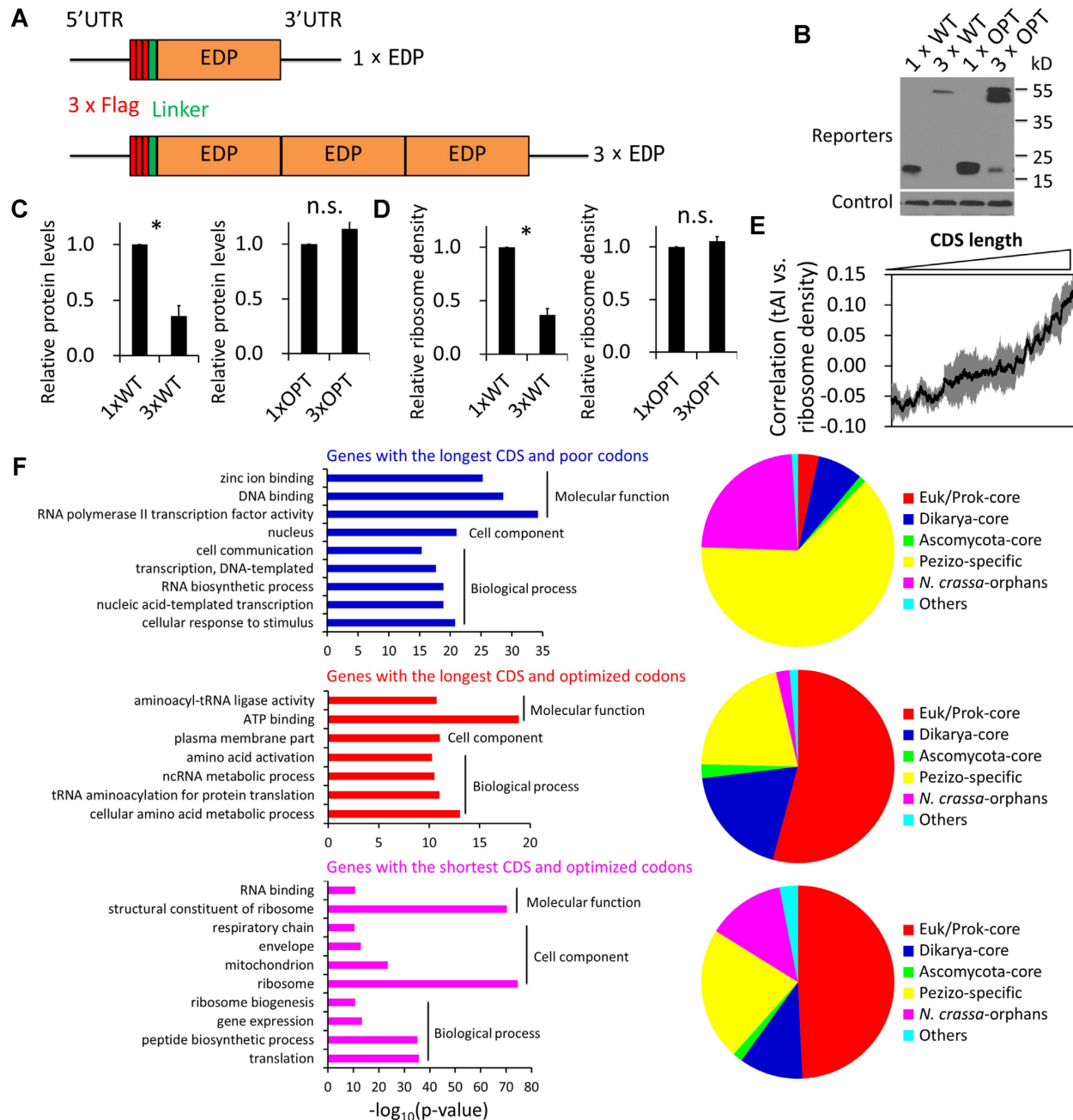
**Figure 1.** The negative correlation between protein length and abundance is codon usage-dependent and conserved in eukaryotes. (A) Scatter plots showing the correlations (Pearson's correlation coefficients) between protein length and abundance for *Neurospora* genes genome-wide (left panel) and for genes with high CBIs (middle panel) and low CBIs (right panel) in the WT *Neurospora* strain. The proteome-wide relative protein abundances (iBAQ values) were determined by MS. (B) Line plots showing the correlations between protein length and abundance as a function of codon usage bias measured by gene tAI values in different eukaryotic organisms. Proteome-wide protein abundance data are described and cited in Materials and Methods. All genes with detected protein levels were ranked by their tAIs, and the Pearson's correlation coefficients were calculated in continuous scanning windows from low to high tAIs. Each scanning window has 500 genes for *N. crassa*, *S. cerevisiae*, *D. melanogaster*, *C. elegans* and 1000 genes for different tissues of *M. musculus*. Methods used to quantify the relative protein levels are indicated.

1 × WT reporter (Supplementary Figure S4B). Because we previously showed that optimal codons can dramatically reduce ribosome densities on mRNAs in *Neurospora* due to increased elongation speed (21,23), this result suggests that the reduction of ribosome density caused by increased elongation speed due to codon optimization more than counters the increase of ribosome density caused by increased translation initiation efficiency.

To examine the CDS length-dependent effect on translation genome-wide, we determined the ribosome densities within CDS regions of all predicted *Neurospora* genes using the ribosome profiling results of the WT strain and calculated the correlations between tAIs and ribosome densities as a function of CDS lengths. It is important to note that codon usage has a major impact on elongation rates and optimal codons can dramatically reduce ribosome densities on mRNAs in *Neurospora* (21,23). Thus, ribosome density

measurement will overestimate the ribosome flux for mRNAs with poor codon usage and underestimate the flux for those with optimal codon usage. The correlations between gene tAIs and ribosome densities are weakly negative for short mRNAs (Figure 2E). As CDS length increases, however, the correlations gradually become positive, suggesting that optimal codon usage of long mRNAs positively correlates with ribosome density (Figure 2E). Because optimal codons result in fast elongation rates which lower ribosome density, the weak positive correlation actually indicates a strong positive effect of optimal codon usage on ribosome flux. Thus, codon usage affects ribosome flux on mRNAs in a CDS length-dependent manner: codon optimality preferentially enhances translation efficiency/ribosome flux of long mRNAs. These results also indicate the existence of a feedback mechanism from translation elongation to regulate translation initiation.





**Figure 2.** CDS length-dependent codon usage effect on protein expression and the separation of protein functions based on codon usage and CDS length. (A) Graphical depiction of the constructs used for the expression of the WT or OPT EDP reporters. The codon sequences for the N-terminal 3 × Flag tag, the linker (8 × Gly) at the N terminus, the 5'-UTR and 3'-UTR are identical in all constructs. (B) Western blot analysis showing that the expression of the indicated EDP reporters. A nonspecific constitutive band detected by the anti-Flag antibody was used as the control. (C) Densitometric analyses of the relative EDP protein levels from three independent experiments. The WT and OPT reporters are compared separately and the level of 1 × WT/OPT EDP protein was set as 1.0 for each panel. (D) Comparison of the relative ribosome density levels between the indicated reporter mRNAs. The relative ribosome density was measured by the relative RPFs normalized by the relative mRNA levels detected by qRT-PCR. The WT and OPT reporters are compared separately and the 1 × WT/OPT EDP ribosome density level was set as 1.0 for each panel. Data in panel C and D are means with standard deviations (SD) ( $n = 3$ ). \* indicates  $p < 0.05$  as determined by Student's two-tailed  $t$  test. n.s., not significant. (E) A line plot showing the correlations between codon usage (gene tAI) and ribosome density in 500-gene scanning windows as a function of CDS length (from short to long). (F) Left panels: Gene functional enrichment analyses showing the separation of protein functions based on codon usage and CDS length. Genes in the *N. crassa* genome were divided into different groups based on codon usage and CDS length (see Supplementary Figure S5). The most statistically enriched GO categories for genes in the indicated groups were shown. See Supplementary Table S2 for the complete list. Right panels: Pie charts showing the proportions of genes (corresponding to the left panels) classified into mutually exclusive lineage specificity groups based on their conservation in other organisms.

### Codon usage and CDS length-dependent separation of gene functions in the genome

Because codon usage differentially affects protein expression level in a CDS length-dependent manner, we hypothesize that large proteins with critical functions would have optimal codon usage profiles to allow their efficient synthesis. To examine this, we grouped all predicted *N. crassa* genes based on their CBI values and CDS lengths and performed gene functional enrichment analyses for four different groups of genes (each with 1000 genes) (Supplementary Figure S5): (i) those with the longest CDS regions among those with strong codon usage biases ( $CBI \geq 0.3$ ), (ii) those with the longest CDS regions among those with non-optimal codon usage biases ( $CBI \leq 0.15$ ), (iii) those with the shortest CDS regions among those with strong codon usage biases and (iv) those with the shortest CDS regions among those with non-optimal codon usage biases. There was no significant functional enrichment ( $P$ -value  $< 1e-10$ ) for genes in group (iv). In contrast, many genes with similar functions or in the same biological process were significantly enriched in the other three groups (Figure 2F, left panels and Supplementary Table S2). As predicted, the genes with long CDS regions and optimal codon usage are enriched for functional categories associated with essential cellular processes such as amino acid activation and amino acid metabolic process, tRNA aminoacylation, plasma membrane components, and non-coding RNA metabolic process (Figure 2F left panels and Supplementary Table S2). In contrast, the genes with the long CDS regions and poor codon usage profiles are mostly enriched for functional categories involved in the responses to environmental stimulus, cell communication, and transcriptional regulation (Figure 2F left panels and Supplementary Table S2). The genes with the shortest CDS regions and optimal codons are significantly enriched for functional categories related to translation, ribosomal proteins and ribosome biogenesis, mitochondrial components and respiratory chain (Figure 2F, left panels and Supplementary Table S2).

We also classified the genes of the three groups in Figure 2F, left panels into six mutually exclusive lineage specificity groups based their conservation in other organisms (82): (i) eukaryote/prokaryote-core (genes with homologs in non-fungal eukaryotes and/or prokaryotes), (ii) dikarya-core (genes with homologs in *Basidiomycota* and *Ascomycota* species), (iii) *Ascomycota*-core, (iv) *Pezizomycotina*-specific, (v) *N. crassa*-orphan genes and (vi) others (genes with homologs identified in prokaryotes or non-fungal eukaryotes in addition to *Pezizomycotina*, but not in members of the *Basidiomycota*, *Saccharomycotina* or *Taphrinomycotina*). As shown in the right panels of Figure 2F, the genes with long CDS regions and optimal codon usage are mostly genes in class (i) and (ii), indicating that they are conserved beyond fungi and likely have functions critical for cell survival. In contrast, the genes with long CDS regions and poor codon usage profiles are mostly *N. crassa*-specific and *Pezizomycotina*-specific. Systematic deletion studies of *Neurospora* genes previously revealed that the essential genes are mostly genes that are conserved beyond fungi, while the *Neurospora*-specific genes are not critical for cell survival

(65). Together, these results are consistent with our hypothesis that optimal codon usage is a mechanism that allows large proteins required for critical cellular functions to be efficiently produced.

### Ribosome stalling at a stage between pre-accommodation and pre-translocation induces eIF2 $\alpha$ phosphorylation

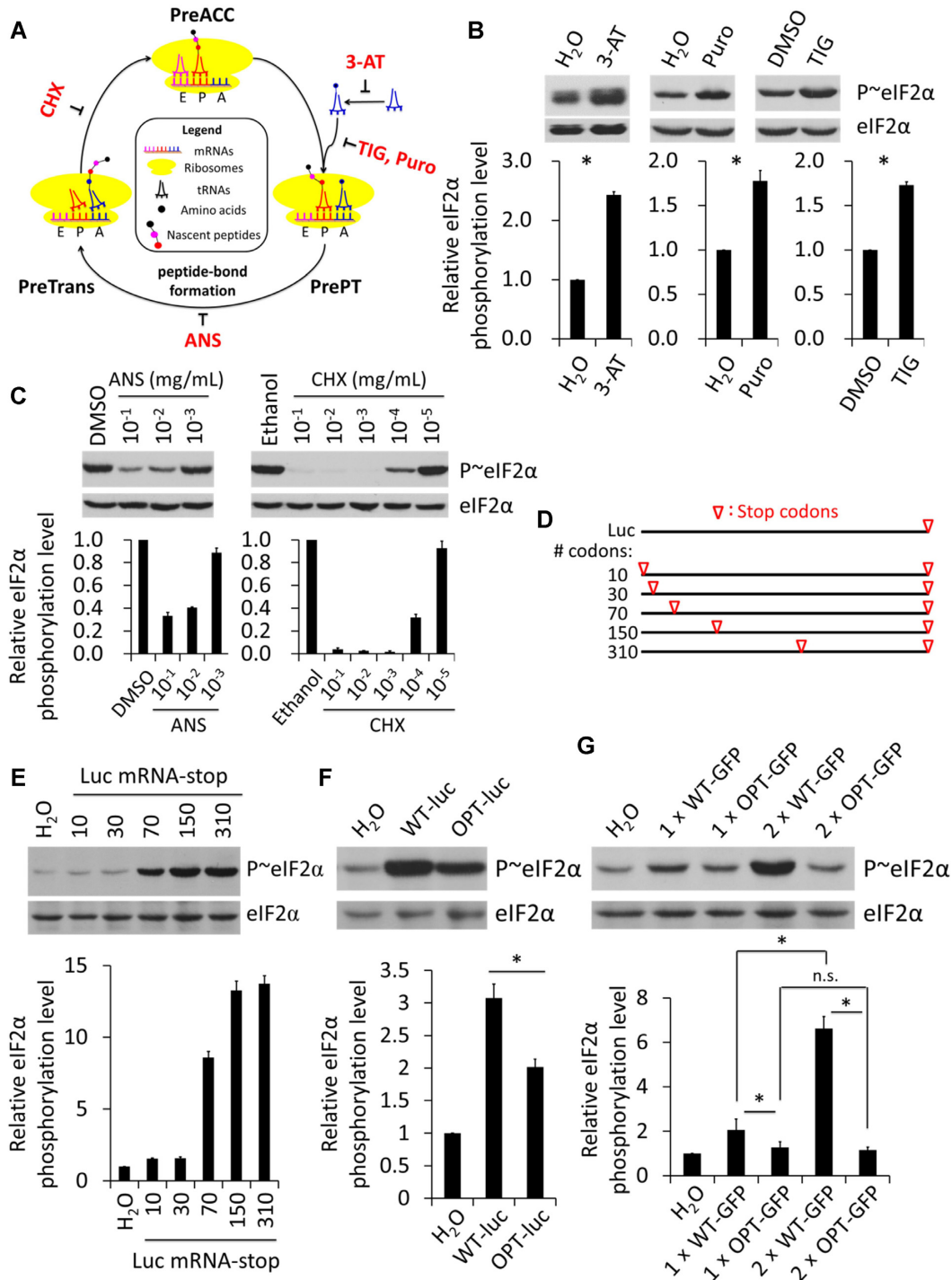
The effect of codon usage on ribosome flux suggests that codon usage-dependent elongation can feedback to regulate translation initiation under nutrient replete growth conditions. Phosphorylation of eIF2 $\alpha$  is an important regulatory mechanism of translation initiation, and is known to be induced by many types of stress conditions to result in global inhibition of translation initiation of many mRNAs (42–44,50). Since codon usage has been shown to play an important role in determining elongation speed and rare codons cause ribosome pausing with empty A site in *Neurospora* (21,23), we first examined whether ribosome stalling can trigger eIF2 $\alpha$  phosphorylation by using different pharmacological inhibitors that can block translation elongation at different steps of the eukaryotic translation elongation cycle (Figure 3A).

We first treated *Neurospora* cultures with 3-AT, a competitive inhibitor of the product of *his-3* gene, which results in accumulation of uncharged tRNAs and cellular amino acid starvation. This treatment resulted in a significant elevation of eIF2 $\alpha$  phosphorylation (Figure 3B). Puromycin (Puro, tyrosyl-tRNA-like) and tigecycline (TIG, tetracycline-like) can cause ribosome stalling at the stage between pre-accommodation and pre-translocation in the translation elongation cycle (83–85) (Figure 3A). Treatment of *Neurospora* cultures with either agent also enhanced eIF2 $\alpha$  phosphorylation *in vivo* (Figure 3B). In contrast, treatments with anisomycin (ANS) and cycloheximide (CHX), which inhibit peptide bond formation and translocation, respectively, resulted in dramatic dose-dependent decreases of eIF2 $\alpha$  phosphorylation (Figure 3C). These results suggest that ribosome stalling at the stage between pre-accommodation and pre-translocation but not at other stages of the elongation cycle induces eIF2 $\alpha$  phosphorylation. This phenomenon may be caused by distinct ribosome conformations when ribosomes stall at different functional states. Consistent with this notion, alteration of ribosome conformation at distinct elongation states caused by CHX, TIG and ANS treatments was previously demonstrated experimentally (84). During the preparation of this manuscript, a similar conclusion on the effects of some of these inhibitors was also reached in yeast (86).

### Induction of eIF2 $\alpha$ phosphorylation by mRNA translation is dependent on CDS length and codon usage

Although eIF2 $\alpha$  phosphorylation can be induced by treatment of cultures with translation inhibitors, it is not clear whether it can be regulated by codon usage or CDS length under normal growth (nutrient replete) conditions. To examine these possibilities, we took advantage of the *Neurospora* cell free *in vitro* translation system that was previously shown to accurately reflect protein translation *in vivo*





**Figure 3.** eIF2α phosphorylation can be induced by either translation inhibitors or poor codon usage in a CDS length-dependent manner. (A) Schematic representation of the eukaryotic translation elongation cycle and the effects of the indicated chemical inhibitors of translation. PreAcc, pre-accommodation; PrePT, pre-peptide bond formation; PreTrans, pre-translocation. (B, C) Western blot analyses and their quantifications showing the relative levels of eIF2α phosphorylation and eIF2α after treating the WT *Neurospora* cultures with the indicated inhibitors for 20 min. For experiment in each panel, equal amount of solvent used for dissolving the corresponding chemical was used as control. (D) Schematic diagram of the luciferase mRNA-stop constructs used for *in vitro* transcription to synthesize luciferase mRNAs with an in-frame stop codon at different indicated codon positions after the start codon within its CDS. (E–G) Western blot analyses and their quantifications showing the relative levels of eIF2α phosphorylation and eIF2α in the *Neurospora in vitro* translation products. The indicated mRNA templates were translated in the *Neurospora in vitro* translation extracts for 15 min before the reactions were stopped. Data are means with SD ( $n = 3$ ). \* $P < 0.05$ , as determined by Student's two-tailed  $t$  test.

(21,87,88). Cellular mRNAs were depleted from this system by micrococcal nuclease digestion so that the translation of a single species of mRNAs and its impact on eIF2 $\alpha$  phosphorylation can be examined. We synthesized a series of capped and polyadenylated WT *luciferase* (*luc*) mRNAs with an in-frame stop codon at different positions ranging from the 10th codon to 310th codon from the start codon (Figure 3D). Quantification of eIF2 $\alpha$  phosphorylation after translation of these mRNAs revealed that there was a CDS length-dependent effect on eIF2 $\alpha$  phosphorylation: mRNAs with long CDS regions result in higher level of eIF2 $\alpha$  phosphorylation than those with short CDS regions (Figure 3E).

To determine whether eIF2 $\alpha$  phosphorylation is dependent on codon usage, we evaluated the eIF2 $\alpha$  phosphorylation level in the presence of the WT *luc* mRNA or the OPT *luc* mRNA. As expected, the translation of the WT mRNA resulted in a significantly higher level of eIF2 $\alpha$  phosphorylation than the translation of the OPT version (Figure 3F and Supplementary Figure S6A). To confirm that this result is not gene specific, we synthesized and translated the WT or OPT versions of mRNAs encoding one or two copies of the *GFP* CDS regions (1  $\times$  GFP and 2  $\times$  GFP, respectively) (Supplementary Figure S6B). Both of the WT *GFP* mRNAs induced significantly higher levels of eIF2 $\alpha$  phosphorylation than the OPT mRNAs (Figure 3G). Strikingly, the induction of eIF2 $\alpha$  phosphorylation by the WT 2  $\times$  GFP was much higher than that by the WT 1  $\times$  GFP, while the OPT 2  $\times$  GFP mRNAs had little effect on eIF2 $\alpha$  phosphorylation compared to the OPT 1  $\times$  GFP. Together, these results demonstrate that translation elongation can induce eIF2 $\alpha$  phosphorylation in a CDS length-dependent and codon usage-dependent manner in the absence of translation stress. Thus, codon usage and CDS length can potentially regulate translation initiation by affecting eIF2 $\alpha$  phosphorylation.

#### Loss of eIF2 $\alpha$ kinase CPC-3 preferentially up-regulates protein expression for mRNAs with long CDS and poor codon usage

eIF2 $\alpha$  phosphorylation does not have to cause global down-regulation of translation initiation (44,50). Our findings that eIF2 $\alpha$  phosphorylation depends on codon usage and CDS length in the absence of translation stress suggest that this is an elongation-dependent feedback mechanism that may alter translation efficiency of specific mRNAs. We hypothesize that long mRNAs and those enriched with rare codons cause local accumulation of phosphorylated eIF2 $\alpha$ , resulting in specific rather than a general suppression of mRNA translation. *cpc-3* (*cross pathway control-3*, NCU01187) encodes the *Neurospora* homolog of the yeast and mammalian GCN2, and is the only known kinase responsible for eIF2 $\alpha$  phosphorylation in *Neurospora* (52). As expected, eIF2 $\alpha$  phosphorylation is completely abolished in the  $\Delta cpc-3$  strain (Supplementary Figure S1). We compared the expression of the four EDP reporters described above in the WT and  $\Delta cpc-3$  strains. Although the deletion of *cpc-3* did not affect the expression levels of the WT or OPT 1  $\times$  EDP or OPT 3  $\times$  EDP, it significantly increased the protein level of the WT 3  $\times$  EDP (Figure 4A, B and Supple-

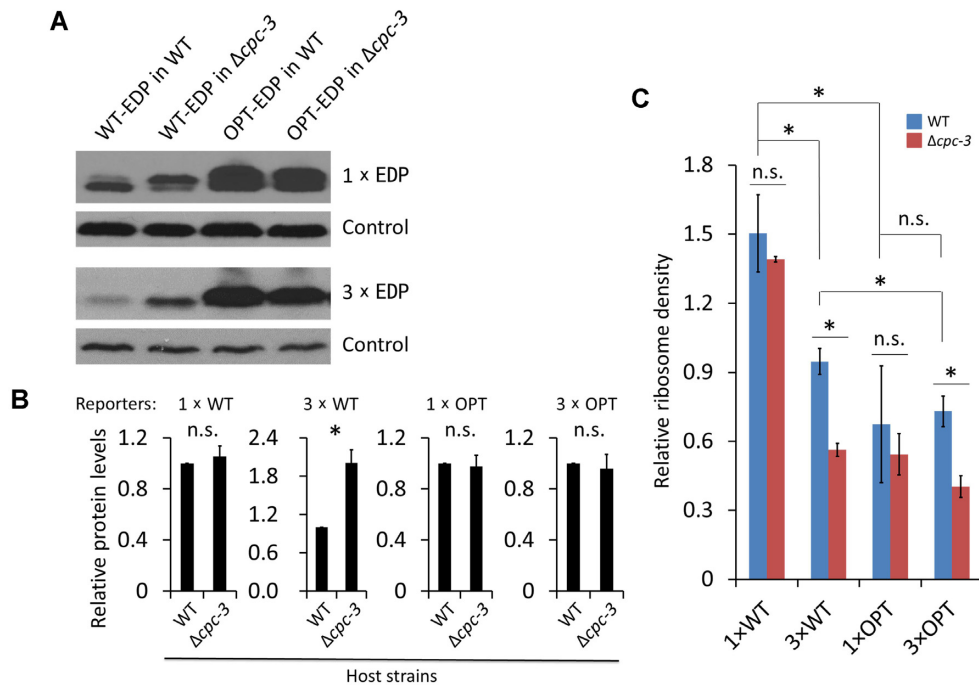
mentary Figure S7). These results suggest that eIF2 $\alpha$  phosphorylation preferentially inhibits the translation of long CDS mRNAs with poor codon usage, resulting in their specific translation inhibition rather than a general translation inhibition. Although deletion of *cpc-3* increased the protein level of WT 3  $\times$  EDP, its level was still lower than that of the WT 1  $\times$  EDP level (Supplementary Figure S7), suggesting that the effect of CDS length on protein abundance is determined by both CPC-3-dependent and CPC-3-independent mechanisms. A CPC-3-independent mechanism may be involved in the negative influence of CDS length on translation due to less efficient ribosome recycling for mRNAs with long CDS than for short CDS mRNAs (39,41). It should also be noted that the negative effect of 'CDS length' on protein level depends on codon usage, because unlike the WT reporters, the protein levels of the 1  $\times$  OPT and 3  $\times$  OPT EDP are comparable (Figure 4A and Supplementary Figure S7), which is consistent with the proteomic data analysis results (Figure 1). This result indicates that the negative effect of CDS length on protein production is attenuated for mRNAs with optimal codon usage.

#### Deletion of *cpc-3* resulted in a decrease of ribosome density of the reporter mRNAs with long CDS

To understand how CPC-3 influences translation *in vivo*, we also performed ribosome profiling experiment in the  $\Delta cpc-3$  strains expressing the different EDP reporters under nutrient replete condition and compared the relative ribosome density of the EDP reporter mRNAs in the WT and  $\Delta cpc-3$  strains by normalizing the number of RPFs on CDS regions with mRNA levels. Consistent with the result in Figure 2D, the relative ribosome density of 3  $\times$  WT EDP was significantly decreased compared to that of the 1  $\times$  WT EDP in the WT strain, while that of 3  $\times$  OPT EDP is comparable to that of 1  $\times$  OPT EDP (Figure 4C). In addition, the ribosome densities of 1  $\times$  and 3  $\times$  WT EDP mRNAs were both significantly higher than that of their OPT counterparts in the WT strain (Figure 4C), indicating that codon optimization results in faster translation elongation speed, which reduced ribosome density despite the strong up-regulation of OPT mRNA translation (21). Surprisingly, compared to the WT strain, the relative ribosome density of the 3  $\times$  WT EDP but not 1  $\times$  WT EDP was significantly decreased in the  $\Delta cpc-3$  strain (Figure 4C). The up-regulation of 3  $\times$  WT EDP protein level but a decrease of ribosome density on its mRNA in the  $\Delta cpc-3$  strain suggests that, in addition to its role on translation initiation, CPC-3 may also have a role that preferentially inhibits translation elongation on mRNAs with long CDS and poor codon usage. As a result, the increase of elongation speed on 3  $\times$  WT EDP mRNA in the  $\Delta cpc-3$  strain more than counters the effect of the increase of translation initiation, resulting in a decrease of ribosome density.

#### CPC-3 slows down translation elongation rate in a codon usage-dependent manner

To confirm our conclusion above and determine the effect of CPC-3 genome-wide, we calculated the gene-specific ribosome densities for all predicted *Neurospora* genes using the ribosome profiling and accompanying RNA-seq results of the WT and  $\Delta cpc-3$  strains. We found that among



**Figure 4.** Codon usage and CDS length-dependent effect on protein expression level in the  $\Delta cpc-3$  strain. (A) Western blot analysis results showing the expression levels of the Flag-tagged 1× EDP (WT or OPT) and 3× EDP (WT or OPT) reporters in the WT or  $\Delta cpc-3$  strains. A nonspecific constitutive band detected by the anti-Flag antibody was used as the control. (B) Densitometric analyses of the western blot results of the EDP levels from three independent experiments described above. The EDP protein level in the WT strain for each panel was set as 1.0. (C) Comparison of the relative ribosome density levels between the indicated reporter mRNAs in the WT strain and the  $\Delta cpc-3$  strain. The relative ribosome density was measured by the relative RPFs normalized by the relative mRNA levels detected by qRT-PCR. Data in panel B-C are means with SD ( $n = 3$ ). \* indicates  $P < 0.05$  as determined by Student's two-tailed  $t$  test. n.s., not significant.

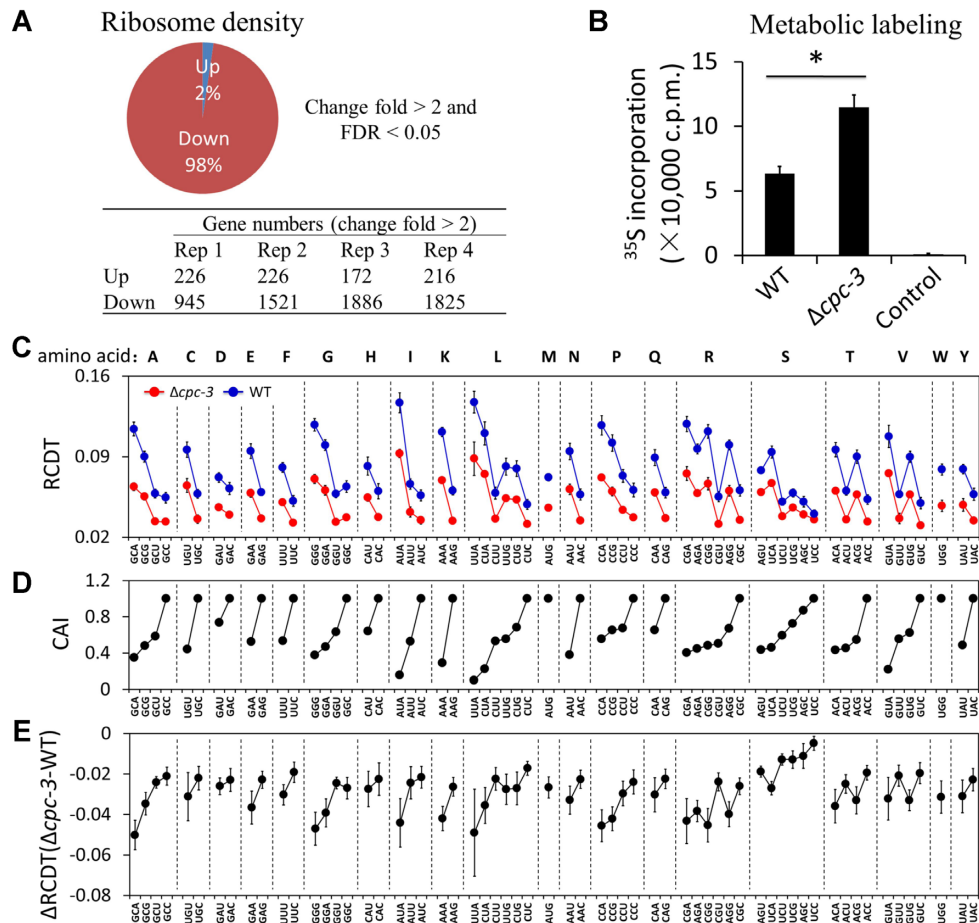
genes with more than 2-fold changes (FDR < 0.05) in ribosome density in the  $\Delta cpc-3$  strain compared to the WT strain, 98% genes had a decreased ribosome density whereas only 2% had an increased ribosome density (Figure 5A). To further confirm this result, we also performed polysome profiling experiments, which showed that the ratio of polysomes/monosome in the  $\Delta cpc-3$  strain was lower than that in the WT strain (Supplementary Figure S8A and B), consistent with the ribosome profiling result.

The global decrease of ribosome densities on mRNAs in the  $\Delta cpc-3$  strain is unexpected due to the well-established role of CPC-3 and eIF2 $\alpha$  phosphorylation in inhibiting translation initiation. To confirm the inhibitory role of CPC-3 and eIF2 $\alpha$  phosphorylation in translation, we performed  $^{35}\text{S}$ -methionine pulse labeling experiment to compare the overall translation efficiencies between the WT and  $\Delta cpc-3$  strains grown at 30°C. As expected, the  $\Delta cpc-3$  strain had a significant increase of  $^{35}\text{S}$ -methionine incorporation level than the WT strain (Figure 5B), indicating that CPC-3 and eIF2 $\alpha$  phosphorylation indeed inhibit general translation efficiency in *Neurospora*. Just like what we found for the 3× WT EDP reporter mRNA, the increase of general translation efficiency but reduction of ribosome densities on most of mRNAs in the  $\Delta cpc-3$  strain suggests an overall increased translation elongation rate in the  $\Delta cpc-3$  strain that can more than counter the effect of the increase of translation initiation on ribosome density.

To determine the role of CPC-3 in translation elongation rate, we calculated the RCDTs for all 61 amino acid-encoding codons in the WT and  $\Delta cpc-3$  strains using the ribosome profiling results. Consistent with our previous studies (21,23,58), there was a clear codon usage bias in RCDTs for all codon families in the WT strain. The most preferred synonymous codon was always the one with the lowest RCDT in each codon family, while rare codons had the highest RCDTs (Figure 5C and D). Although the codon usage biases in RCDTs did not change in the  $\Delta cpc-3$  strain, RCDTs were reduced for all codons (Figure 5C and D). This indicates that there was a global increase in translation elongation rate in the  $\Delta cpc-3$  strain, resulting in the decreased ribosome density on most mRNAs despite the increase of translation initiation. We next examined whether the effect of CPC-3 on translation elongation is codon usage-dependent. We observed that the decrease of RCDTs in the  $\Delta cpc-3$  strain compared to the WT strain was always greater for the rare codons than for the most preferred codons in all codon families (Figure 5E), indicating that CDC-3 preferentially slows translation elongation at rare codons.

To further confirm this conclusion, we utilized the *Neurospora* cell-free translation system and the WT and OPT luciferase reporters, which was previously used to demonstrate the codon usage effect on elongation speed (21). Because luciferase is known to be folded co-translationally and becomes functional within a few seconds after the completion of translation, the time of first appearance (TFA) val-

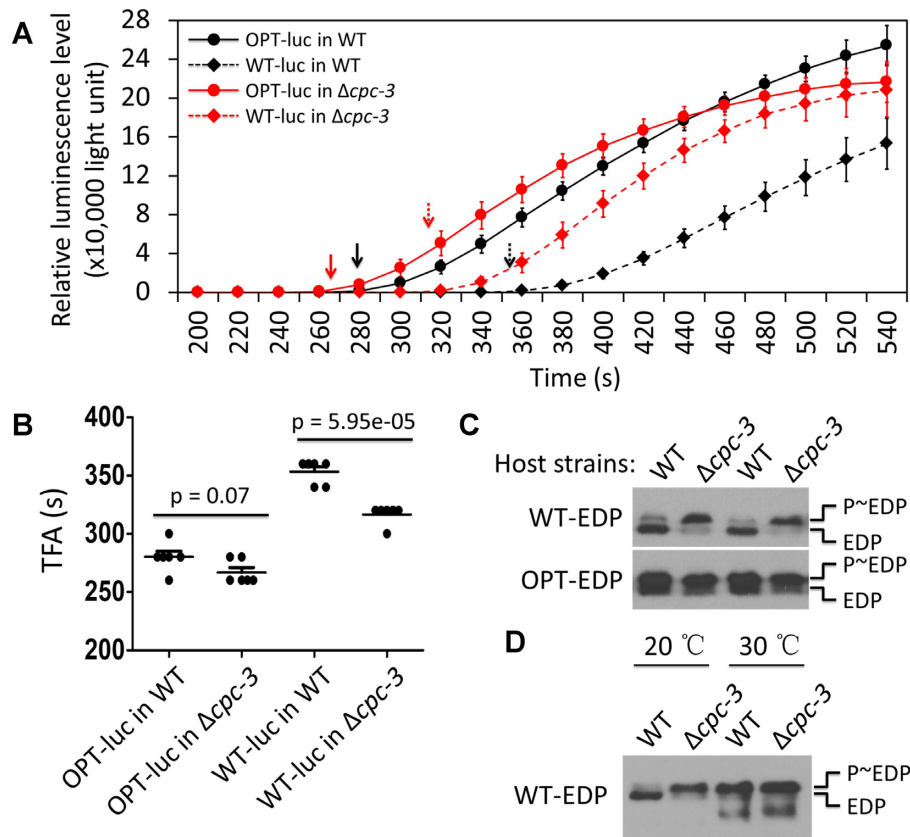




**Figure 5.** Ribosome profiling results show that CPC-3 slows down codon decoding rates in a codon usage-dependent manner. (A) Upper panel: A pie chart showing the percentages of mRNAs with reduced (red) or increased (blue) ribosome densities on mRNAs in the  $\Delta cpc-3$  strain comparing to the WT strain. Genes with significantly increased and decreased ribosome density were identified by using the following cut-off: the change fold of the average of the four replicates >2 and FDR <0.05. The *P*-values for the four replicates were determined by Student's two-tailed *t* test, which were further adjusted by FDR. Lower panel: A table showing the numbers of genes with change fold of ribosome density >2 (up- or down-regulated) in the four replicates. (B) Metabolic labeling experiment indicated an increased global protein synthesis rate in the  $\Delta cpc-3$  strain.  $^{35}\text{S}$ -Methionine was added into the cultures for 45 min before sample collection. For the control culture, CHX (a final concentration of 0.1 mg/ml) was added 2 min before adding  $^{35}\text{S}$ -methionine. Data are means with SD ( $n = 3$ ). \*  $P < 0.05$ ; as determined by Student's two-tailed *t* test. (C) Comparison of RCDTs in the WT and  $\Delta cpc-3$  strains ( $n = 3$ ). (D) The CAI values of the corresponding codons in each codon family. (E) Quantification of the differences of RCDTs between the  $\Delta cpc-3$  strain and the WT strain showed that CPC-3 preferentially impacts the elongation rates of rare codons in each codon family. Data are means with SD ( $n = 3$ ).

ues of luciferase signal for the WT and OPT *luc* mRNAs reflect differences in translation elongation rates (21,58,89). In addition, translation initiation time was previously estimated to be less than several seconds (37,47,90). Thus, the TFA changes should reflect changes in elongation rates. Similar to what we reported previously (21), the TFA of the OPT *luc* mRNA was significantly shorter than that of the WT *luc* in the WT extracts (Figure 6A and B). In the  $\Delta cpc-3$  extracts, the TFA values of both WT and OPT *luc* mRNAs were reduced, confirming the effect of CPC-3 on translation elongation. Importantly, the impact of loss of CPC-3 on translation elongation rate was clearly codon usage-dependent: The TFA was significantly faster (by  $\sim 40$  s) for the WT *luc* mRNA in the  $\Delta cpc-3$  extracts than in the WT extracts, but for the OPT *luc* mRNA, the TFA was only marginally reduced in the  $\Delta cpc-3$  extracts and was not statistically different from that in the WT extracts (Figure 6A and B).

When the WT  $1 \times$  EDP reporter was expressed in the WT strain, we noticed that there were several protein species with different mobilities in SDS-PAGE gel (Figure 4A). Phosphatase treatment of the protein extracts indicated that the bands that migrated more slowly were phosphorylated EDP (Supplementary Figure S9). EDP was mostly in the hypo-phosphorylated form in the WT strain but was mostly hyper-phosphorylated in the  $\Delta cpc-3$  strain (Figure 4A and 6C). The expression of the OPT  $1 \times$  EDP, however, resulted in hyper-phosphorylation of EDP in both of the WT and  $\Delta cpc-3$  strains (Figure 6C). This codon usage-dependent protein phosphorylation profile change is very similar to what we previously observed when the codon usage profiles of genes encoding circadian clock proteins FRQ and PER from *Neurospora* and *Drosophila*, respectively, were changed (16,30,63). These results suggest that the EDP phosphorylation profile is affected by the co-translational protein folding process that is sensi-

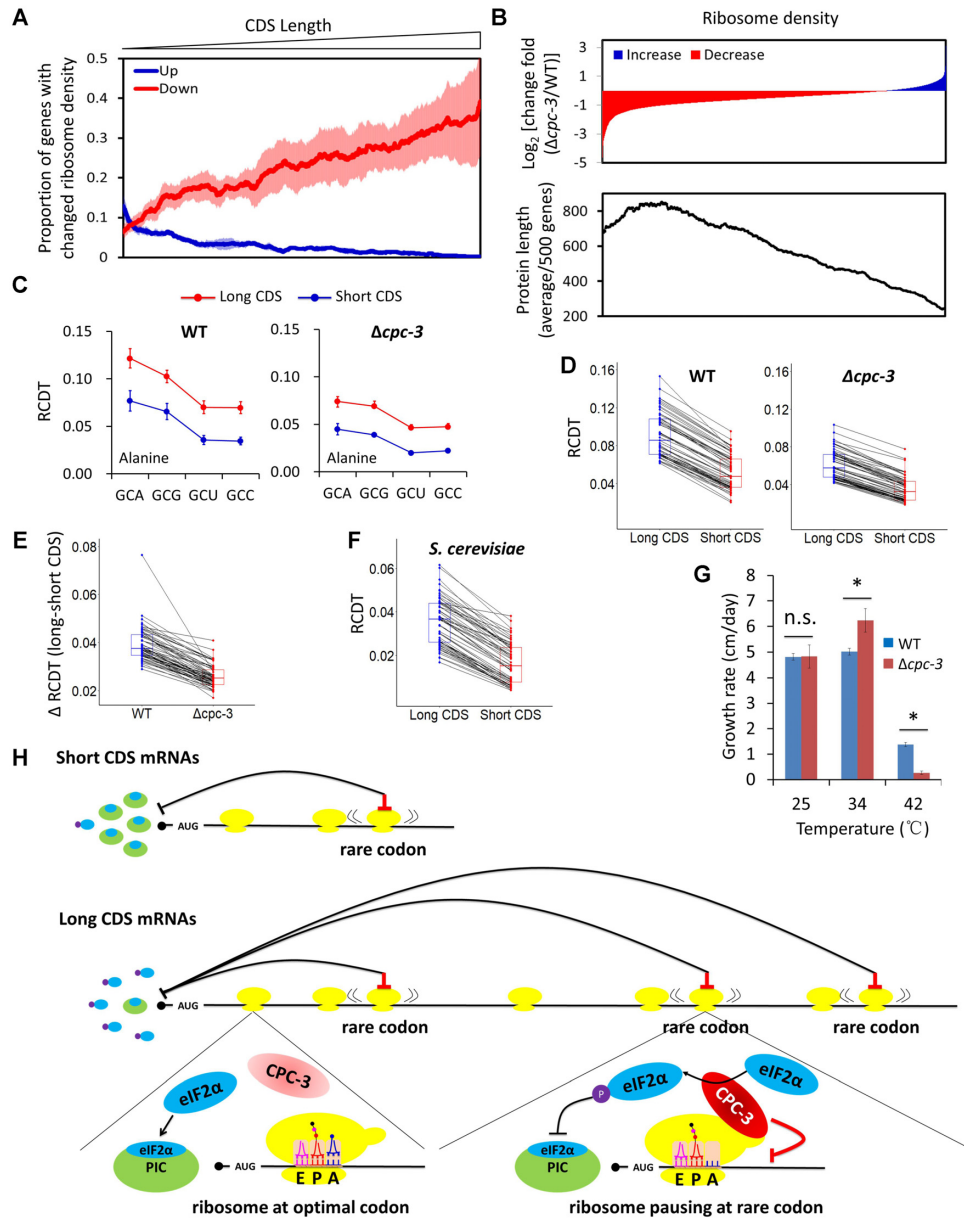


**Figure 6.** Codon usage-dependent increase of translation elongation rate in the  $\Delta cpc-3$  strain demonstrated by *in vitro* translation assay and an *in vivo* protein phosphorylation reporter. (A) *In vitro* translation experiments of the indicated strains using the *Neurospora* cell-free translation system showed the comparison of the temporal profiles of luciferase signal level of different luciferase mRNAs at 26°C. The arrows indicate the TFAs of the luciferase signal for indicated mRNAs. Data are represented as mean  $\pm$  SD. (B) Scatter plots comparing the TFAs between the indicated *Luc* mRNAs in the WT or  $\Delta cpc-3$  lysates (n = 6). Means indicated by horizontal bars. P-values was shown, as determined by Student's two-tailed *t* test. (C) Western blot analysis results showing that the EDP phosphorylation profiles of the indicated reporters were determined by codon usage and host strains. The phosphorylated and un-phosphorylated EDP bands were indicated. (D) Western blot analysis results showing that high temperature promotes the EDP phosphorylation of the WT-EDP reporter in the WT strain.

tive to translation elongation rate regulated by codon usage. Rapid translation elongation caused by either codon optimization or deletion of *cpc-3* results in altered EDP structure that promotes its phosphorylation. To further confirm this, we compared the WT 1  $\times$  EDP expression profiles for cultures grown at 20 and 30°C. We have shown previously that higher temperature increases translation elongation rate (21). As expected, in the WT strain, the protein expressed from WT 1  $\times$  EDP was hypo-phosphorylated at 20°C but became hyper-phosphorylated at 30°C (Figure 6D). On the other hand, EDP expressed from WT 1  $\times$  EDP mRNA was hyper-phosphorylated at both temperatures in the *cpc-3* strain. Together, these results show that CPC-3 not only regulates translation initiation via eIF2 $\alpha$  phosphorylation but also translation elongation in a codon usage-dependent manner. Thus, CPC-3 plays an important role in determining the codon usage effect on elongation speed so that optimal codons are decoded much faster than rare codons. Our results here also caution the use of ribosome density as a reflection of translation efficiency because elongation rate can have a major impact on ribosome density.

### CPC-3 influences translation kinetics in a CDS length-dependent manner

Because the CPC-3-mediated eIF2 $\alpha$  phosphorylation is dependent on both codon usage and CDS length, we examined whether the role of CPC-3 on translation kinetics is also influenced by CDS length. We calculated the proportions of genes with up-regulated (change fold of  $\Delta cpc-3$ /WT > 2) or down-regulated (change fold of  $\Delta cpc-3$ /WT < 0.5) ribosome density in a 500-gene scanning window. After ranking genes by their CDS lengths from short to long, we found that as CDS length gradually increased, the proportions of genes with down-regulated ribosome density increased markedly (Figure 7A). In contrast, the proportions of genes with up-regulated ribosome density decreased as CDS length increased in the same corresponding window (Figure 7A). When we ranked genes by their log<sub>2</sub> [change fold ( $\Delta cpc-3$ /WT)] of ribosome density from low to high and determined their averages of CDS lengths in 500-gene scanning windows, it was clear that the genes with down-regulated ribosome density are mostly those with long CDS mRNAs (>700 aa), whereas the genes with up-regulated or unchanged ribosome density tend to be short CDS



**Figure 7.** CPC-3 regulates CDS length-dependent codon decoding rates. (A) Line plots showing that the proportions of genes with decreased ribosome density in the  $\Delta cpc-3$  strain (change fold of  $\Delta cpc-3/WT < 0.5$ ) increased as protein length increased. Genes were ranked by their CDS lengths from short to long and the proportions of genes with decreased/increased ribosome density in a scanning 500 gene-window were calculated and plotted. Data are means with SD ( $n = 4$ ). Shaded areas indicate error bars in each scanning window. (B) Genes with decreased ribosome density preferentially code large proteins while genes with increased ribosome density prefer to code shorter proteins. Top panel: genes were ranked by ribosome density changes [ $\log_2(\text{change fold of } \Delta cpc-3/WT)$ ] from the greatest decrease to the greatest increase. Bottom panel: the averages of protein lengths in 500-gene scanning windows corresponding to the ribosome density change above were calculated and plotted. (C) The RCDTs of the alanine codons within mRNAs with CDS  $>600$  aa and those  $<300$  aa in the WT and  $\Delta cpc-3$  strains. Results for all other codons are shown in Supplementary Figure S11. (D) Pairwise comparison of RCDTs of all codons in the long and short CDS mRNAs in the WT and  $\Delta cpc-3$  strains. The same codons were linked by black lines. (E) The differences of RCDTs of all codons (pairwise comparison) between long and short CDS mRNAs [ $\Delta RCDT$  (long-short CDS)] are always larger in the WT strain than that in the  $\Delta cpc-3$  strain. (F) Pairwise comparison of RCDTs of all codons in long and short CDS mRNAs in *S. cerevisiae*. The previously published ribosome profiling and RNA-seq data were used for this analysis (91). (G) Comparison of daily growth rates of the WT and  $\Delta cpc-3$  strains by race tube assay at 25, 34 and 42°C. The asterisk indicates  $P < 0.05$  ( $n = 6$ ) as determined by Student's two-tailed *t*-test. (H) A schematic illustration of a model explaining the mechanism of translation elongation feeding back to regulate translation initiation and elongation speed in a codon usage and CDS length-dependent manner. Bottom left: ribosome does not pause at optimal codons during elongation. Bottom right: Rare codons cause ribosome pausing during elongation, which potentially promotes the interaction between CPC-3 and ribosomes, resulting in CPC-3 activation and phosphorylation of eIF2 $\alpha$ -GDP. The phosphorylated eIF2 $\alpha$ -GDP prevents the recycling of eIF2 $\alpha$ -GDP to become eIF2 $\alpha$ -GTP, thus inhibiting the formation of pre-initiation complex (PIC) and therefore translation initiation. Translation of an mRNA with a short CDS triggers less rare codon-mediated ribosome pausing, thus less CPC-3 activation and less the eIF2 $\alpha$  phosphorylation-mediated inhibition of initiation. Translation of an mRNA with a long CDS potentially triggers more rare-codon mediated ribosome pausing events, resulting in high local concentration of phosphorylated eIF2 $\alpha$ -GDP, which inhibits ribosome recycling and translation re-initiation. In addition, CPC-3 also inhibits translation elongation rate in a codon usage-dependent manner so that codon decoding rates for optimal codons are faster than those for non-optimal codons.



mRNAs (Figure 7B). These results are also consistent with the results of the four EDP reporters (Figure 4C), which showed that the relative ribosome density was significantly decreased for the 3× EDP mRNAs but not for the 1× EDP mRNAs in the  $\Delta cpc-3$  strain. Together, these results suggest that the inhibitory effect of CPC-3 on translation kinetics is dependent on both codon usage and CDS length.

We then performed proteomic quantitative MS analyses to identify differentially expressed proteins between the  $\Delta cpc-3$  strain and its complementation strain by TMT mass tagging technology (Supplementary Table S3). The result showed that, after ranking proteins by their lengths from short to long, the proportions of up-regulated proteins (FDR < 0.05) increase in a 1000-gene scanning window as protein size increases (Supplementary Figure S10A). In addition, comparison of the protein length profiles showed that the up-regulated proteins are preferentially larger proteins (with an average of 645 aa) than the predicted proteome (with an average of 450 aa) (Supplementary Figure S10B). These results suggest that CPC-3 preferentially inhibits the expression of large proteins. It should be noted, however, our MS analysis preferentially identifies abundant proteins and failed to detect the vast majority of proteins encoded by mRNAs with poor codon usage (13).

#### Codon decoding rates are CDS length-dependent and are regulated by CPC-3

It was previously assumed that the same codons can be recognized and translated with similar efficiency on different mRNAs (21,22). The CDS length-dependent effect on ribosome density prompts us to examine whether codon decoding rate is also affected by CDS length. Thus, we compared RCDTs of all codons using the ribosome profiling data of the WT strain for mRNAs with long CDS regions (>600 aa) or short CDS regions (<300 aa). Remarkably, all codons have higher RCDTs for long CDS mRNAs than those of short CDS mRNAs (Figure 7C, D and Supplementary Figure S11). However, the difference of RCDT of each codon between long and short CDS mRNAs was much smaller in the  $\Delta cpc-3$  strain than in the WT strain (Figure 7E). These results suggest that CPC-3 also regulates the CDS length-dependent effect on codon decoding rate. By analyzing the previously published ribosome profiling results in *S. cerevisiae* (91), the similar CDS length-dependent effect on codon decoding rates was also observed for all codons (Figure 7F), suggesting that the effect of CDS length on elongation rate is conserved in eukaryotes. Because of the role of elongation rate in regulating co-translational protein folding and because folding of large proteins is more complicated than folding of small proteins, the CDS length-dependent regulation on translation is likely an adaptive mechanism that slows translation elongation to facilitate optimal co-translational folding of large proteins.

#### CPC-3 deletion resulted in increased sensitivity to heat shock treatment

Because translation kinetics influences co-translational protein folding (6,32,92,93), the deletion of *cpc-3* in *Neurospora* should broadly affect co-translational protein folding as

indicated by the EDP phosphorylation profile change. If so, proteins in the  $\Delta cpc-3$  strain may be more sensitive to conditions that trigger protein misfolding, resulting in impaired cell growth. To examine this possibility, we compared the growth rates of the WT and  $\Delta cpc-3$  strains at normal growth temperatures (20–34°C) and at 42°C. The 42°C treatment induces a heat shock response in *Neurospora* and impairs cell growth. The heat shock treatment should also make nascent proteins prone to be misfolded if their co-translational folding processes are not optimal. Although the two strains had a similar growth rate at 25°C, the  $\Delta cpc-3$  strain grew more rapidly than the WT strain at 34°C, which may be due to elevated protein translation (Figure 7G). At 42°C, however, the growth of the  $\Delta cpc-3$  strain was almost completely inhibited whereas the WT strain still exhibited modest growth. This result indicates that the deletion of *cpc-3* causes increased sensitivity to heat shock treatment, which is consistent with the role of CPC-3 in regulating translation kinetics, which in turn influences protein folding and function.

## DISCUSSION

In this study, we showed that codon optimality regulates protein translation in a CDS length-dependent manner by regulating both translation initiation rate and elongation speed. Analyses of the proteomic results from *Neurospora*, yeast, fly, worm and mouse showed that protein abundance negatively correlates with protein length genome-wide. The negative correlation, however, is dependent on codon usage: As codon optimality increases, the negative correlation progressively weakens. The conserved nature of this observation suggests a common mechanism mediated by codon usage that regulates protein synthesis in a protein size-dependent manner. Using gene reporters with different codon usage biases and different CDS lengths, we showed that non-optimal codon usage preferentially reduced the production levels of large proteins and that optimal codon usage eliminated the length-dependent effect on protein production in *Neurospora*. Gene functional enrichment analysis showed that there is a functional separation of gene functions based on codon usage and CDS length: The genes encoding long mRNAs with optimal codons are significantly enriched for functional categories of essential cellular processes, whereas those encoding long mRNAs with non-optimal codon usage are enriched for functional categories involved in regulatory processes. These results suggest that optimal codon usage is a mechanism that permits efficient production of large proteins critical for cell survival.

Further, we showed that codon optimality regulates ribosome density and ribosome flux on mRNA genome-wide in a CDS length-dependent manner. We showed that codon usage- and CDS length-dependent eIF2 $\alpha$  phosphorylation occurs in the absence of translation stress, suggesting a mechanism for how codon usage and CDS length regulate translation by feeding back on translation initiation (Figure 7H). We propose that rare codons cause ribosomes with empty A sites to pause. Such ribosome pausing results in CPC-3 activation and eIF2 $\alpha$  phosphorylation. For translation of mRNAs with the same codon usage, the paus-

ing occurs more often on long CDS mRNAs than that on short CDS mRNAs due to the existence of more rare codons, resulting in higher level of eIF2 $\alpha$  phosphorylation, which can inhibit translation initiation by blocking the formation of the pre-initiation complex. However, as in yeast and mammals, CPC-3 may also have substrates other than eIF2 $\alpha$ , such as the methionyl-tRNA synthetase. The phosphorylation of methionyl-tRNA synthetase has been shown to inhibit methionyl-tRNA synthetase activity, thus reinforcing the transient inhibition of translation initiation exerted by eIF2 $\alpha$  phosphorylation (94,95). Thus, the eIF2 $\alpha$  phosphorylation-independent function of CPC-3 may also contribute to its functions in translation.

Unlike the eIF2 $\alpha$  phosphorylation induction under stress conditions that results in the integrated stress response and global repression of translation (42,53–56), the codon usage-dependent and CDS length-dependent induction of eIF2 $\alpha$  phosphorylation is mRNA specific and does not cause a global increase of eIF2 $\alpha$  phosphorylation under nutrient replete growth conditions. Indeed, we showed that the deletion of *cpc-3*, which results in loss of eIF2 $\alpha$  phosphorylation, preferentially increased the abundances of large proteins encoded by mRNAs with non-optimal codon usage (Figure 4A and B). This result suggests that a high local concentration of phosphorylated eIF2 $\alpha$  specifically inhibits translation initiation of long CDS mRNAs with non-optimal codon usage. This notion is also consistent with our discovery that the correlation between codon usage and ribosome density increases as CDS length increases (Figure 2E). It is important to note that such an effect on ribosome density occurs despite of the known dramatic opposing effect of codon usage on ribosome occupancy because of its role in elongation rate in *Neurospora* (21,23), suggesting that codon usage feeds back on translation initiation in an mRNA-specific manner.

Consistent with our model that translation elongation feeds back to influence mRNA-specific translation initiation, it was previously proposed that translation initiation and elongation coordinate with each other to optimize protein production: mRNAs that encode high-abundance proteins usually have high translation initiation rates, fast elongation rates, and optimal codon usage (90). In addition, low levels of eIF2 $\alpha$  phosphorylation may have specific rather than broad effects on translation (50,62). Importantly, it was previously shown that certain chemical modifications of mRNA transcribed *in vitro* can specifically enhance its translation in cells through attenuating eIF2 $\alpha$  phosphorylation and increasing translation initiation (96). The specific effect of eIF2 $\alpha$  phosphorylation can be caused by the compartmentalization of translation or the local ribosome recycling for translation re-initiation of circularized mRNAs (97–100).

Our data indicate that the presence of rare codons activates CPC-3 to phosphorylate eIF2 $\alpha$ . The yeast and mammalian homolog GCN2 has been shown to be associated with ribosomes and such association is important for GCN2 activation (101–104). The interaction of GCN2 with the ribosomal P-stalk can potentially activate GCN2 in the absence of uncharged tRNA (59,60). Therefore, it is likely that rare codons trigger ribosome pausing, which may promote the interaction between GCN2 (or CPC-3) and ribo-

somes, resulting in the kinase activation and the subsequent eIF2 $\alpha$  phosphorylation. It was also previously shown that ribosomes have different conformations at different stages of the elongation cycle (84). It is possible that the ribosome conformation at a specific functional state with an empty A site promotes the interaction between ribosomes and CPC-3 and the latter activation.

As expected, deletion of *cpc-3* resulted in a general increase of protein synthesis, confirming the roles of eIF2 $\alpha$  phosphorylation in inhibiting translation initiation. Unexpectedly, however, deletion of *cpc-3* in *Neurospora* also had a major impact on translation elongation rate, and this effect was also dependent on codon usage and CDS length. The increased translation elongation rate in the  $\Delta$ *cpc-3* strain was demonstrated by three independent methods: the relative codon decoding rates determined by ribosome profiling, the *in vitro* translation assay that was used to directly compare translation elongation rates, and the *in vivo* protein conformation reporter that is sensitive to elongation rate change. Although the elongation rates of all codons were increased in the  $\Delta$ *cpc-3* strain, the effects were codon usage-dependent: Deletion of *cpc-3* preferentially increased elongation rates of mRNAs rich in rare codons, indicating that CPC-3 regulates elongation rates in a codon usage-dependent manner. Thus, in the WT strain, CPC-3 amplifies the codon usage effect on elongation speed so that codon decoding rates for optimal codons are much faster than those for rare codons.

The negative correlations between CDS length and protein abundance, translation initiation rate, and ribosome density suggest that increasing ORF length may decrease the translation initiation efficiency (22,36–40). Our results here suggest that both CPC-3-dependent and CPC-3-independent mechanisms are involved in the CDS length-dependent regulation on protein production. The CPC-3-independent mechanism may be due to the less efficient mRNA circularization, ribosome re-initiation or ribosome recycling for long CDS mRNAs than for short CDS mRNAs (39,41). Because protein production rate should be mostly determined by translation initiation rate on mRNAs unless there are significant amounts of translation abortion events (40), the up-regulated protein synthesis rates (including the 3  $\times$  WT EDP reporter) in the  $\Delta$ *cpc-3* strain suggest their increased translation initiation. However, if there are strong ribosome stalling or premature termination during translation elongation, the increase of the translation elongation can also promote translation efficiency. Our results showed that CPC-3 plays an important role in regulating translation elongation in addition to its role in regulating translation initiation. Therefore, for the feedback mechanism mediated by codon usage and CDS length, the effects of CPC-3 and eIF2 $\alpha$  phosphorylation on translation initiation should play an important role in regulating protein synthesis levels. In addition, the role of CPC-3 in translation elongation can also contribute to translation efficiency by regulating ribosome stalling or premature translation termination events (21,23).

Codon decoding rates are also regulated by CDS length: the rates are slower for mRNAs with long CDS regions and faster for mRNAs with short CDS regions. This phenomenon was observed in both *N. crassa* and *S. cerevisiae*,

suggesting a conserved mechanism regulating translation elongation speed in eukaryotes. Because the elongation rate regulates co-translation folding process, and large proteins have more structural domains and should be more prone to be misfolded than small proteins (93,105–107), a slow elongation rate likely promotes optimal co-translational folding of large proteins.

In higher eukaryotes, in addition to GCN2, the eIF2 $\alpha$  phosphorylation at Ser 51 can also be mediated by protein kinase R, PKR-like endoplasmic reticulum kinase and heme-regulated inhibitor (42,50,51). Therefore, it is possible that these additional kinases may also contribute to the feedback process from translation elongation to initiation. In addition, although our results demonstrated the involvement of CPC-3 in regulating the feedback mechanism from elongation to initiation in a codon usage and CDS length-dependent manner, CPC-3 independent mechanism may also exist.

Although how CPC-3 slows down elongation rate is not known, GCN2 in both yeast and mammalian cells has been shown to interact with the translation elongation factor eEF1A and this interaction keeps GCN2 inactive under nutrient-replete conditions (108–110). It is possible that this interaction also negatively influences the ability of eEF1A to deliver cognate aminoacyl tRNAs to the ribosomal A-site during elongation. Together, our results here suggest that translation elongation can feed back on both translation initiation and elongation kinetics through a mechanism that depends on codon usage and CDS length to allow optimal synthesis of proteins of different sizes.

## DATA AVAILABILITY

Ribosome profiling and RNA-seq data have been submitted to the NCBI Gene Expression Omnibus under accession number GSE168595. Customized scripts used for ribosome profiling analyses were deposited at <https://github.com/lxlsc0715/scripts-for-ribosome-profiling-and-RNA-seq>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank the members of our laboratory for technical assistance and discussion. We thank Andrew Lemoff and Mohammad Goodarzi at UT Southwestern core facility for MS analyses.

*Author contributions:* X.L. and Y.L. conceived the project. X.L., Q.Y. and Y.L. designed and planned experiments. X.L., Q.Y. and F.Z. performed experiments, collected and analyzed data. Y. L. and X. L. wrote the manuscript. All authors read and approved the manuscript.

## FUNDING

National Institutes of Health [R35GM118118]; Welch Foundation [I-1560 to Y.L.]. Funding for open access charge: NIGMS [R35GM118118].

*Conflict of interest statement.* None declared.

## REFERENCES

- Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, **2**, 13–34.
- Sharp, P.M. and Li, W.H. (1987) The codon adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.*, **15**, 1281–1295.
- Plotkin, J.B. and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, **12**, 32–42.
- Chaney, J.L. and Clark, P.L. (2015) Roles for synonymous codon usage in protein biogenesis. *Annu. Rev. Biophys.*, **44**, 143–166.
- Hanson, G. and Collier, J. (2018) Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.*, **19**, 20–30.
- Liu, Y. (2020) A code within the genetic code: codon usage regulates co-translational protein folding. *Cell Commun. Signal*, **18**, 145.
- Liu, Y., Yang, Q. and Zhao, F. (2021) Synonymous but not silent: the codon usage code for gene expression and protein folding. *Annu. Rev. Biochem.*, **90**, 375–401.
- Nieuwkoop, T., Finger-Bou, M., van der Oost, J. and Claessens, N.J. (2020) The ongoing quest to crack the genetic code for protein production. *Mol. Cell*, **80**, 193–209.
- Lyu, X. and Liu, Y. (2020) Nonoptimal codon usage is critical for protein structure and function of the master general amino acid control regulator CPC-1. *mBio*, **11**, e02605-20.
- Friberg, M., von Rohr, P. and Gonnet, G. (2004) Limitations of codon adaptation index and other coding DNA-based features for prediction of protein expression in *Saccharomyces cerevisiae*. *Yeast*, **21**, 1083–1093.
- Duret, L. and Mouchiroud, D. (1999) Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 4482–4487.
- Carlini, D.B. and Stephan, W. (2003) *In vivo* introduction of unpreferred synonymous codons into the *Drosophila Adh* gene results in reduced levels of ADH protein. *Genetics*, **163**, 239–243.
- Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C.H., Fu, J., Chen, S. and Liu, Y. (2016) Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E6117–E6125.
- Gustafsson, C., Govindarajan, S. and Minshull, J. (2004) Codon bias and heterologous protein expression. *Trends Biotechnol.*, **22**, 346–353.
- Xu, Y., Ma, P., Shah, P., Rokas, A., Liu, Y. and Johnson, C.H. (2013) Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature*, **495**, 116–120.
- Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J.M., Sachs, M.S. and Liu, Y. (2013) Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature*, **495**, 111–115.
- Hense, W., Anderson, N., Hutter, S., Stephan, W., Parsch, J. and Carlini, D.B. (2010) Experimentally increased codon bias in the *Drosophila Adh* gene leads to an increase in larval, but not adult, alcohol dehydrogenase activity. *Genetics*, **184**, 547–555.
- Lampson, B.L., Pershing, N.L., Prinz, J.A., Lacsina, J.R., Marzluff, W.F., Nicchitta, C.V., MacAlpine, D.M. and Counter, C.M. (2013) Rare codons regulate KRas oncogenesis. *Curr. Biol.*, **23**, 70–75.
- Fu, J., Dang, Y., Counter, C. and Liu, Y. (2018) Codon usage regulates human KRAS expression at both transcriptional and translational levels. *J. Biol. Chem.*, **293**, 17929–17940.
- Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sorensen, K.D. *et al.* (2014) A dual program for translation regulation in cellular proliferation and differentiation. *Cell*, **158**, 1281–1292.
- Yu, C.H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M.S. and Liu, Y. (2015) Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol. Cell*, **59**, 744–754.
- Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J.A., Plotkin, J.B. and Bartel, D.P. (2016) Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.*, **14**, 1787–1799.
- Yang, Q., Yu, C.H., Zhao, F., Dang, Y., Wu, C., Xie, P., Sachs, M.S. and Liu, Y. (2019) eRF1 mediates codon usage effects on mRNA



- translation efficiency through premature termination at rare codons. *Nucleic Acids Res.*, **47**, 9243–9258.
24. Hussmann, J.A., Patchett, S., Johnson, A., Sawyer, S. and Press, W.H. (2015) Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. *PLoS Genet.*, **11**, e1005732.
  25. Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R. *et al.* (2015) Codon optimality is a major determinant of mRNA stability. *Cell*, **160**, 1111–1124.
  26. Wu, Q., Medina, S.G., Kushawah, G., DeVore, M.L., Castellano, L.A., Hand, J.M., Wright, M. and Bazzini, A.A. (2019) Translation affects mRNA stability in a codon-dependent manner in human cells. *Elife*, **8**, e45396.
  27. Mordret, E., Dahan, O., Asraf, O., Rak, R., Yehonadav, A., Barnabas, G.D., Cox, J., Geiger, T., Lindner, A.B. and Pilpel, Y. (2019) Systematic detection of amino acid substitutions in proteomes reveals mechanistic basis of ribosome errors and selection for translation fidelity. *Mol. Cell*, **75**, 427–441.
  28. Mordstein, C., Savaisaar, R., Young, R.S., Bazile, J., Talmanc, L., Luft, J., Liss, M., Taylor, M.S., Hurst, L.D. and Kudla, G. (2020) Codon usage and splicing jointly influence mRNA localization. *Cell Syst.*, **10**, 351–362.
  29. Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V. and Gottesman, M.M. (2007) A “silent” polymorphism in the *MDR1* gene changes substrate specificity. *Science*, **315**, 525–528.
  30. Fu, J., Murphy, K.A., Zhou, M., Li, Y.H., Lam, V.H., Tabuloc, C.A., Chiu, J.C. and Liu, Y. (2016) Codon usage affects the structure and function of the *Drosophila* circadian clock protein PERIOD. *Genes Dev.*, **30**, 1761–1775.
  31. Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., Rodnina, M.V. and Komar, A.A. (2016) Synonymous codons direct cotranslational folding toward different protein conformations. *Mol. Cell*, **61**, 341–351.
  32. O’Brien, E.P., Vendruscolo, M. and Dobson, C.M. (2014) Kinetic modelling indicates that fast-translating codons can coordinate cotranslational protein folding by avoiding misfolded intermediates. *Nat. Commun.*, **5**, 2988.
  33. Zhou, Z., Dang, Y., Zhou, M., Yuan, H. and Liu, Y. (2018) Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation. *Elife*, **7**, e33569.
  34. Yang, Q., Lyu, X., Zhao, F. and Liu, Y. (2021) Effects of codon usage on gene expression are promoter context dependent. *Nucleic Acids Res.*, **49**, 818–831.
  35. Santiago, T.C., Purvis, I.J., Bettany, A.J. and Brown, A.J. (1986) The relationship between mRNA stability and length in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **14**, 8347–8360.
  36. Dao Duc, K. and Song, Y.S. (2018) The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation. *PLoS Genet.*, **14**, e1007166.
  37. Sharma, A.K., Sormanni, P., Ahmed, N., Ciryam, P., Friedrich, U.A., Kramer, G. and O’Brien, E.P. (2019) A chemical kinetic basis for measuring translation initiation and elongation rates from ribosome profiling data. *PLoS Comput. Biol.*, **15**, e1007070.
  38. Li, J.J., Chew, G.L. and Biggin, M.D. (2019) Quantitative principles of cis-translational control by general mRNA sequence features in eukaryotes. *Genome Biol.*, **20**, 162.
  39. Fernandes, L.D., Moura, A.P.S. and Ciandrini, L. (2017) Gene length as a regulator for ribosome recruitment and protein synthesis: theoretical insights. *Sci. Rep.*, **7**, 17409.
  40. Shah, P., Ding, Y., Niemczyk, M., Kudla, G. and Plotkin, J.B. (2013) Rate-limiting steps in yeast protein translation. *Cell*, **153**, 1589–1601.
  41. Rogers, D.W., Bottcher, M.A., Traulsen, A. and Greig, D. (2017) Ribosome reinitiation can explain length-dependent translation of messenger RNA. *PLoS Comput. Biol.*, **13**, e1005592.
  42. Sonenberg, N. and Hinnebusch, A.G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, **136**, 731–745.
  43. Jackson, R.J., Hellen, C.U. and Pestova, T.V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.*, **11**, 113–127.
  44. Hinnebusch, A.G. (2014) The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.*, **83**, 779–812.
  45. Jin, H., Bjornsson, A. and Isaksson, L.A. (2002) Cis control of gene expression in *E. coli* by ribosome queuing at an inefficient translational stop signal. *EMBO J.*, **21**, 4357–4367.
  46. Diamant, A., Feldman, A., Schochet, E., Kupiec, M., Arava, Y. and Tuller, T. (2018) The extent of ribosome queuing in budding yeast. *PLoS Comput. Biol.*, **14**, e1005951.
  47. Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M.F. and Haar, T. (2014) Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J.*, **33**, 21–34.
  48. Kudla, G., Murray, A.W., Tollervey, D. and Plotkin, J.B. (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, **324**, 255–258.
  49. Pop, C., Rouskin, S., Ingolia, N.T., Han, L., Phizicky, E.M., Weissman, J.S. and Koller, D. (2014) Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol. Syst. Biol.*, **10**, 770.
  50. Boye, E. and Grallert, B. (2020) eIF2alpha phosphorylation and the regulation of translation. *Curr. Genet.*, **66**, 293–297.
  51. Masson, G.R. (2019) Towards a model of GCN2 activation. *Biochem. Soc. Trans.*, **47**, 1481–1488.
  52. Karki, S., Castillo, K., Ding, Z., Kerr, O., Lamb, T.M., Wu, C., Sachs, M.S. and Bell-Pedersen, D. (2020) Circadian clock control of eIF2alpha phosphorylation is necessary for rhythmic translation initiation. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 10935–10945.
  53. Pakos-Zebrucka, K., Koryga, I., Mnich, K., Ljubic, M., Samali, A. and Gorman, A.M. (2016) The integrated stress response. *EMBO Rep.*, **17**, 1374–1395.
  54. Yang, R., Wek, S.A. and Wek, R.C. (2000) Glucose limitation induces GCN4 translation by activation of Gcn2 protein kinase. *Mol. Cell Biol.*, **20**, 2706–2717.
  55. Dong, J., Qiu, H., Garcia-Barrio, M., Anderson, J. and Hinnebusch, A.G. (2000) Uncharged tRNA activates GCN2 by displacing the protein kinase moiety from a bipartite tRNA-binding domain. *Mol. Cell*, **6**, 269–279.
  56. Deng, J., Harding, H.P., Raught, B., Gingras, A.C., Berlanga, J.J., Scheuener, D., Kaufman, R.J., Ron, D. and Sonenberg, N. (2002) Activation of GCN2 in UV-irradiated cells inhibits translation. *Curr. Biol.*, **12**, 1279–1286.
  57. Ishimura, R., Nagy, G., Dotu, I., Chuang, J.H. and Ackerman, S.L. (2016) Activation of GCN2 kinase by ribosome stalling links translation elongation with translation initiation. *Elife*, **5**, e14295.
  58. Lyu, X., Yang, Q., Li, L., Dang, Y., Zhou, Z., Chen, S. and Liu, Y. (2020) Adaptation of codon usage to tRNA I34 modification controls translation kinetics and proteome landscape. *PLoS Genet.*, **16**, e1008836.
  59. Jimenez-Diaz, A., Remacha, M., Ballesta, J.P. and Berlanga, J.J. (2013) Phosphorylation of initiation factor eIF2 in response to stress conditions is mediated by acidic ribosomal P1/P2 proteins in *Saccharomyces cerevisiae*. *PLoS One*, **8**, e84219.
  60. Inglis, A.J., Masson, G.R., Shao, S., Perisic, O., McLaughlin, S.H., Hegde, R.S. and Williams, R.L. (2019) Activation of GCN2 by the ribosomal P-stalk. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 4946–4954.
  61. Harding, H.P., Ordonez, A., Allen, F., Parts, L., Inglis, A.J., Williams, R.L. and Ron, D. (2019) The ribosomal P-stalk couples amino acid starvation to GCN2 activation in mammalian cells. *Elife*, **8**, e50149.
  62. Stonyte, V., Boye, E. and Grallert, B. (2018) Regulation of global translation during the cell cycle. *J. Cell Sci.*, **131**, jcs220327.
  63. Zhou, M., Wang, T., Fu, J., Xiao, G. and Liu, Y. (2015) Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol. Microbiol.*, **97**, 974–987.
  64. Zhao, F., Zhou, Z., Dang, Y., Na, H., Adam, C., Lipzen, A., Ng, V., Grigoriev, I.V. and Liu, Y. (2021) Genome-wide role of codon usage on transcription and identification of potential regulators. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2022590118.
  65. Colot, H.V., Park, G., Turner, G.E., Ringelberg, C., Crew, C.M., Litvinkova, L., Weiss, R.L., Borkovich, K.A. and Dunlap, J.C. (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 10352–10357.
  66. Bardiya, N. and Shiu, P.K. (2007) Cyclosporin A-resistance based gene placement system for *Neurospora crassa*. *Fungal Genet. Biol.*, **44**, 307–314.

67. Wu, C., Amrani, N., Jacobson, A. and Sachs, M.S. (2007) The use of fungal *in vitro* systems for studying translational regulation. *Methods Enzymol.*, **429**, 203–225.
68. Lee, H.C., Chang, S.S., Choudhary, S., Aalto, A.P., Maiti, M., Bamford, D.H. and Liu, Y. (2009) qiRNA is a new type of small interfering RNA induced by DNA damage. *Nature*, **459**, 274–277.
69. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N. and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics*, **13**, 2513–2526.
70. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337–342.
71. Krey, J.F., Scheffer, D.I., Choi, D., Reddy, A., David, L.L., Corey, D.P. and Barr-Gillespie, P.G. (2018) Mass spectrometry quantitation of proteins from small pools of developing auditory and vestibular cells. *Sci Data*, **5**, 180128.
72. Krey, J.F., Wilmarth, P.A., Shin, J.B., Klimek, J., Sherman, N.E., Jeffery, E.D., Choi, D., David, L.L. and Barr-Gillespie, P.G. (2014) Accurate label-free protein quantitation with high- and low-resolution mass spectrometers. *J. Proteome Res.*, **13**, 1034–1044.
73. Sabi, R., Volovitch Daniel, R. and Tuller, T. (2017) stAICalc: tRNA adaptation index calculator based on species-specific weights. *Bioinformatics*, **33**, 589–591.
74. Li, G.W., Burkhardt, D., Gross, C. and Weissman, J.S. (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, **157**, 624–635.
75. Aradska, J., Bulat, T., Sialana, F.J., Birner-Gruenberger, R., Erich, B. and Lubec, G. (2015) Gel-free mass spectrometry analysis of *Drosophila melanogaster* heads. *Proteomics*, **15**, 3356–3360.
76. Samson, M., Jow, M.M., Wong, C.C., Fitzpatrick, C., Aslanian, A., Saucedo, I., Estrada, R., Ito, T., Park, S.K., Yates, J.R. 3rd *et al.* (2014) The specification and global reprogramming of histone epigenetic marks during gamete formation and early embryo development in *C. elegans*. *PLoS Genet.*, **10**, e1004588.
77. Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villen, J., Haas, W., Sowa, M.E. and Gygi, S.P. (2010) A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell*, **143**, 1174–1189.
78. Bennetzen, J.L. and Hall, B.D. (1982) Codon selection in yeast. *J. Biol. Chem.*, **257**, 3026–3031.
79. dos Reis, M., Savva, R. and Wernisch, L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**, 5036–5044.
80. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. and Weissman, J.S. (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.
81. Ingolia, N.T., Lareau, L.F. and Weissman, J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.
82. Kasuga, T., Mannhaupt, G. and Glass, N.L. (2009) Relationship between phylogenetic distribution and genomic features in *Neurospora crassa*. *PLoS One*, **4**, e5286.
83. Jenner, L., Starosta, A.L., Terry, D.S., Mikolajka, A., Filonava, L., Yusupov, M., Blanchard, S.C., Wilson, D.N. and Yusupova, G. (2013) Structural basis for potent inhibitory activity of the antibiotic tigeicycline during protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 3812–3816.
84. Wu, C.C., Zinshteyn, B., Wehner, K.A. and Green, R. (2019) High-resolution ribosome profiling defines discrete ribosome elongation states and translational regulation during cellular stress. *Mol. Cell*, **73**, 959–970.
85. Enam, S.U., Zinshteyn, B., Goldman, D.H., Cassani, M., Livingston, N.M., Seydoux, G. and Green, R. (2020) Puromycin reactivity does not accurately localize translation at the subcellular level. *Elife*, **9**, e60303.
86. Yan, L.L. and Zaher, H.S. (2021) Ribosome quality control antagonizes the activation of the integrated stress response on colliding ribosomes. *Mol. Cell*, **81**, 614–628.
87. Hood, H.M., Neafsey, D.E., Galagan, J. and Sachs, M.S. (2009) Evolutionary roles of upstream open reading frames in mediating gene regulation in fungi. *Annu. Rev. Microbiol.*, **63**, 385–409.
88. Wu, C., Dasgupta, A., Shen, L., Bell-Pedersen, D. and Sachs, M.S. (2018) The cell free protein synthesis system from the model filamentous fungus *Neurospora crassa*. *Methods*, **137**, 11–19.
89. Kolb, V.A., Makeyev, E.V. and Spirin, A.S. (1994) Folding of firefly luciferase during translation in a cell-free system. *EMBO J.*, **13**, 3631–3637.
90. Riba, A., Di Nanni, N., Mittal, N., Arhne, E., Schmidt, A. and Zavolan, M. (2019) Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 15023–15032.
91. Young, D.J., Guydosh, N.R., Zhang, F., Hinnebusch, A.G. and Green, R. (2015) Rli1/ABCE1 recycles terminating ribosomes and controls translation reinitiation in 3' UTRs *in vivo*. *Cell*, **162**, 872–884.
92. Purvis, I.J., Bettany, A.J., Santiago, T.C., Coggins, J.R., Duncan, K., Eason, R. and Brown, A.J. (1987) The efficiency of folding of some proteins is increased by controlled rates of translation *in vivo*. A hypothesis. *J. Mol. Biol.*, **193**, 413–417.
93. O'Brien, E.P., Vendruscolo, M. and Dobson, C.M. (2012) Prediction of variable translation rate effects on cotranslational protein folding. *Nat. Commun.*, **3**, 868.
94. Kwon, N.H., Kang, T., Lee, J.Y., Kim, H.H., Kim, H.R., Hong, J., Oh, Y.S., Han, J.M., Ku, M.J., Lee, S.Y. *et al.* (2011) Dual role of methionyl-tRNA synthetase in the regulation of translation and tumor suppressor activity of aminoacyl-tRNA synthetase-interacting multifunctional protein-3. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 19635–19640.
95. Murguía, J.R. and Serrano, R. (2012) New functions of protein kinase Gcn2 in yeast and mammals. *IUBMB Life*, **64**, 971–974.
96. Svitkin, Y.V., Cheng, Y.M., Chakraborty, T., Presnyak, V., John, M. and Sonenberg, N. (2017) N1-methyl-pseudouridine in mRNA enhances translation through eIF2 $\alpha$ -dependent and independent mechanisms by increasing ribosome density. *Nucleic Acids Res.*, **45**, 6023–6036.
97. Chouaib, R., Safieddine, A., Pichon, X., Imbert, A., Kwon, O.S., Samacoits, A., Traboulsi, A.M., Robert, M.C., Tsanov, N., Coleno, E. *et al.* (2020) A dual protein-mRNA localization screen reveals compartmentalized translation and widespread co-translational RNA targeting. *Dev. Cell*, **54**, 773–791.
98. Pichon, X., Bastide, A., Safieddine, A., Chouaib, R., Samacoits, A., Basyuk, E., Peter, M., Mueller, F. and Bertrand, E. (2016) Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. *J. Cell Biol.*, **214**, 769–781.
99. Wells, S.E., Hillner, P.E., Vale, R.D. and Sachs, A.B. (1998) Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell*, **2**, 135–140.
100. Marshall, E., Stansfield, I. and Romano, M.C. (2014) Ribosome recycling induces optimal translation rate at low ribosomal availability. *J. R. Soc. Interface*, **11**, 20140589.
101. Sattlegger, E. and Hinnebusch, A.G. (2000) Separate domains in GCN1 for binding protein kinase GCN2 and ribosomes are required for GCN2 activation in amino acid-starved cells. *EMBO J.*, **19**, 6622–6633.
102. Sattlegger, E. and Hinnebusch, A.G. (2005) Polyribosome binding by GCN1 is required for full activation of eukaryotic translation initiation factor 2 $\alpha$  kinase GCN2 during amino acid starvation. *J. Biol. Chem.*, **280**, 16514–16521.
103. Ramirez, M., Wek, R.C. and Hinnebusch, A.G. (1991) Ribosome association of GCN2 protein kinase, a translational activator of the GCN4 gene of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **11**, 3027–3036.
104. Zhu, S. and Wek, R.C. (1998) Ribosome-binding domain of eukaryotic initiation factor-2 kinase GCN2 facilitates translation control. *J. Biol. Chem.*, **273**, 1808–1814.
105. Cassaignau, A.M.E., Cabrita, L.D. and Christodoulou, J. (2020) How does the ribosome fold the proteome? *Annu. Rev. Biochem.*, **89**, 389–415.
106. Holtkamp, W., Kocic, G., Jager, M., Mittelstaet, J., Komar, A.A. and Rodnina, M.V. (2015) Cotranslational protein folding on the ribosome monitored in real time. *Science*, **350**, 1104–1107.
107. Guinn, E.J., Tian, P., Shin, M., Best, R.B. and Marqusee, S. (2018) A small single-domain protein folds through the same pathway on and off the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 12206–12211.

108. Ramesh,R. and Sattlegger,E. (2020) Domain II of the translation elongation factor eEF1A is required for Gcn2 kinase inhibition. *FEBS Lett.*, **594**, 2266–2281.
109. Visweswaraiah,J., Lageix,S., Castilho,B.A., Izotova,L., Kinzy,T.G., Hinnebusch,A.G. and Sattlegger,E. (2011) Evidence that eukaryotic translation elongation factor 1A (eEF1A) binds the Gcn2 protein C terminus and inhibits Gcn2 activity. *J. Biol. Chem.*, **286**, 36568–36579.
110. Silva,R.C., Sattlegger,E. and Castilho,B.A. (2016) Perturbations in actin dynamics reconfigure protein complexes that modulate GCN2 activity and promote an eIF2 response. *J. Cell Sci.*, **129**, 4521–4533.