

Massively parallel identification of sequence motifs triggering ribosome-associated mRNA quality control

Katharine Y. Chen^{1,2}, Heungwon Park¹ and Arvind Rasi Subramaniam^{1,*}

¹Basic Sciences Division and Computational Biology Section of the Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA 98109, USA

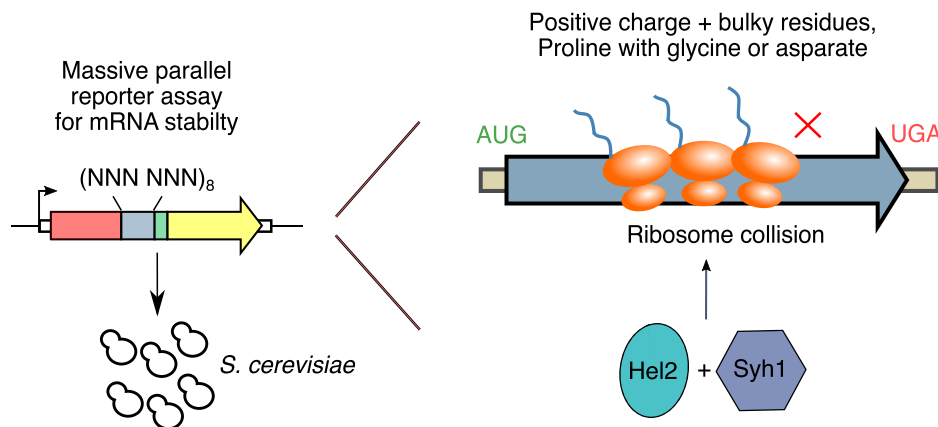
²Molecular and Cellular Biology Program, University of Washington, Seattle, WA 98195, USA

*To whom correspondence should be addressed. Tel: +1 206 667 4679; Fax: +1 206 606 1025; Email: rasi@fredhutch.org

Abstract

Decay of mRNAs can be triggered by ribosome slowdown at stretches of rare codons or positively charged amino acids. However, the full diversity of sequences that trigger co-translational mRNA decay is poorly understood. To comprehensively identify sequence motifs that trigger mRNA decay, we use a massively parallel reporter assay to measure the effect of all possible combinations of codon pairs on mRNA levels in *S. cerevisiae*. In addition to known mRNA-destabilizing sequences, we identify several dipeptide repeats whose translation reduces mRNA levels. These include combinations of positively charged and bulky residues, as well as proline-glycine and proline-aspartate dipeptide repeats. Genetic deletion of the ribosome collision sensor Hel2 rescues the mRNA effects of these motifs, suggesting that they trigger ribosome slowdown and activate the ribosome-associated quality control (RQC) pathway. Deep mutational scanning of an mRNA-destabilizing dipeptide repeat reveals a complex interplay between the charge, bulkiness, and location of amino acid residues in conferring mRNA instability. Finally, we show that the mRNA effects of codon pairs are predictive of the effects of endogenous sequences. Our work highlights the complexity of sequence motifs driving co-translational mRNA decay in eukaryotes, and presents a high throughput approach to dissect their requirements at the codon level.

Graphical abstract



Introduction

Translation and decay of mRNA are fundamental stages of gene expression whose interplay is crucial for determining steady-state protein levels in the cell. The protein coding region of mRNA has been recently recognized as an important determinant of mRNA stability (1–10). Ribosome elongation rates can vary along the protein coding region, which is sensed by diverse regulatory factors to trigger mRNA decay (7–20). Dysregulation of mRNA decay pathways has been linked to neurological diseases, autoinflammatory diseases, and cancer (21–26).

Several motifs in the protein coding region of eukaryotic mRNAs have been associated with changes in mRNA sta-

bility (3,4,6,9–13,27,28). Non-optimal codons decrease ribosome elongation rates and trigger Not5-dependent mRNA deadenylation and decay (3,29–31). Strong ribosome stalls caused by polybasic residues, poly-tryptophan sequences, and rare codon repeats trigger ribosome collisions and Hel2-dependent ribosome-associated mRNA quality control (henceforth RQC) (6,7,10,19,20,28,32,33). Poly-proline sequences decrease ribosome elongation rate, but such slowdowns are thought to be resolved by eIF5A and not trigger mRNA quality control (34,35). Ribosome profiling studies have identified several dipeptide and tripeptide motifs that are enriched at sites of ribosome stalls and collisions (36–39). However, whether such motifs are sufficient to trigger

Received: September 21, 2023. Revised: February 28, 2024. Editorial Decision: March 30, 2024. Accepted: April 4, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

mRNA quality control is not known. Ribosome stalling motifs in endogenous protein coding sequences often depend on a complex combination of amino acid residues in the nascent peptide (40–44), and thus their relation to the simple repeat stalling sequences studied in reporter assays is not clear.

We recently developed a massively parallel reporter assay to identify coding sequence motifs triggering mRNA decay in human cells (27). Using this assay, we found that translation of a diverse set of dipeptide repeats composed of bulky and positively charged amino acids are sufficient to trigger mRNA decay in human cells. Nevertheless, the molecular mechanism by which translation of these dipeptide repeats triggers mRNA decay in human cells remains unknown. Further, the extent to which translation of bulky and positively charged residues serves as an evolutionarily conserved signal for mRNA decay in other eukaryotes is unclear. Since co-translational mRNA decay pathways have been extensively studied in the budding yeast *Saccharomyces cerevisiae* (7,45–48), we sought to use this as an experimental model to dissect the molecular mechanism and sequence requirements of coding sequence-dependent mRNA decay. By extending our massively parallel reporter assay from human cells to *S. cerevisiae*, we identify several mRNA-destabilizing dipeptide motifs including combinations of bulky and positively charged residues, as well as proline-glycine and proline-aspartic acid dipeptide repeats. We define Hel2-dependent RQC as the major pathway regulating mRNA decay triggered by translation of these dipeptide repeats. Using deep mutational scanning, we further characterize the biochemical requirements at the codon level for bulky and positively charged dipeptide repeats in triggering Hel2-dependent mRNA decay. Together, our results highlight the diversity of coding sequence motifs triggering co-translational mRNA decay in *S. cerevisiae*, define the biochemical requirements for their mRNA-destabilizing effects, and reveal the extent of evolutionary conservation of these motifs across eukaryotes.

Materials and methods

Parent vector construction

Plasmids constructed and used in this study are listed in [Supplementary Table S1](#). Oligonucleotides used in this study are listed in [Supplementary Table S3](#). Plasmid assembly was carried out using standard molecular biology techniques as described below. All polymerase chain reaction (PCR) reactions were performed using Phusion polymerase (Thermo Fisher F530S) or Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher F548L) according to manufacturer's instructions. Restriction enzymes were obtained from Thermo Fisher and FastDigest (FD) variants were used when available.

The chrI-integrating parent vector pHPSC1120 used for this study was constructed from pHPSC417 used in our previous work (20). In comparison to pHPSC417, pHPSC1120 contains an additional Illumina Read1 primer binding site and T7 promoter sequences for deep sequencing of inserts and barcode sequences and for *in vitro* transcription from genomic DNA, respectively. The Illumina R1 sequencing primer binding and T7 promoter sequences were PCR-amplified using oHP558 as the forward primer, oHP530 as a bridge primer, and oHP529 as a reverse primer, and cloned into BamHI-linearized pHPSC417 using Gibson assembly. The –1 frameshifted parent vector pHPSC1114 was also constructed

from pHPSC417 using the same strategy as for pHPSC1120 but with a different forward primer oHP528 that incorporates the frameshift. All plasmids were verified by Sanger sequencing.

Variable oligo pool design

Pool 1

Pool 1 includes the 8 × dicodon library (4096 codon pair inserts) and the endogenous gene fragments library (1904 inserts). The 8 × dicodon library (Figure 1A) encodes all possible codon pair (6 nucleotides) combinations, for a total of 4096 codon pairs. Each codon pair is repeated eight times to create 48 nucleotide (nt) inserts. The endogenous gene fragments library includes 1904 endogenous fragments, each 48 nt in length (Figure 6A). Endogenous gene fragments were selected as 253 nt to 300 nt of each ORF. Only ORFs designated as 'Verified' by the *Saccharomyces* Genome Database (SGD) in the R64-1-1 release were included (http://sgd-archive.yeastgenome.org/sequence/S288C_reference/genome_releases/). Every second gene in descending order of RNA expression (49) was included in this library to encompass a wide range of expression levels. All 6000 inserts are flanked with the same 29 nt 5' homology arm and 24 nt 3' homology arm. The oligo pool (oAS385) was ordered from Twist Biotechnologies.

Pool 2

The FK₈ deep mutational scanning library (Figure 5A) was constructed from a starting sequence composed of phenylalanine and lysine codons repeated eight times in tandem (48 nt inserts). The phenylalanine codons TTT and TTC and the lysine codons AAA and AAG were used interchangeably throughout the insert to avoid producing a repetitive mRNA sequence. At each of the 16 positions, an NNN sequence was used to randomize the codon. The oligo pool (oKC224) was ordered as an oPool from Integrated DNA Technologies.

Pool 3

Pool 3 includes various 8 × codon pair repeats encoding dipeptides determined to be destabilizing in Figure 3C. As in Pool 1, each codon pair is repeated eight times to create 48 nt inserts. The oligo pool (oKC265) was ordered from Integrated DNA Technologies.

Pool 4

Pool 4 includes various 8 × codon pair repeats encoding dipeptides that were not destabilizing in the original screen. As in Pool 1, each codon pair is repeated eight times to create 48 nt inserts. The oligo pool of these negative control pairs (oKC147) was ordered from Integrated DNA Technologies.

Plasmid library construction

For the 8 × dicodon library, oligo pool 1 (described above) was PCR-amplified with oKC97 and oHP531. For the –1 frameshifted 8 × dicodon library, pool 1 was PCR-amplified with oHP532 and oHP531. As described above, oHP531 encodes a 24 nt random barcode region, comprised of 8 × VNN repeats. Barcoded oligo pools were cloned into BamHI-linearized pHPSC1120 and pHPSC1114 by Gibson assembly. Assembled plasmid pools were transformed at high efficiency into NEB 10-Beta *Escherichia coli* cells, and plated as 1:10 serial dilutions. 500 000 colonies were scraped from plates

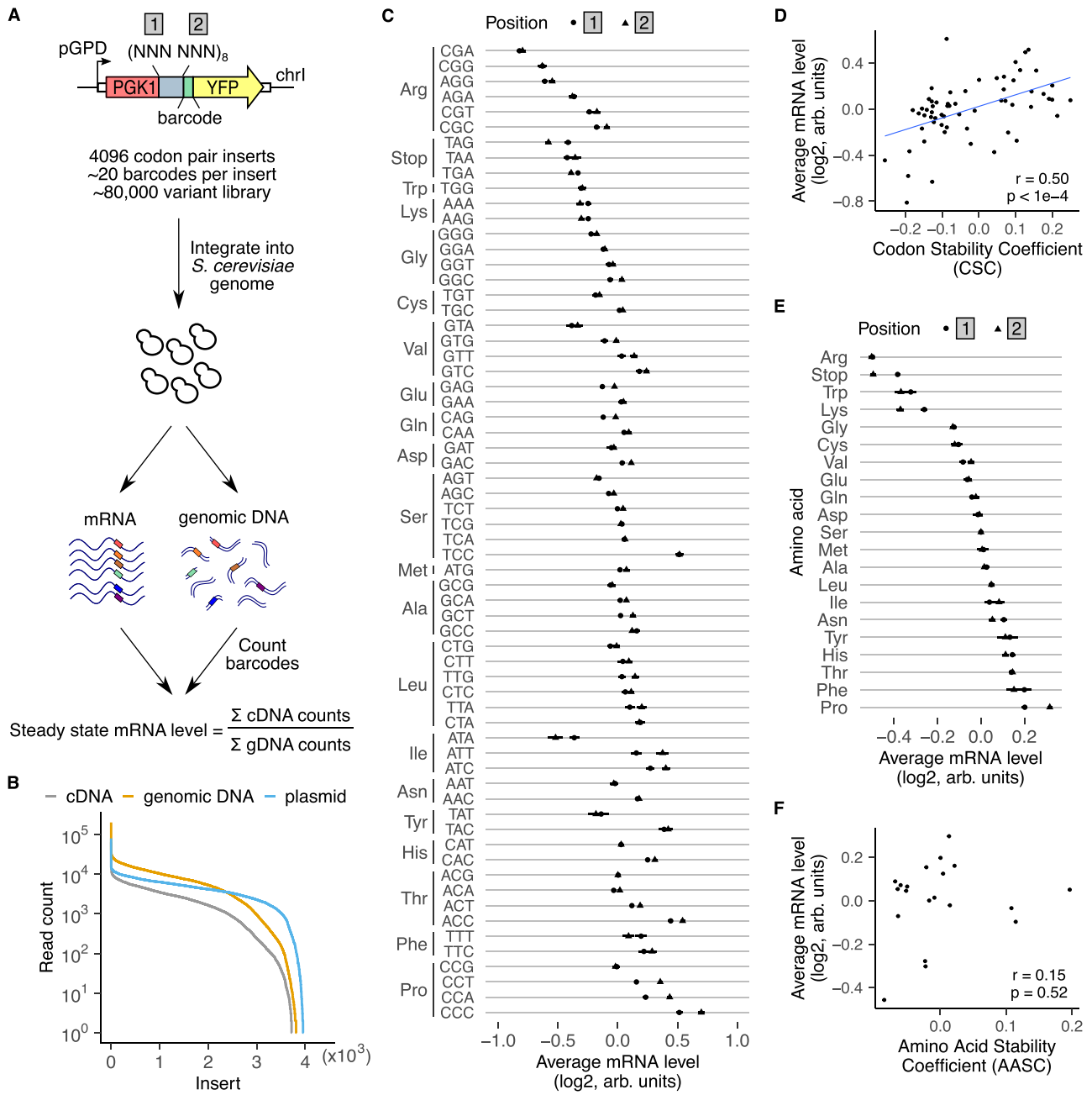


Figure 1. A massively parallel reporter assay for mRNA effects in *S. cerevisiae*. **(A)** Assay design. Each element in the library includes one of 4096 possible combinations of codon pairs repeated eight times. Each repeat is inserted in-frame between *PGK1* and *YFP*, and is followed by a random 24 nt barcode without in-frame stop codons (median of 20 barcodes/insert). The 80 000 variant library is integrated as a pool into a noncoding region of chromosome I. The barcodes in cDNA and genomic DNA are counted by high throughput amplicon sequencing. Relative steady state mRNA effect of each insert is calculated by first normalizing cDNA counts by genomic DNA counts for all barcodes linked to that insert and then by median-normalizing across all codon pairs. **(B)** Distribution of reads per codon pair insert for cDNA, genomic DNA, and plasmid libraries. **(C)** Average mRNA level of reporters with indicated codons in position 1 (circles) or position 2 (triangles) of the codon pair. **(D)** Average mRNA effects of individual codons compared against codon stability coefficients derived from endogenous *S. cerevisiae* mRNAs (3). **(E)** Average mRNA level of reporters encoding the indicated amino acid in position 1 (circles) or position 2 (triangles) of the codon pair. Error bars in C and E represent standard deviation over all variants containing the codon or amino acid at each position. Average mRNA levels in C and E are median-normalized over all codons or amino acids at each position. **(F)** Same as (D), except for amino acids compared against amino acid stability coefficients (4).

for extraction in order to bottleneck the number of unique variants.

Pool 2 was PCR-amplified with oKC97 and oKC225 and cloned into BamHI-linearized pHPSC1120 by Gibson assembly. The assembled plasmid pool was transformed at high efficiency into NEB 10-Beta *E. coli* cells. 70 000 colonies were scraped from plates for extraction in order to bottleneck the number of unique variants.

For the small 8 × dicodon library in various RQC deletion strains (Supplementary Figure S4A), oligo pool 3 and 4 (described above) were PCR-amplified individually with oKC97 and oHP531. PCR products were pooled together at an equimolar ratio and cloned into BamHI-linearized pHPSC1120 by Gibson assembly. The assembled plasmid pool was transformed at high efficiency into NEB 10-Beta *E. coli* cells. 10 000 colonies were scraped from different plates and extracted separately in order to keep the barcode and insert combinations unique between different plasmid libraries.

Individual plasmid construction

To generate the *PGK1-YFP* reporters used for flow cytometry of individually selected codon pairs, the desired codon pair inserts were amplified using two rounds of PCR from a pooled plasmid template pHPSC1136 not used in this study. Unique primers (oKC129-142) were used to amplify the six desired inserts. Homology arms were added to the six amplified inserts using oKC97 and oKC123 primers. Amplified inserts were cloned into BamHI-linearized pHPSC1120 by Gibson assembly to produce pHPSC1144, pHPSC1145, pHPSC1146, pHPSC1147, pHPSC1149, pHPSC1150 plasmids. All individual plasmids were verified by Sanger sequencing.

To create the small barcoded pool for mRNA measurement validation (Supplementary Figure S2A,E), oKC97 and oKC148 oligos were used to barcode and amplify inserts from the following plasmids (described above): pHPSC1144, pHPSC1145, pHPSC1146, pHPSC1147, pHPSC1149, pHPSC1150. oKC148 encodes a 24 nt random barcode region, comprised of 8 × VNN repeats. Barcoded inserts were then pooled at equimolar concentrations and cloned into BamHI-linearized pHPSC1120 by Gibson assembly. The assembled plasmid pool was transformed at high efficiency into NEB 10-Beta *E. coli* cells. 2000 colonies were scraped from plates for extraction in order to bottleneck the number of unique variants. Two colonies were picked and Sanger sequenced to obtain the identity of the insert and barcode pair of the two spike-in plasmids, pHPSC1159-sc2 and pHPSC1159-sc5.

Strain construction

All *S. cerevisiae* strains used in this study are listed in Supplementary Table S2. Integration of pooled plasmids into the *S. cerevisiae* genome was performed by transforming 30–200 µg of NotI-linearized plasmid library into 1–5e9 cells according to the LiAc/SS carrier DNA/PEG method (50). Following heat shock, cells were transferred into a 5× volume of a 1:1 solution of 20% dextrose and synthetic complete (SC) media lacking uracil with 2% dextrose (SCD-URA) and spun at 1850g for 5 minutes. Cell pellets were gently resuspended in 100 ml of fresh SCD-URA and allowed to recover overnight at 30°C shaking at 200 rpm. After 20–24 h, 1e9 cells were passaged into 100ml fresh SCD-URA and grown overnight at 30°C shaking at 200 rpm. This process was repeated for a total of 72 h of selection in SCD-URA before making glycerol

stocks from saturated cultures. Integration of individual constructs into the *S. cerevisiae* genome was performed by transforming 0.5–1.0 µg of linearized plasmid according to the LiAc/SS carrier DNA/PEG method (50). Single yeast colonies were selected on SCD agar plates lacking uracil after 48–72 h growth at 30°C.

Harvesting pooled library cells

Glycerol stocks of cells containing pooled reporter strains were thawed and grown overnight in 20–50 ml YEPD at starting OD₆₀₀ between 0.1 and 0.5 at 30°C with shaking at 200 rpm. The saturated cultures were diluted approximately 200-fold (for starting OD₆₀₀ of 0.1) and spike-in strains (scKC190 and scKC191) were introduced into each culture at a concentration approximately the same as each library variant based on OD₆₀₀ density. Cultures were grown for 4–6 h at 30°C with shaking at 200 rpm until mid-log phase (OD₆₀₀ between 0.4 and 0.6), then transferred to ice-water baths. Each culture was split into 50 ml aliquots (approximately ≥200 million cells) in pre-chilled conical tubes and spun down at 3000g, 4°C, for 5 min. The supernatant was removed and the cell pellets were flash-frozen in a dry ice-ethanol bath and stored at –80°C.

Harvesting glucose-depleted cells

Glycerol stocks of cells containing the pHPSC1142 pooled reporter library were thawed and grown overnight as described above. Saturated cultures were diluted and spike-in strains (scKC190 and scKC191) were introduced as described above. Cells were grown for 4 h at 30°C with shaking at 200rpm until OD₆₀₀ of 0.4. Cells were spun down at 3000 rpm for 2 min and washed with 30 ml H₂O twice, then resuspended into YEP (no glucose). Glucose-depleted cells were grown for 1 hour at 30°C with shaking at 200rpm. After 1 hour of growth, cells were harvested by spinning in 50ml pre-chilled tubes at 3000g, 4°C, for 5 min. The supernatant was removed and the cell pellets were flash-frozen in a dry ice-ethanol bath and stored at –80°C.

Harvesting pooled RQC deletion strains

Glycerol stocks of RQC deletion strains each containing one of the small 8 × dicodon libraries (pHPSC1165-71) were thawed and grown overnight individually in 20–50 ml YEPD at starting OD₆₀₀ between 0.1 and 0.5 at 30°C with shaking at 200rpm. Saturated cultures were pooled together at 1:1 based on optical density measurements and diluted to OD₆₀₀ of 0.1. The pooled culture was grown to mid-log phase and harvested as described above.

Library genomic DNA extraction

For genomic DNA extraction, between 400 million to 1.2 billion cells (two to six flash-frozen pellets) were lysed and extracted using the YeaStar Genomic DNA kit (Zymo 11-323), following the manufacturer's instructions, with 240 µl YD digestion buffer and 10 µl R-Zymolyase per pellet. Extracted genomic DNA was sheared for 10 min (30 s on, 30 s off, on 'High' setting) on ice using a Diagenode Bioruptor. Sheared gDNA was cleaned using DNA Binding Buffer (Zymo ZD4004-1-L) and UPrep Spin Columns (Genesee Scientific 88-143). Sheared and cleaned gDNA was then *in vitro* transcribed into RNA (denoted gRNA below and in analysis

code) starting from the T7 promoter region in the insert cassette, similar to previous approaches (27,51), using the HiScribe T7 High Yield RNA Synthesis Kit (NEB E2040S). Transcribed gRNA was cleaned using the RNA Clean and Concentrator kit (Zymo R1013).

Library mRNA extraction

At least 200 million cells (one flash-frozen pellet) per sample was resuspended in 400 μ l Trizol (Thermo Fisher 15596-026) in a 1.5-ml tube and vortexed with 500 μ l of glass beads (Sigma G8772) at 4°C for 10 min (2 min on, 1 minute on ice). RNA was extracted from the resulting lysate using the Direct-zol RNA Miniprep Kit (Zymo R2070) following manufacturer's instructions.

mRNA and genomic DNA barcode sequencing

For pHPSC1142, pHPSC1117 and pHPSC1160 libraries, between 0.5–10 μ g of mRNA and gRNA for each library was reverse transcribed into cDNA using SuperScript IV (Thermo Fisher 18090050) and a primer annealing to the Illumina R1 primer binding site (oPB354). A 170 nt region surrounding the 24 nt barcode was PCR-amplified from the resulting cDNA in two rounds. Round 1 PCRs used cDNA template comprising 1/5th of the PCR reaction volume and primers oPB354 and oHP534. Round 1 PCR cycle numbers were adjusted as needed to obtain adequate product concentration while avoiding overamplification (between 5 and 15 cycles), then cleaned using DNA Binding Buffer (Zymo ZD4004-1-L) and UPrep Micro Spin Columns (Genesee Scientific 88-343). Cleaned samples were then used as template for Round 2 PCR, and cycles were again adjusted to avoid overamplification (between 4 and 8 cycles). Round 2 PCRs used Round 1 PCR product comprising between 1/10th to 1/5th of the PCR reaction volume and oAS111 with indexed forward primers (oAS112-135 and oHP281-290). Amplified samples were run on a 2% agarose gel and fragments of the correct size were purified using ADB Agarose Dissolving Buffer (Zymo D4001-1-100) and UPrep Micro Spin Columns (Genesee Scientific 88-343). Concentrations of gel-purified samples were measured using a Qubit dsDNA HS Assay Kit (Q32851) with a Qubit 4 Fluorometer. Samples were sequenced using an Illumina NextSeq 2000 in 1 \times 50, 2 \times 50, or 1 \times 100 mode (depending on other samples pooled with the sequencing library). For the pHPSC1142 libraries, samples were sequenced with standard Read 1, standard Read 2, and standard i7/i5 index sequencing primers. A subset of these libraries were sent for re-sequencing to obtain greater read depth and sequenced with standard Read 1, custom Read 2 oAS1638 (to maintain compatibility with other libraries in the pool), and standard i7/i5 index sequencing primers. For the pHPSC1117 libraries, samples were sequenced with the standard Read 1 sequencing primer and standard index sequencing primers. For the pHPSC1160 libraries, samples were sequenced with standard Read 1, standard Read 2 and standard index sequencing primers.

For the FK₈ library (pHPSC1163/pHPSC1164), between 0.5 and 10 μ g of mRNA and gRNA were reverse transcribed into cDNA using SuperScript IV and a primer annealing to the Illumina R1 primer binding site that contains a 7 nt unique molecular identifier (UMI) (oKC235). A 195 nt region surrounding the 48 nt insert was PCR-amplified from the resulting cDNA in one round using oPN776 and indexed for-

ward primers (oKC230-233, oKC239-242). PCR cycle numbers were adjusted as needed to obtain adequate product concentration while avoiding overamplification (between 10 and 17 cycles). Amplified samples were size-selected and quantified as described previously. Samples were sequenced using an Illumina NextSeq 2000 in 1 \times 70 mode using standard Read 1, custom i7 sequencing primer oKC256, standard i5, and custom Read 2 sequencing primer oKC236.

The 8 \times dicodon library (pHPSC1142) in glucose-depleted cells and the small 8 \times dicodon libraries (pHPSC1165-71) in pooled RQC deletion strains were reverse transcribed following the same procedure and primer as pHPSC1163/pHPSC1164 described above. A 219 nt region surrounding the 48 nt insert and 24 nt barcode was PCR-amplified from the resulting cDNA in one round using oPN776 and indexed forward primers (oKC230-233, oKC239-242). PCR cycle numbers were adjusted as needed to obtain adequate product concentration while avoiding overamplification (between 8 to 16 cycles). Amplified samples were size-selected and quantified as described previously. Samples were sequenced using an Illumina NextSeq 2000 in 1 \times 70 mode using standard Read 1, custom i7 sequencing primer oKC256, standard i5, and custom Read 2 sequencing primer oKC236.

Insert-barcode linkage sequencing

8–10 ng of plasmid pools (pHPSC1142, pHPSC1160, pHPSC1117, pHPSC1165-71) were used in PCR using Phusion polymerase (Thermo Fisher F530S) or Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher F548L). Round 1 PCR was carried out for up to 10 cycles, with 8–10 ng plasmid pool template comprising 1/5th of the PCR reaction volume, using primers oPB354 and oHP534. Round 1 PCRs were cleaned using DNA Binding Buffer (Zymo ZD4004-1-L) and UPrep Micro Spin Columns (Genesee Scientific 88-343). Cleaned samples were used as template for Round 2 PCR, carried out to between 4 to 8 cycles, using oAS111 and indexed forward primers (oAS112-135 and oHP281-290). Amplified samples were purified after size selection and quantified as described above. Samples were sequenced using an Illumina NextSeq 2000 in 2 \times 50 or 1 \times 100 mode. For the pHPSC1142 library, sequencing was performed using standard Read 1 sequencing primer, standard index sequencing primers, and custom Read 2 sequencing primer oAS1637. For the pHPSC1117 library, sequencing was performed using standard Read 1 sequencing primer and standard index sequencing primers. For the pHPSC1160 library, sequencing was performed using standard Read 1, standard Read 2, and standard index sequencing primers. For the pHPSC1165-71 libraries, sequencing was performed using standard Read 1, standard Read 2 and standard index sequencing primers.

Flow cytometry

Five single *S. cerevisiae* colonies integrated with plasmids described above were inoculated into separate wells of 96-well plates containing 150 μ l of SCD-URA medium in each well and grown overnight at 30°C with shaking at 800 rpm. The saturated cultures were diluted 100-fold into 150 μ l of fresh SCD-URA medium and grown for 5–6 h at 30°C with shaking at 800 rpm. The plates were placed on ice and analyzed using the 96-well attachment of a BD FACS Aria or Symphony cytometer. Forward scatter (FSC), side scatter (SSC), YFP flu-

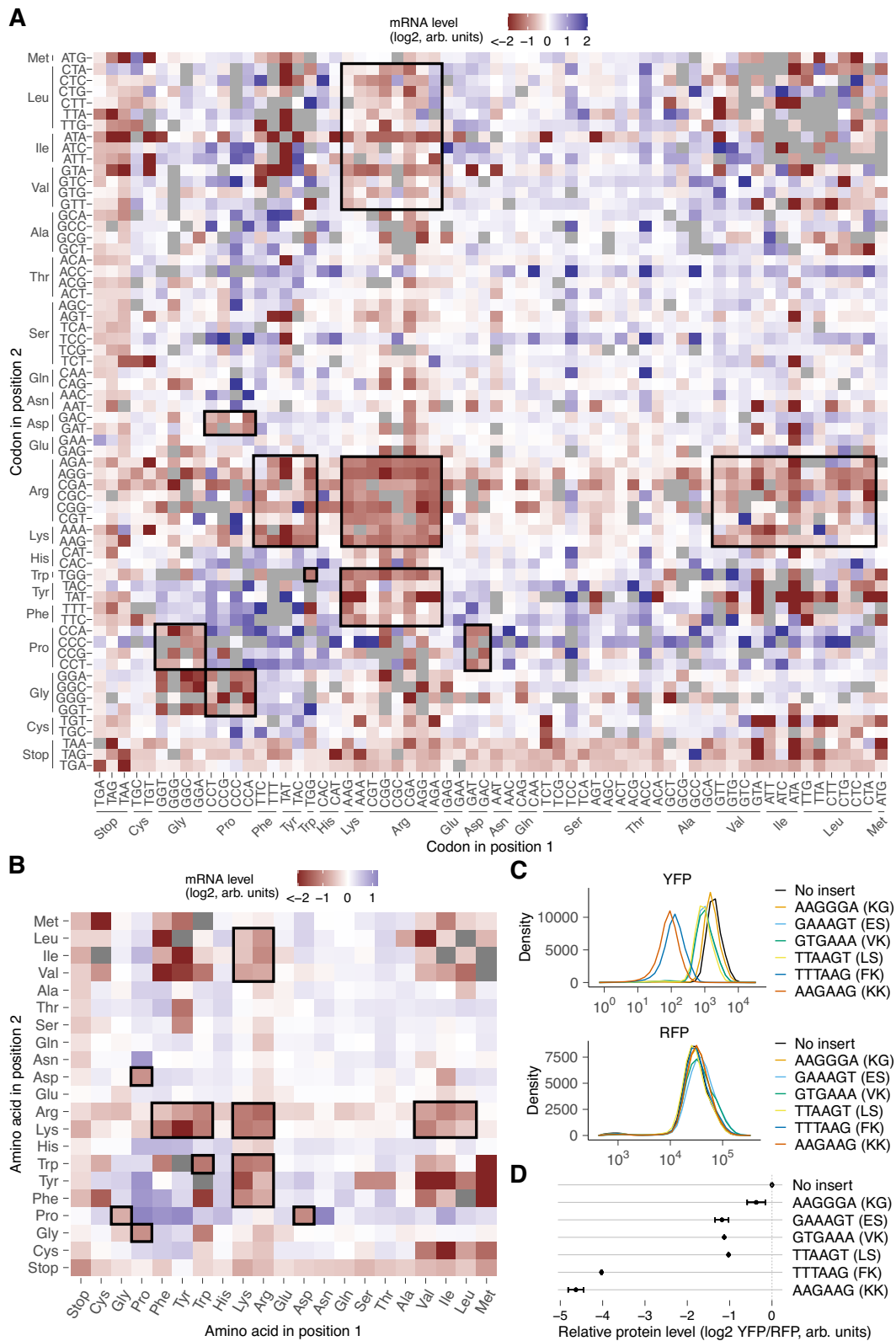


Figure 2. Identification of codon pairs and dipeptides that reduce mRNA levels. **(A)** mRNA level of inserts encoding each codon pair repeat. Codons at the first or second position of each pair are shown along the horizontal or vertical axes, respectively. Missing codon pairs are in grey. Synonymous codon pair families with lower mRNA levels are outlined in black. **(B)** mRNA level of inserts encoding each dipeptide repeat. Amino acids at the first or second position of each dipeptide are shown along the horizontal or vertical axes, respectively. Missing dipeptides are in grey. Dipeptide groups with lower mRNA levels are outlined in black. **(C)** Protein expression from individual *PGK1*-YFP reporters measured by flow cytometry (Top). A control RFP reporter integrated at a different locus was also quantified (Bottom). **(D)** Quantification of median YFP signal in (C) relative to the constitutively expressed RFP reporter. Error bars represent standard error of the mean across 5 biological replicates. GAAAGT (ES) is a frameshift control for GTGAAA (VK), and TTAAGT (LS) is a frameshift control for TTTAAG (FK).

orescence (FITC), and RFP fluorescence (PE.Texas.Red) were measured for 10 000 cells in each well. The resulting data in individual .fcs files for each well were combined into a single tab-delimited text file. YFP expression was first normalized to RFP expression per cell (henceforth referred to as YFP/RFP), then used to calculate the median value of each well. For the no-insert control, the median YFP/RFP values of all wells were averaged together. The median YFP/RFP value per replicate for all strains were then normalized to the average no-insert control value by taking the \log_2 difference. The average and standard error of this ratio across replicates were calculated (Figure 2D).

Computational analyses

Pre-processing steps for high-throughput sequencing were implemented as Snakemake workflows run within Singularity containers on an HPC cluster. All container images used in this study are publicly available as Docker images at <https://github.com/orgs/rasilab/packages>. Python (v3.9.15) and R (v4.2.2) programming languages were used for all analyses unless mentioned otherwise.

Barcode to insert assignment

The raw data from insert-barcode linkage sequencing are in FASTQ format. All pertinent reads were concatenated into one FASTQ file using `fasterq-dump -concatenate-reads`, and inserts and barcodes were extracted and counted using `awk` (mawk implementation, v1.3.4). Only insert-barcode combinations where the insert matches a reference sequence in the list of reference sequences using `awk` were retained. Barcodes were aligned against themselves using `bowtie2` with options `-L 19 -N 1 -all -norc -no-unal -f`. This self-alignment was used to exclude barcodes that are linked to different inserts or that are linked to the same barcode but are aligned against each other by `bowtie2`. In the latter case, the barcode with the lower count is discarded in `filter_barcodes.ipynb`. The final list of insert-barcode pairs is written as a comma-delimited .csv file for aligning barcodes from genomic DNA and mRNA sequencing below.

Barcode counting in genomic DNA and mRNA

The raw data from sequencing barcodes in genomic DNA and mRNA is in FASTQ format. All pertinent reads were concatenated into one FASTQ file, and barcodes were extracted and counted using `awk`. For barcodes that are present in the filtered barcodes .csv file from linkage sequencing, the barcode count and associated insert are printed into a .csv file for subsequent analyses in R. For libraries containing both barcodes and UMIs, only distinct barcode-UMI combinations where the barcode is present in the filtered barcodes .csv file from linkage sequencing are retained. The number of UMIs per barcode and associated insert are printed into a .csv file for subsequent analyses in R.

mRNA quantification and statistical analyses for barcode sequencing

Only barcodes with a minimum of 10 reads and inserts with a minimum of 2–4 barcodes were included. The mRNA level for each insert was calculated as the mean \log_2 ratio of the summed mRNA barcode counts to the summed gRNA barcode counts using 100 bootstrap samples. The standard devi-

ation was calculated across all barcodes for each insert using 100 bootstrap samples. For libraries with a large number of variants (e.g. $\geq 70\,000$) mRNA levels were median-normalized within each library. For libraries with a smaller number of variants (e.g. 1000–2000), libraries were normalized to spike-in strain barcode counts or library size (RPM).

For the small $8 \times$ dicodon libraries in pooled RQC deletion strains, only barcodes with a minimum of 10 reads, inserts with a minimum of 2 barcodes and 100 reads, and dipeptides with a minimum of 500 reads were included. The mRNA level for each insert was calculated as the mean \log_2 ratio of the summed mRNA barcode counts to the summed gRNA barcode counts. The mRNA level for each dipeptide was calculated as the average log fold change across inserts. Because all strains were harvested as a pool, no spike-in strains were used and reads were instead normalized by total reads in the library.

For all other experiments, the standard error of the mean was calculated using the `std.error` function from the `plotrix` R package. P-values for statistically significant differences were calculated using the `t.test` or `wilcox.test` R functions as appropriate for each figure (see figure captions).

Insert counting and mRNA quantification

For the FK_8 deep mutational scanning library, inserts were sequenced directly and thus barcodes were not counted or used for statistical analysis. Instead, inserts and UMIs were extracted and counted using `awk`. Only insert-UMI combinations where the insert matches a reference sequence in the list of reference sequences using `awk` were retained. Subsequent insert-UMI counts were summed across the mRNA and gRNA samples. mRNA levels for each insert were calculated as the \log_2 ratio of the summed mRNA insert-UMI counts to the summed gRNA insert-UMI counts, and then averaged across the two biological replicates. Resultant mRNA levels were then normalized against mRNA levels of spike-in strains to allow comparison between wild-type, *hel2* Δ and *upf1* Δ cells.

Results

A massively parallel reporter assay for mRNA effects in *S. cerevisiae*

To study the effect of coding sequence motifs on mRNA levels in *S. cerevisiae* in an unbiased manner, we modified a pooled reporter assay that we previously developed in mammalian cells (27) (Figure 1A). In our design for *S. cerevisiae*, a tandem $8 \times$ repeat of all possible codon pairs (4096 pairs in total) is inserted between the *PGK1* and *YFP* coding sequences. The $8 \times$ repetition amplifies the effect of each codon pair on mRNA levels. Each codon pair repeat is followed by a 24 nucleotide random barcode without stop codons, which enables their accurate quantification without sequence-dependent biases. Barcode sequences linked to each codon pair insert are identified by sequencing the plasmid library. We integrated the plasmid library into a noncoding region of chromosome I of *S. cerevisiae*, extracted mRNA and genomic DNA, and counted barcodes by high throughput amplicon sequencing. Barcode counts in the cDNA normalized by corresponding counts in the genomic DNA provide a relative measure of the steady-state mRNA level of each codon pair insert in our library (Supplementary Table S4). We further normalized mRNA levels by the median value across all inserts in the library to ac-

count for different sequencing depths and to facilitate comparison across experiments.

We recovered a median of 20 barcodes linked to each codon pair insert in the cDNA and genomic DNA libraries out of the 100 barcodes per insert in the plasmid library (Supplementary Figure S1A). We identified barcodes linked to 97% of all codon pairs in the plasmid library and 91% in the cDNA and genomic DNA libraries (Figure 1B), indicating our assay's ability to capture most of the codon pair motifs. Missing codon pairs in the plasmid library have a high GC content (Supplementary Figure S1B), suggesting that they are either resistant to cloning or toxic for *E. coli* growth. Many of the remaining missing codon pairs in the cDNA and genomic DNA from *S. cerevisiae* encode hydrophobic amino acids (Supplementary Figure S1C). Constitutive expression of such dipeptide repeats might be toxic due to their aggregation or membrane insertion.

To test whether our massively parallel assay recapitulates known codon and amino acid effects, we examined the average mRNA levels of individual codons and amino acids (Figure 1C, E). To this end, we calculated the normalized ratio of barcode counts between cDNA and genomic DNA across all codon pairs containing each of the 64 codons or 20 amino acids. We observed a tight overlap of average mRNA levels of each codon or amino acid between positions 1 and 2 of the codon pair (Figure 1C, E). This observation is consistent with the $8 \times$ repetitive nature of our codon pair library, due to which each codon pair insert is similar to its codon-reversed counterpart except for circular permutation of a single codon.

Within several synonymous codon families, codons with lowest mRNA levels in our assay (Figure 1C) correspond to the less frequent codons within that family in the *S. cerevisiae* transcriptome (52–54). These include CGA, CGG and AGG (Arg), ATA (Ile) and CCG (Pro) (Figure 1C), all of which are known to reduce protein expression or trigger mRNA decay in *S. cerevisiae* (3,5,6,20,55,56). In line with these observations, average mRNA levels of codons in our assay positively correlated with codon stability coefficients (CSCs) inferred from stability measurements on endogenous mRNAs in *S. cerevisiae* (3,4) (Figure 1D, $r = 0.50$, $P < 1e-4$). This correlation with CSC is notable given that we vary only a 16 codon region within a 700 codon *PGK1-YFP* coding sequence in our assay.

At the amino acid level, arginine, lysine and tryptophan had the lowest mRNA levels on average (Figure 1E), consistent with the known role for these amino acids in triggering ribosome-associated quality control (6,7,12,20,28,46,57–60). mRNA effects of these amino acids are comparable to that of stop codons, which trigger nonsense-mediated mRNA decay (NMD). In contrast to the codon effects, average mRNA levels of amino acids in our assay do not show significant correlation with amino acid stability coefficients (AASCs) inferred from stability measurements on endogenous mRNAs in *S. cerevisiae* (4) (Figure 1F). This lack of correlation is in line with the limited role of amino acid identity in determining global mRNA stability in *S. cerevisiae* (3,4). Further, AASC values are derived from endogenous mRNAs where amino acids are typically not clustered as they are in our codon pair reporter assay. Thus, the low correlation of our results with AASC (Figure 1F) suggests that clustering of amino acids rather than the identity of the amino acid itself drives the observed mRNA effects in our assay. Conversely, the higher correlation between our results and CSC (Figure 1D) suggests that the codon ef-

fects in our assay are driven by the identity of the codon itself.

Identification of codon pair repeats that reduce mRNA levels

Inclusion of all possible codon pair repeats in our library allowed us to next study the effect of pairwise codon and amino acid combinations on mRNA levels (Figure 2A, B). We found a strong correlation ($r = 0.92$, $P < 1e-10$) between mRNA effects of codon pairs and their reverse counterparts, indicating the robustness of our measurements (Supplementary Figure S1D). We identified several families of synonymous codon pairs that consistently reduced mRNA levels relative to the remaining inserts in the library (black outlines, Figure 2A, B). Among the most destabilizing codon families were those encoding lysine, arginine, and tryptophan repeats, in agreement with the average destabilizing effect of these amino acids (Figure 1E). Several hydrophobic and aromatic pairs (MLIV \times FYW) as well as glycine pairs had low mRNA levels (Figure 2A). As noted earlier, many of these pairs also had low read counts in genomic DNA, suggesting that their mRNA effects might arise indirectly from cell growth inhibition rather than due to changes in mRNA stability.

Our assay revealed several dipeptide repeats that have not been previously associated with ribosome stalling or ribosome-associated quality control in *S. cerevisiae* (Figure 2A, B). These include several combinations of bulky and positively charged amino acids such as phenylalanine-lysine (FK/KF), tryptophan-arginine (WR/RW), tyrosine-arginine (YR/Ry), and tyrosine-lysine (YK/KY). Interestingly, within the tyrosine-lysine and tyrosine-arginine groups, the strongest mRNA-reducing effects are for the purine-containing codons (AAA, AAG, AGA and AGG), possibly due to nucleotide-specific interactions at the decoding center of the ribosome as has been observed in other stalling sequences (8,40). Some combinations of hydrophobic and positively charged amino acids such as arginine-leucine (LR/RL) and arginine-isoleucine (IR/RI) were also destabilizing. Notably, we found similar mRNA-destabilizing combinations of positively charged amino acids with bulky and hydrophobic amino acids in human cells (27), indicating that these sequences may be broadly destabilizing across eukaryotes. We confirmed the requirement of bulkiness for reducing mRNA levels in a targeted experiment by replacing phenylalanine with the smaller non-polar glycine in combination with lysine (Supplementary Figure S2A). Using flow cytometry, we found FK dipeptide repeats reduced YFP reporter levels similar to the known RQC-inducing KK repeat (Figure 2C, D). Moreover, protein levels of a control RFP reporter expressed from a different chromosomal locus was unaffected by FK repeat expression, indicating that it does not perturb global gene expression (Figure 2C).

Proline-glycine (PG/GP) and proline-aspartic acid (PD/DP) repeats were also among the mRNA-destabilizing codon pairs in our assay (black outlines, Figure 2A, B). Unlike combinations of bulky and positively charged amino acids, these repeats did not reduce mRNA levels in human cells (27). Conversely, amino acid combinations such as arginine-histidine and serine-phenylalanine that destabilize mRNAs in human cells (27) did not reduce mRNA levels in our assay in *S. cerevisiae*. Finally, dipeptides comprised of bulky and positively charged amino acids as well as proline-glycine and

proline-aspartic acid dipeptides are enriched at sites of ribosome collisions in *S. cerevisiae* and mammalian cells (36–38). This observation suggests that the mRNA-destabilizing effects of such dipeptide repeats in our assay arises from ribosome slowdown when these peptide motifs are synthesized during mRNA translation.

Dipeptide-induced mRNA destabilization requires translation

We used three different approaches to assay whether translation of dipeptide repeats is necessary for their mRNA-destabilizing effects.

First, we computationally tested whether the presence of codon pairs in the correct *PGK1-YFP* reading frame is necessary for the mRNA effects of the corresponding dipeptide repeats (Figure 3A). We compared the mRNA level of the insert with a codon pair in the +0 frame against the mRNA levels of the insert where the sequence is shifted over by one, two or three nucleotides (+1, +2, +3 frame respectively). mRNA effects of dipeptide repeats encoded in the correct +0 frame showed much lower correlation with the mRNA effects in the wrong +1 and +2 frames than with the correct +3 frame. We note that the +3 frameshift is essentially the same frame as the in-frame codon pair but with the codon positions interchanged. Thus, the simple presence of nucleotide sequences coding for destabilizing dipeptide repeats in the mRNA is not sufficient to reduce mRNA levels; they need to be present in the correct translated frame. Consistent with this observation, we found low correlation between mRNA levels of codon pair inserts and basic measures of nucleotide diversity such as GC content or GC3 content (Supplementary Figure S2B).

Second, we tested whether global inhibition of translation is sufficient to rescue the mRNA-destabilizing effects of dipeptide repeats. Glucose deprivation is known to rapidly inhibit translation initiation in yeast (61,62). Therefore, we grew *S. cerevisiae* cells containing the original codon pair library (Figure 1A) in media without glucose for one hour, and quantified relative mRNA levels of inserts by high throughput sequencing as before. At the codon level, glucose deprivation increased the relative mRNA levels of inserts containing arginine and lysine codons, consistent with their mRNA effects arising at the translational level (Figure 3B). Glucose deprivation also increased the relative mRNA levels of several dipeptide-encoding inserts that were destabilizing under normal growth (Figure 3C). These include the known RQC-inducing polybasic sequences RR, RK, KR, and KK, as well as the novel destabilizing dipeptide repeats such as KW, FK, RW, PD and PG that we identified in our original screen. Intriguingly, stop codon-containing inserts had lower mRNA levels during glucose deprivation even though nonsense-mediated mRNA decay of these inserts also requires translation. This might be because NMD is triggered following just one or few rounds of translation (63), while a high rate of translation initiation is necessary for collision-driven mRNA decay (20). Thus, the inhibition of translation during glucose deprivation will have a greater effect on the rate of collision-dependent decay than on NMD, leading to relatively lower mRNA levels of stop codon-containing inserts during glucose deprivation (Supplementary Figure S4B).

Third, we tested whether experimentally altering the translated reading frame of codon pair inserts is sufficient to abrogate their mRNA-destabilizing effects, which would rule

out transcription or RNA processing as possible mechanisms. Therefore, we inserted 2 base pairs upstream of the codon pair insert, leaving all other aspects of the reporter identical to the original library, and assayed for mRNA effects as before (Figure 3D). The 2 bp insertion shifts all codon pair inserts to the –1 frame, but does not introduce a stop codon upstream of the codon pair inserts. At the aggregate level, the –1 frameshifted library loses the previously observed correlation with codon stability coefficients (Figure 3E, compare against Figure 1D), consistent with the codon effects predominantly arising from translation. Similarly, most dipeptide repeats that destabilize mRNAs in the original library had higher relative mRNA levels in the –1 frameshifted library (Figure 3F). Note that the WW dipeptide-coding repeat did not pass our read cutoff filter in both the glucose deprivation and the –1 frameshifting experiment (Figure 3C, F).

Overall, for the set of 19 dipeptides that we reliably identified as destabilizing in our assay, we found that 15 of them showed rescued mRNA levels with all three methods (Supplementary Figure S3A). Thus, our computational and experimental frameshifting assays, along with our glucose depletion experiment, establish the translation dependence of the mRNA effects of most of the destabilizing dipeptide repeats identified in our original screen.

Ribosome-associated quality control regulates mRNA destabilization by dipeptide motifs

Given the translational dependence of mRNA destabilization by dipeptide repeats, we sought to identify the co-translational regulatory pathways mediating these effects. Ribosome stalling at poly-lysine, poly-arginine, and poly-tryptophan repeats triggers ribosome-associated quality control (RQC) of nascent peptides and mRNAs in *S. cerevisiae* (6,28,58–60). The E3 ubiquitin ligase Hel2 (*S. cerevisiae* homolog of human ZNF598), which binds collided ribosomes at extended ribosome stalls, is necessary for RQC induction at these sequences (10,32,59,60,64–66) (Figure 4A). Syh1 (GI-GYF2 in humans) has also been recently implicated in a Hel2-independent pathway of mRNA decay of reporters with repeats of the rare codon CGA (67–69) (Figure 4A). To test the requirement for these factors in reducing the mRNA levels at the novel destabilizing dipeptide repeats identified in our screen, we integrated our original 4096-codon pair library into *S. cerevisiae* strains with *HEL2* or *SYH1* deletion, and measured relative mRNA levels as before (Figure 4B).

We compared by linear regression the relative mRNA levels in the *hel2*Δ and *syh1*Δ strains against the wild-type strain to identify inserts with altered mRNA levels (Figure 4C, D). In the *hel2*Δ strain, 14 dipeptides had 1.5-fold or greater increase in relative mRNA levels compared to the wild-type strain (red points, Figure 4C). These include the known RQC-inducing repeats, KK, RR, WW, RK and KR. *HEL2* deletion also restored the mRNA levels of several bulky and positively charged dipeptide repeats (FK/KF, WR/RW, WK/KW) as well as proline-aspartic acid (PD/DP) and proline-glycine (PG/GP) repeats (Figure 4E, Supplementary Figure S3C). By contrast, *SYH1* deletion did not restore the mRNA levels of any dipeptide repeat (Figure 4D, E). This is likely because Syh1 acts as a compensatory mechanism when Hel2-mediated RQC is inactive (67).

To further assess the involvement of other RNA decay factors in mediating the effects of these pairs, we inte-

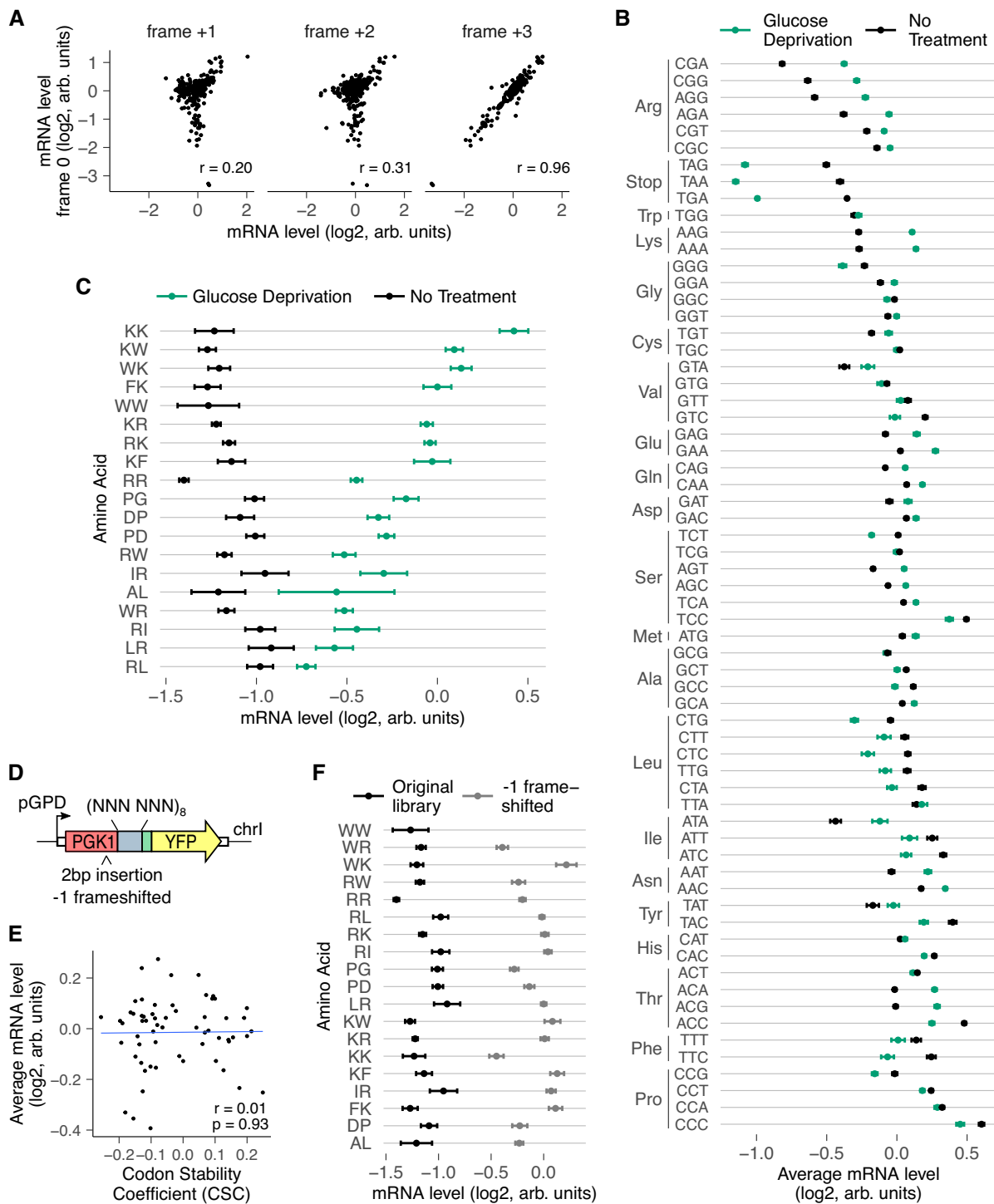


Figure 3. mRNA effects of dipeptide repeats require in-frame translation. **(A)** mRNA level of reporters encoding 320 different dipeptide repeats (excluding stop codon-containing dipeptides and pairs that did not pass read count cutoffs) compared between the correct reading frame (frame 0, vertical axis) and computationally-shifted +1, +2 or +3 reading frames (horizontal axes). r indicates Pearson correlation coefficient. **(B)** Average mRNA level of reporters with indicated codons averaged across positions 1 and 2 of the codon pair library during normal growth and glucose depletion. mRNA levels were median-normalized separately for each growth condition. Error bars represent standard deviation over all variants containing the codon at either position. **(C)** mRNA level of reporters encoding indicated dipeptides during normal growth and glucose deprivation. mRNA levels were median-normalized separately for each growth condition. Only dipeptide inserts with a minimum of 10 reads per barcode, 4 barcodes per insert, and low variability between barcodes are included here and in further analysis. Error bars represent standard deviation over barcodes linked to the indicated dipeptide repeat. **(D)** Schematic of frameshifted codon pair library. Two base pairs were inserted upstream of the codon pair insert in the 4096 codon pair library to create a -1 frameshift in the codon pair. Libraries were integrated and sequenced as in Figure 1A. **(E)** Average mRNA effects of individual codons in the -1 frameshifted library compared against codon stability coefficients (3). **(F)** mRNA levels of destabilizing dipeptides in the original in-frame library and in the -1 frameshifted library. Error bars calculated as in (C).

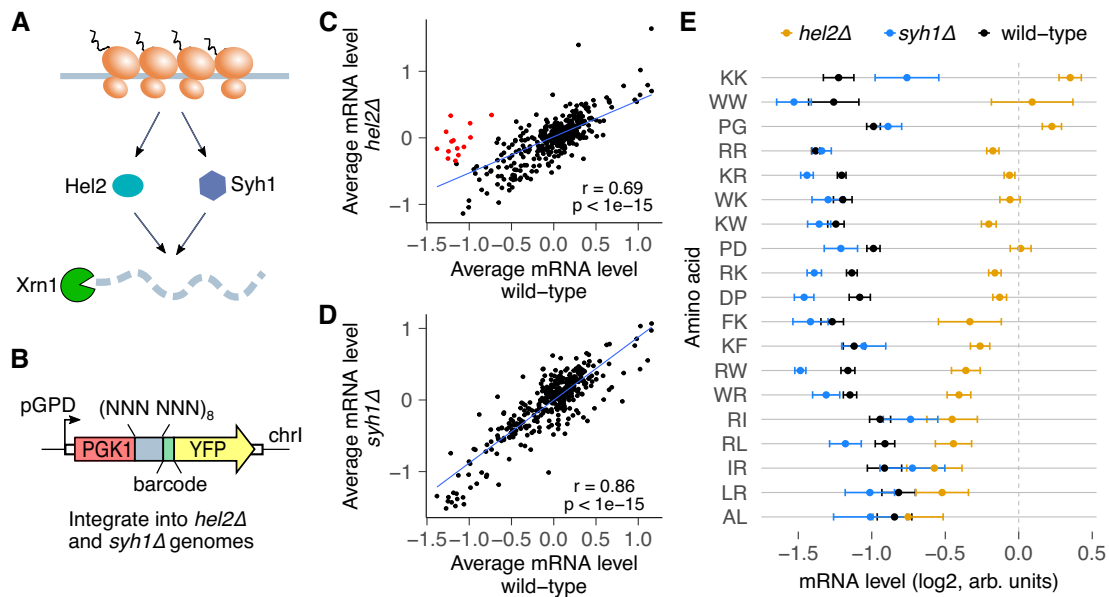


Figure 4. Ribosome collision sensor Hel2 regulates the mRNA effects of dipeptide repeats. **(A)** The RQC factors Hel2 and Syh1 are known to respond to collided ribosomes and trigger mRNA decay through Xrn1. **(B)** The codon pair library in Figure 1A was integrated into *hel2Δ* and *syh1Δ* cells, and mRNA levels were quantified as before. **(C)** mRNA levels for dipeptide repeats compared between *hel2Δ* and wild-type cells. mRNA levels were calculated as in (C), and median-normalized separately for each strain. Dipeptide repeats with residuals less than -2 from the linear regression line are marked in red. **(D)** Same plot as in C, but for *syh1Δ* cells. No dipeptide repeats are preferentially stabilized in *syh1Δ* cells with residuals less than -2 from the linear regression line. **(E)** mRNA levels for wild-type mRNA-destabilizing dipeptides (from C) for *hel2Δ* and *syh1Δ* cells. Error bars represent standard deviation over barcodes linked to the indicated dipeptide repeat.

grated a small pool of select codon pairs into *hel2Δ*, *syh1Δ*, *hel2Δsyh1Δ*, *xrn1Δ* and *cue2Δ* cells. Codon pairs in this small pool included many of the pairs that reduced mRNA levels in our initial screen along with negative control pairs. Consistent with the results of the full library, mRNA levels for none of the destabilized dipeptides were rescued in *syh1Δ* cells, while they were rescued in *hel2Δ* cells (Supplementary Figure S4A). mRNA levels for destabilized dipeptides in the *hel2Δsyh1Δ* double deletion strain were higher than in any of the single deletion strains, consistent with the known mutually compensatory roles of these two factors. *cue2Δ* and *xrn1Δ* cells also had varying but generally higher mRNA levels for destabilizing reporters than wild-type cells, consistent with previous work (33). mRNA destabilization by RI/IR pairs were not rescued by either *HEL2* or *SYH1* deletion in the full library, and were also not rescued by any of the other decay factors tested in the small library.

Together, these results reveal that Hel2-mediated RQC regulates most but not all mRNA-destabilizing effects of dipeptide repeats identified in our original screen, and suggest that Syh1 and Hel2 can act compensatorily to regulate these dipeptides.

Deep mutational scanning identifies critical residues mediating mRNA destabilization by dipeptide motifs

Ribosome-associated quality control often depends on interactions between specific residues in the nascent peptide and various regions of the ribosome such as the peptidyl-transferase center (PTC) and the uL4/uL22 constriction point in the exit tunnel (8,28,40,43,44). To dissect the mechanism by which the FK dipeptide repeat triggers mRNA destabilization, we developed a deep mutational scanning assay using re-

porter mRNA level as a readout (Figure 5A). Specifically, we mutated each location in the 16-codon insert encoding (FK)₈ to all 64 codons to generate a pooled library of 1024 variants. We cloned these variants between the *PGK1* and *YFP* coding sequences, integrated them into the genomes of wild-type and *hel2Δ* cells, and measured variant frequency in cDNA and genomic DNA by high throughput amplicon sequencing. We used the ratio of cDNA to genomic DNA to quantify the relative mRNA levels of each variant, and further normalized to spike-in control strains to enable comparison across different genotypes (see Materials and methods). We confirmed reproducibility of mRNA levels between biological replicate transformations into *S. cerevisiae* of the same plasmid library (Figure 5B).

Visualizing the relative mRNA levels of (FK)₈ mutants as a function of mutation identity and location yields several interesting observations (Figure 5C). First, mRNA levels of stop codon-containing variants decrease only when stop codons are present until position 11 of the (FK)₈ repeat. This suggests that nonsense-mediated mRNA decay does not occur on mRNAs undergoing extended ribosome stalling. Indeed, low coverage deep mutational scanning confirmed that deletion of *UPF1*, the primary effector of NMD, does not rescue reporter mRNA levels when stop codons occur beyond position 11 of the (FK)₈ repeat (Supplementary Figure S5A). Since *hel2Δ* cells also exhibit the same pattern of stop codon effects as wild-type cells (Supplementary Figure S5B), the observed lack of NMD is not simply due to kinetic competition between RQC and NMD, but rather a consequence of ribosome stalling driven by the first 10 codons of the (FK)₈-encoding region.

Second, mRNA levels for nearly all mutations from positions 1 to 6 are as low as the wild-type sequence. This observation is consistent with 10 residues in (FK)₈ being the

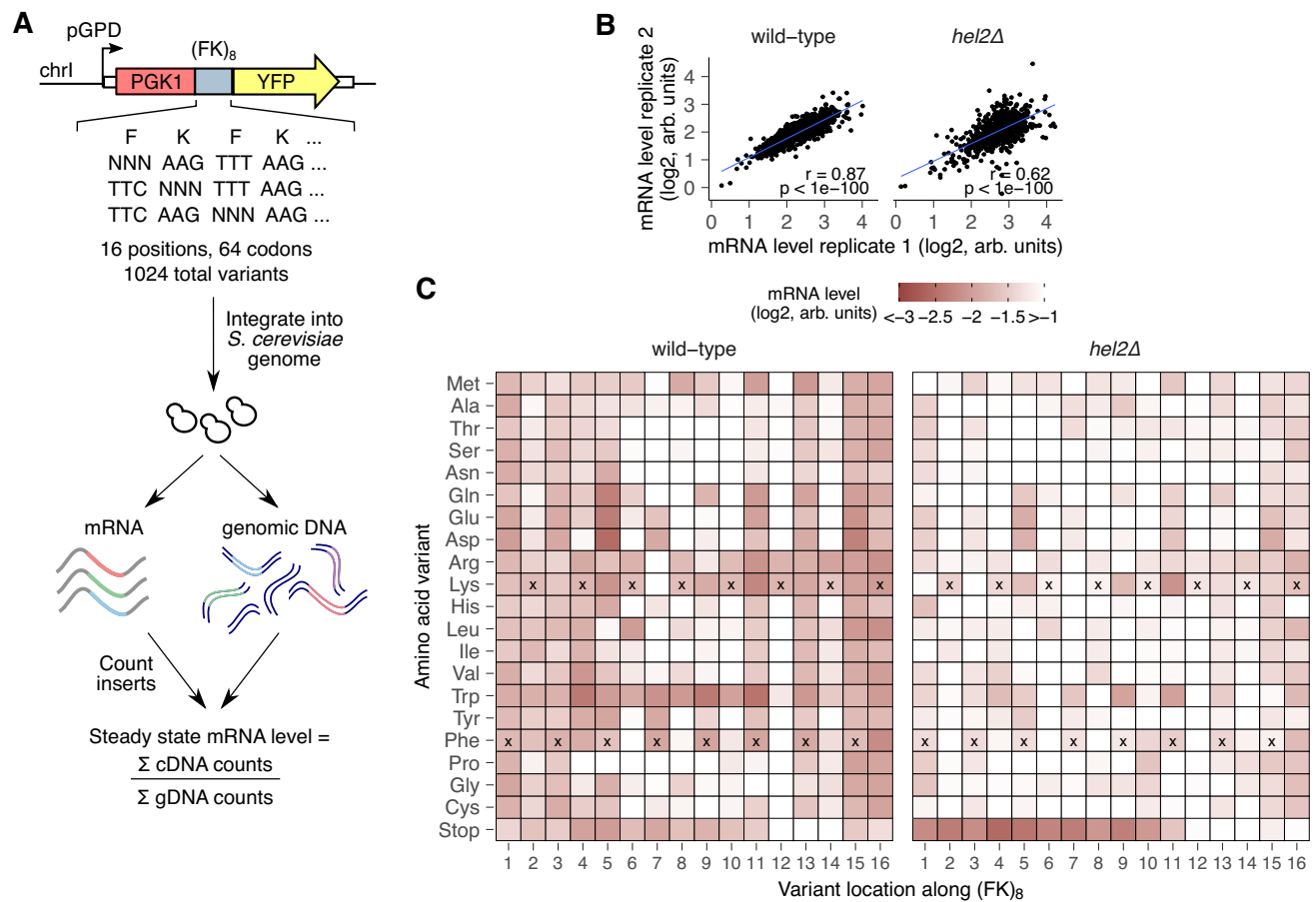


Figure 5. Deep mutational scanning identifies amino acids critical for mRNA effects of a destabilizing dipeptide repeat. **(A)** Schematic of deep mutational scanning (DMS) of the FK dipeptide repeat. Each location in an (FK)₈-encoding insert was randomized to all 64 codons. This 1024-variant library was cloned as a pool between *PGK1* and *YFP*, and genomically integrated into wild-type and *hel2Δ* strains. Inserts were quantified in cDNA and genomic DNA by high throughput amplicon sequencing. **(B)** Pearson correlation between biological replicates for each variant in the (FK)₈ DMS library. **(C)** mRNA level for inserts containing the indicated amino acid mutation (vertical axis) at the indicated position (horizontal axis). mRNA levels are averaged across replicates and normalized within each genotype using spike-in control strains. The wild-type amino acid variant is marked with black crosses at each location.

minimum RQC-inducing length, because mutating any of the first six residues will preserve this minimum length downstream of the mutated position. Pro is the only target mutation within the first six positions that consistently rescues mRNA levels, likely by limiting the conformational flexibility of the nascent peptide (70–72). Third, location 12 (and to a lesser extent location 14) within (FK)₈ are the sole positions that require positively charged Arg or Lys to trigger Hel2-dependent RQC. At several other locations where the original amino acid is positively charged (such as at positions 6, 8 and 10), mutation to the bulkiest Trp residue can still trigger RQC, while mutations to other aromatic amino acids (Phe and Tyr) are insufficient. Fourth, at some locations where the original amino acid is bulky (such as at positions 9 and 11), mutating to the bulkier Trp or to positively charged Arg or Lys maintains RQC. The two preceding observations imply that positive charge and bulkiness play interchangeable roles at several locations within the (FK)₈ repeat in triggering RQC. Finally, at position 7, where the original amino acid is Phe, mutations to other aromatic amino acids (Trp or Tyr) or to a negatively charged residue (Glu or Asp) triggers RQC, while positive charge is insufficient. Thus, the interchangeability of

bulkiness with positive charge in triggering RQC is not universal, but rather depends on the location within the stalling peptide.

We next compared the aggregate effect of all mutations at each location of the (FK)₈ repeat on mRNA levels between wild-type and *hel2Δ* cells (Supplementary Figure S5C). We excluded stop-codon containing mutants from this analysis to avoid convoluting NMD and RQC effects. The positions with the highest mutational effect differences between the two strains are at the ends of the stalling sequence: positions 1–6, 15 and 16 of (FK)₈. This observation is consistent with our earlier interpretation that translation of approximately 10 residues of (FK)₈ is necessary to drive Hel2-dependent mRNA decay. Conversely, positions 10, 9, and 12 had the least mutational effect differences between the two strains, revealing that these positions are most important for triggering Hel2-dependent RQC. Finally, *HEL2* deletion did not fully rescue the mRNA effects of any (FK)₈ terminal mutants (positions 1, 15, 16), suggesting that Hel2-dependent RQC activity is saturated at longer repeat lengths, and mRNA decay proceeds through multiple compensatory pathways.

Codon pair library predicts mRNA effects of endogenous sequences

Though a few mRNA sequences are known to stall ribosomes and trigger RQC in reporter studies (40,41,73), the sequence motifs that underpin endogenous mRNA stability are not well understood. For example, the simple presence of polybasic stretches or rare codons is not sufficient to trigger quality control on endogenous yeast mRNAs (40,74). Thus, we sought to test whether our codon pair assay could predict mRNA effects of sequence motifs in endogenous *S. cerevisiae* genes. To this end, we assayed 1904 fragments, each 48 nucleotides long, from endogenous mRNAs spanning a wide range of expression levels (49) using the same reporter design as the codon pair library (Figure 6A). We integrated this endogenous fragment library into wild-type cells and counted barcodes by high throughput amplicon sequencing as before. Compared to the codon pair library, mRNA levels in the endogenous fragment library were more tightly distributed around the median, indicating more muted effects on mRNA stability (Figure 6B). We next calculated the codon stability coefficient (CSC) values for each of the 64 codons using mRNA levels either from the codon pair library or the endogenous fragment library (3). We found strong correlation ($r = 0.67$, $P < 1e-8$) between the two libraries, indicating that mRNA effects of codon pair repeats predict mRNA effects of endogenous sequence motifs in wild-type cells (Figure 6C). We next integrated the endogenous fragment library into *hel2* Δ cells and tested how Hel2-dependent RQC affects the relationship between CSC values calculated from the codon pair and the endogenous fragment libraries. We found that *hel2* Δ cells still exhibited a significant correlation ($r = 0.49$, $P < 1e-4$) between the two libraries, though to a lesser extent than in wild-type cells (Figure 6D). This is consistent with the differences in mRNA stability of endogenous sequences arising from the additive effects of codons rather than from RQC occurring at specific amino acid sequences.

Discussion

Here, we use a massively parallel approach to identify and dissect sequence motifs underlying mRNA instability in *S. cerevisiae*. In addition to validating known codon and amino acid effects on mRNA stability, we identify several sequence motifs that have not been previously associated with mRNA decay (Supplementary Table S4). These include combinations of bulky and positively charged amino acids, and proline with aspartate and glycine, all of which trigger translation-dependent mRNA decay through the Hel2-dependent RQC pathway. By combining our massively parallel assay with deep mutational scanning, we dissect the codon-level biochemical requirements for triggering mRNA decay by a bulky and positively charged dipeptide repeat. Despite the apparent simplicity of the codon pair repeat library, we find that it captures the mRNA effects of endogenous coding sequence fragments from the *S. cerevisiae* transcriptome.

Our codon pair library confirms the role of codon optimality as a major determinant of mRNA stability in *S. cerevisiae*, and provides insights into the resulting hierarchy of effects. We observe several synonymous codon families within which aggregate mRNA levels differ based on the hierarchy of codon optimality (3,55) (Figure 1C), but have different absolute effects. The non-optimal codons ATA (Ile), GTA (Val),

and TAT (Tyr) are highly destabilized relative to their optimal counterparts. By contrast, the optimal codon TCC (Ser) is preferentially stabilized relative to its non-optimal counterparts. Both the arginine and proline synonymous codon families are stratified based on codon optimality even though these amino acids have opposite average effects on mRNA stability (Figure 1E, Arg – destabilizing, Pro – stabilizing). Thus, codon optimality effects on mRNA stability act in parallel and independent of amino acid identity. Consistent with codon optimality-mediated mRNA decay being a co-translational process (29,75,76), translational shutoff by glucose depletion rescues the mRNA-destabilizing effects of eight out of the 10 most non-optimal codons (ATA, CGA, AGG, GTA, ACG, AGT, AAA, AGC) (3) (Figure 3B). Finally, the effects of codon optimality on mRNA stability in our codon pair library are driven by mutations within a short 16 codon region despite being part of a 700 codon *PGK1-YFP* mRNA. This is likely because the *PGK1-YFP* region is efficiently translated (77), while the tandem and repetitive nature of the codon pairs amplifies their effect on ribosome slowdown and recruitment of mRNA-destabilizing factors.

While polybasic and poly-tryptophan sequences are known to trigger RQC in *S. cerevisiae*, our codon pair assay reveals combinations of bulky (Val, Ile, Leu, Phe, Tyr, Trp) and positively charged (Arg, Lys) amino acids as a general trigger of mRNA decay (Figure 2A, B). Interestingly, combinations of Val, Ile, Leu and Phe with Arg and Lys were also found to destabilize mRNA in human cells (27), indicating their evolutionary conservation as mRNA-destabilizing sequences across eukaryotes. Supporting these findings, ribosome profiling in human cells revealed an enrichment in disome occupancy at sites that followed an Arg-X-Lys pattern, with highest disome density occurring when X was Phe, Ile or Leu (36). We find that positively charged amino acids in combination with the bulkiest side chains (Phe, Trp) trigger RQC-dependent mRNA decay in *S. cerevisiae*, while less bulky side chains (Val, Ile, Leu) decrease mRNA levels in a RQC-independent manner (Figure 4E, A). Gamble *et al.* found that several codon pairs for Arg with Ile and Leu are likely destabilized due to codon rarity and the requirement of some of these codons to be wobble decoded (56). Of the 17 codon pairs that inhibited translation in Gamble *et al.*, we were able to reliably measure mRNA levels of 13 codon pairs, of which 12 are in the lowest quartile of mRNA levels in our assay (Supplementary Figure S6A). This is consistent with ribosome stalling at these codon pairs triggering mRNA decay when the codon pairs occur in a cluster (78).

In our codon pair assay, combinations of proline with aspartate and glycine (PD/DP, PG/GP) decrease mRNA levels in a Hel2-dependent manner (Figures 2A, B, 4E, Supplementary Figure S3). While poly-proline sequences stall ribosomes due to inefficient peptide bond formation, these sequences are not known to induce RQC and are instead translated with the assistance of eIF5A (34). Consistent with these previous findings, proline-proline combinations, and all other proline-containing combinations except for with aspartate and glycine, are stabilizing in our assay. Conversely, no other aspartate or glycine containing codon pairs except the ones with proline are destabilizing. While increased ribosome occupancy has been observed at proline, aspartate, and glycine codons in both *S. cerevisiae* and human cells (36,79,80), our results suggest that these effects may be driven by combinations of these amino acids rather than by their individual oc-

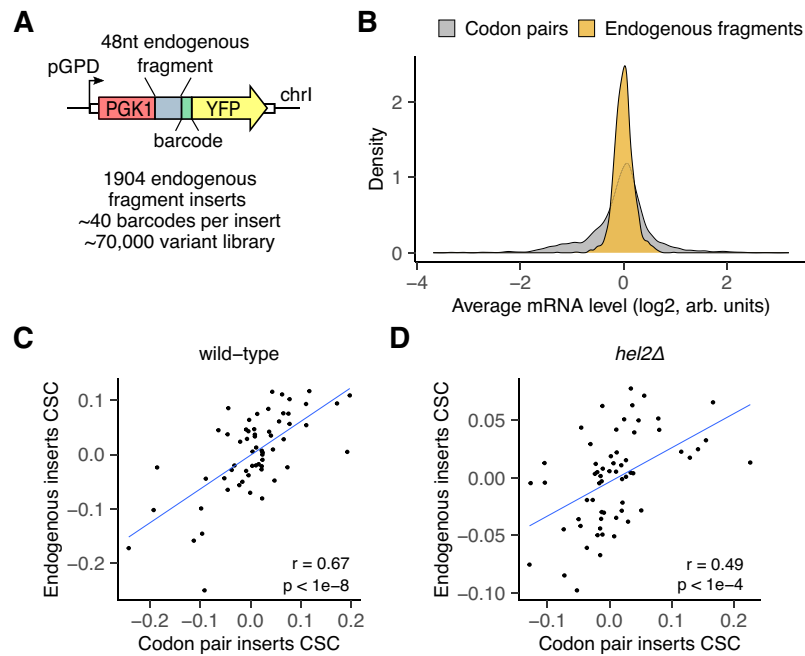


Figure 6. Codon pair measurements predict effects of endogenous mRNA sequences. **(A)** Schematic of endogenous sequence insert library. Each element in the library includes one of 1904 possible 48nt endogenous fragments. Each sequence is inserted in-frame between *PGK1* and *YFP*, and is followed by a random 24nt barcode without in-frame stop codons (median of 40 barcodes/insert). The 70 000 variant library is genomically integrated into wild-type and *hel2Δ* cells, and mRNA levels are quantified as in Figure 1A. **(B)** Distribution of mRNA levels for endogenous fragments vs codon pair inserts in wild-type cells. **(C)** Correlation between CSC values calculated for each codon from the endogenous fragment library against CSC values derived from the codon pair library in wild-type cells. Pearson correlation coefficient is reported as r . The CSC for each codon is calculated by taking the Pearson correlation coefficient between codon frequency of an insert and its steady state mRNA level. **(D)** Same plot as in (C), but for *hel2Δ* cells.

currence. Consistent with this idea, PD and PPD peptides have increased ribosome occupancy and are under-represented in the *S. cerevisiae* proteome, while PP and GG dipeptides also have increased ribosome occupancy but are over-represented (37). Similarly, PD dipeptides in *E. coli* (81), and PD and PG motifs in mouse embryonic stem cells (38) have increased ribosome occupancy. Thus, PD and PG motifs may have evolutionarily conserved effects on ribosome slowdown through a mechanism distinct from poly-proline stalls, and can trigger Hel2-dependent mRNA decay in *S. cerevisiae*.

Our deep mutational scanning reveals complex codon-level requirements for the (FK)₈ repeat to confer mRNA instability in a Hel2-dependent manner (Figure 5). Strikingly, these results also exhibit several similarities to the composite biochemical requirements for ribosome stalling observed at the known endogenous RQC substrate in *S. cerevisiae*, *SDD1*₁₉₆₋₂₁₂ (FFYEDYLIFDCRAKRRK) (40). First, the strict requirement for positive charge at positions 12 and 14 of the (FK)₈ repeat to trigger mRNA decay matches the requirement for positive charge at positions 207 and 209 of *SDD1*₁₉₆₋₂₁₂, which are thought to perturb the peptidyl-transferase center of the ribosome. Second, the requirement for bulky aromatic residues at position 7 of (FK)₈ is similar to the requirement for aromatic residues at position 201 of *SDD1*₁₉₆₋₂₁₂, which are thought to interact with the uL4/uL22 constriction point of the ribosome. Third, the ability of negatively charged aspartate, and to a lesser extent glutamate, at position 7 of (FK)₈ to preserve stalling resembles the requirement for aspartate at position 200 of *SDD1*₁₉₆₋₂₁₂, though in the *SDD1* case, the requirement for aspartate is strict. Our results show that bulkiness can be compensated by negative or positive charge in stall sequences depending on the position along the sequence.

Specifically, aspartate's prevalence in stalling sequences is evident in ribosome profiling studies from *S. cerevisiae* to humans, which show increases in monosome and disome occupancy at aspartate codons (36,79,80), presumably due to interactions with the negatively charged ribosome exit tunnel. Taken together, our deep mutational scanning results with a simple (FK)₈ repeat recapitulate and generalize the biochemical requirements for ribosome stalling and quality control observed with other endogenous stall sequences. Nevertheless, our mutational scanning results with the single (FK)₈ repeat are insufficient to decipher the mRNA-destabilizing effects of other dipeptides such as (PG)₈ and (PD)₈ identified in our assay or endogenous stalling sequences with critical role for proline rather than positively charged amino acids (82–84).

While we did not intend to focus on NMD for this study, our assay nonetheless identified several patterns related to NMD. Surprisingly, we found that glucose depletion selectively destabilized stop codon-containing mRNAs for all three stop codons (Figure 3B) even though NMD depends on mRNA translation. A likely explanation is that NMD, which requires only a few rounds of translation, experiences a less drastic inhibition during glucose depletion than collision-driven mRNA decay, a decay process requiring high translation initiation rate. Deep mutational scanning of the (FK)₈ dipeptide also revealed the differential kinetics between NMD and RQC when in competition for the same substrates (Figure 5C). Before 10 Phe and Lys residues are translated, stop-codon containing sequences are predominantly degraded by NMD. After this minimum stalling sequence is translated, RQC dominates as the primary regulatory mechanism. A minimum length of 10 Phe and Lys residues of RQC is consistent with 12 repeated tryptophan residues being sufficient

to induce RQC, while greater than 8 residues were required (28).

The results of our combinatorial codon pair and endogenous motif mRNA stability assays suggest that a wider diversity of mRNA sequences impact mRNA stability than previously appreciated. Poly-GP repeats, identified in our study to stall ribosomes and trigger RQC, are translated through repeat associated non-ATG (RAN) translation of the pathogenic G₄C₂ repeat expansion in the *C9ORF72* gene and is a biomarker for C9ORF72-associated ALS (85). Valine-arginine repeats, identified in our study to destabilize mRNAs in a Hel2-independent manner, are also translated through RAN in the mammalian TERRA sequence to form inclusions during disrupted telomere homeostasis (86). Thus the sequences identified in our study have important implications in the maintenance of cellular homeostasis and disease progression.

Data availability

The raw sequencing data generated in this study have been deposited in the Sequence Read Archive under BioProject accession number PRJNA974090, at <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA974090>. Raw data from flow cytometry are available at <http://flowrepository.org/id/FR-FCM-Z6QH>. Code to reproduce figures in the manuscript starting from raw data is publicly available at <https://doi.org/10.5281/zenodo.8365102> and https://github.com/rasilab/chen_2023. Software environments used to run the code in the above GitHub repository are publicly available as Docker containers at <https://github.com/orgs/rasilab/packages>. Biological reagents or methodology clarification can be publicly requested by opening an issue at https://github.com/rasilab/chen_2023/issues.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank members of the Subramaniam lab, the Zid lab and Joshua Arribera for discussions and feedback on the manuscript. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions: K.Y.C. conceived the project, designed research, performed experiments, analyzed data, and wrote the manuscript. H.P. designed research and performed experiments. A.R.S. conceived the project, designed research, analyzed data, wrote the manuscript, supervised the project, and acquired funding.

Funding

National Institutes of Health [R35 GM119835]; National Science Foundation [MCB 1846521 to A.R.S.]; Genomics Shared Resources of the Fred Hutch/University of Washington Cancer Consortium [P30 CA015704]; Scientific Computing at Fred Hutchinson Cancer Center [NIH grants S10-OD-020069 and S10-OD-028685]. Funding for open access charge: NIH [GM119835].

Conflict of interest statement

None declared.

References

1. Hoof, A. van and Wagner, E.J. (2011) A brief survey of mRNA surveillance. *Trends Biochem. Sci.*, **36**, 585–592.
2. Shoemaker, C.J. and Green, R. (2012) Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.*, **19**, 594–601.
3. 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.
4. Forrest, M.E., Pinkard, O., Martin, S., Sweet, T.J., Hanson, G. and Collier, J. (2020) Codon and amino acid content are associated with mRNA stability in mammalian cells. *PLoS One*, **15**, e0228730.
5. Letzring, D.P., Dean, K.M. and Grayhack, E.J. (2010) Control of translation efficiency in yeast by codon–anticodon interactions. *RNA*, **16**, 2516–2528.
6. Letzring, D.P., Wolf, A.S., Brule, C.E. and Grayhack, E.J. (2013) Translation of CGA codon repeats in yeast involves quality control components and ribosomal protein L1. *RNA*, **19**, 1208–1217.
7. Arthur, L.L. and Djuranovic, S. (2018) PolyA tracks, polybasic peptides, poly-translational hurdles. *Wiley Interdiscip. Rev. RNA*, **9**, e1486.
8. Chandrasekaran, V., Juskiewicz, S., Choi, J., Puglisi, J.D., Brown, A., Shao, S., Ramakrishnan, V. and Hegde, R.S. (2019) Mechanism of ribosome stalling during translation of a poly(A) tail. *Nat. Struct. Mol. Biol.*, **26**, 1132–1140.
9. Doma, M.K. and Parker, R. (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, **440**, 561–564.
10. Simms, C.L., Yan, L.L. and Zaher, H.S. (2017) Ribosome collision is critical for quality control during No-go decay. *Mol. Cell*, **68**, 361–373.
11. Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrero, A.L., Parker, R. and Dietz, H.C. (2002) An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science*, **295**, 2258–2261.
12. Ito-Harashima, S., Kuroha, K., Tatematsu, T. and Inada, T. (2007) Translation of the poly(A) tail plays crucial roles in nonstop mRNA surveillance via translation repression and protein destabilization by proteasome in yeast. *Genes Dev.*, **21**, 519–524.
13. Tsuboi, T., Kuroha, K., Kudo, K., Makino, S., Inoue, E., Kashima, I. and Inada, T. (2012) Dom34:hbs1 Plays a general role in quality-control systems by dissociation of a stalled ribosome at the 3' end of aberrant mRNA. *Mol. Cell*, **46**, 518–529.
14. Tesina, P., Lessen, L.N., Buschauer, R., Cheng, J., Wu, C.C.-C., Berninghausen, O., Buskirk, A.R., Becker, T., Beckmann, R. and Green, R. (2020) Molecular mechanism of translational stalling by inhibitory codon combinations and poly(A) tracts. *EMBO J.*, **39**, e103365.
15. Lu, J. and Deutsch, C. (2008) Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.*, **384**, 73–86.
16. Meydan, S. and Guydosh, N.R. (2021) A cellular handbook for collided ribosomes: surveillance pathways and collision types. *Curr. Genet.*, **67**, 19–26.
17. Koutmou, K.S., Schuller, A.P., Brunelle, J.L., Radhakrishnan, A., Djuranovic, S. and Green, R. (2015) Ribosomes slide on lysine-encoding homopolymeric A stretches. *eLife*, **4**, e05534.
18. Simms, C.L., Thomas, E.N. and Zaher, H.S. (2017) Ribosome-based quality control of mRNA and nascent peptides. *Wiley Interdiscip. Rev. RNA*, **8**, e1366.
19. Guydosh, N.R. and Green, R. (2017) Translation of poly(A) tails leads to precise mRNA cleavage. *RNA*, **23**, 749–761.
20. Park, H. and Subramaniam, A.R. (2019) Inverted translational control of eukaryotic gene expression by ribosome collisions. *PLoS Biol.*, **17**, e3000396.

21. Shu,H., Donnard,E., Liu,B., Jung,S., Wang,R. and Richter,J.D. (2020) FMRP links optimal codons to mRNA stability in neurons. *Proc. Natl. Acad. Sci.*, **117**, 30400–30411.
22. Martin,R., Splitt,M., Genevieve,D., Aten,E., Collins,A., de Bie,C.I., Faivre,L., Foulds,N., Giltay,J., Ibitoye,R., *et al.* (2019) De novo variants in CNOT3 cause a variable neurodevelopmental disorder. *Eur. J. Hum. Genet.*, **27**, 1677–1682.
23. De Keersmaecker,K., Atak,Z.K., Li,N., Vicente,C., Patchett,S., Girardi,T., Gianfelici,V., Geerdens,E., Clappier,E., Porcu,M., *et al.* (2013) Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nat. Genet.*, **45**, 186–190.
24. Yang,K., Han,J., Gill,J.G., Park,J.Y., Sathe,M.N., Gattineni,J., Wright,T., Wysocki,C., de la Morena,M.T. and Yan,N. (2022) The mammalian SKIV2L RNA exosome is essential for early B cell development. *Sci. Immunol.*, **7**, eabn2888.
25. Tuck,A.C., Rankova,A., Arpat,A.B., Liechti,L.A., Hess,D., Iesmantavicius,V., Castelo-Szekely,V., Gatfield,D. and Bühler,M. (2020) Mammalian RNA decay pathways are highly specialized and widely linked to translation. *Mol. Cell*, **77**, 1222–1236.e13.
26. Yang,K., Han,J., Asada,M., Gill,J.G., Park,J.Y., Sathe,M.N., Gattineni,J., Wright,T., Wysocki,C.A., Morena,M.T. de la, *et al.* (2022) Cytoplasmic RNA quality control failure engages mTORC1-mediated autoinflammatory disease. *J. Clin. Invest.*, **132**, e146176.
27. Burke,P.C., Park,H. and Subramaniam,A.R. (2022) A nascent peptide code for translational control of mRNA stability in human cells. *Nat. Commun.*, **13**, 6829.
28. Mizuno,M., Ebine,S., Shounai,O., Nakajima,S., Tomomatsu,S., Ikeuchi,K., Matsuo,Y. and Inada,T. (2021) The nascent polypeptide in the 60S subunit determines the Rqc2-dependency of ribosomal quality control. *Nucleic Acids Res.*, **49**, 2102–2113.
29. Buschauer,R., Matsuo,Y., Sugiyama,T., Chen,Y.-H., Alhusaini,N., Sweet,T., Ikeuchi,K., Cheng,J., Matsuki,Y., Nobuta,R., *et al.* (2020) The Ccr4-not complex monitors the translating ribosome for codon optimality. *Science*, **368**, 281–282.
30. Hanson,G., Alhusaini,N., Morris,N., Sweet,T. and Collier,J. (2018) Translation elongation and mRNA stability are coupled through the ribosomal A-site. *RNA*, **24**, 1377–1389.
31. Absmeier,E., Chandrasekaran,V., O'Reilly,F.J., Stowell,J.A.W., Rappilber,J. and Passmore,L.A. (2023) Specific recognition and ubiquitination of translating ribosomes by mammalian CCR4–NOT. *Nat. Struct. Mol. Biol.*, **30**, 1314–1322.
32. Garzia,A., Jafarnejad,S.M., Meyer,C., Chapat,C., Gogakos,T., Morozov,P., Amiri,M., Shapiro,M., Molina,H., Tuschl,T., *et al.* (2017) The E3 ubiquitin ligase and RNA-binding protein ZNF598 orchestrates ribosome quality control of premature polyadenylated mRNAs. *Nat. Commun.*, **8**, 16056.
33. D'Orazio,K.N., Wu,C.C.-C., Sinha,N., Loll-Krippelbein,R., Brown,G.W. and Green,R. (2019) The endonuclease Cue2 cleaves mRNAs at stalled ribosomes during No Go Decay. *eLife*, **8**, e49117.
34. Gutierrez,E., Shin,B.-S., Woolstenhulme,C.J., Kim,J.-R., Saini,P., Buskirk,A.R. and Dever,T.E. (2013) eIF5A Promotes translation of polyproline motifs. *Mol. Cell*, **51**, 35–45.
35. Pavlov,M.Y., Watts,R.E., Tan,Z., Cornish,V.W., Ehrenberg,M. and Forster,A.C. (2009) Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 50–54.
36. Han,P., Shichino,Y., Schneider-Poetsch,T., Mito,M., Hashimoto,S., Udagawa,T., Kohno,K., Yoshida,M., Mishima,Y., Inada,T., *et al.* (2020) Genome-wide survey of ribosome collision. *Cell Rep.*, **31**, 107610.
37. Sabi,R. and Tuller,T. (2017) Computational analysis of nascent peptides that induce ribosome stalling and their proteomic distribution in *Saccharomyces cerevisiae*. *RNA*, **23**, 983–994.
38. 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.
39. Meydan,S. and Guydosh,N.R. (2020) Disome and trisome profiling reveal genome-wide targets of ribosome quality control. *Mol. Cell*, **79**, 588–602.
40. Matsuo,Y., Tesina,P., Nakajima,S., Mizuno,M., Endo,A., Buschauer,R., Cheng,J., Shounai,O., Ikeuchi,K., Saeki,Y., *et al.* (2020) RQT complex dissociates ribosomes collided on endogenous RQC substrate SDD1. *Nat. Struct. Mol. Biol.*, **27**, 323–332.
41. Yanagitani,K., Kimata,Y., Kadokura,H. and Kohno,K. (2011) Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA. *Science*, **331**, 586–589.
42. Nakatogawa,H. and Ito,K. (2002) The ribosomal exit tunnel functions as a discriminating gate. *Cell*, **108**, 629–636.
43. Bhushan,S., Hoffmann,T., Seidelt,B., Frauenfeld,J., Mielke,T., Berninghausen,O., Wilson,D.N. and Beckmann,R. (2011) SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol.*, **9**, e1000581.
44. Shanmuganathan,V., Schiller,N., Magoulopoulou,A., Cheng,J., Braunger,K., Cymer,F., Berninghausen,O., Beatrix,B., Kohno,K., von Heijne,G., *et al.* (2019) Structural and mutational analysis of the ribosome-arresting human XBP1u. *eLife*, **8**, e46267.
45. Sitron,C.S., Park,J.H., Giafaglione,J.M. and Brandman,O. (2020) Aggregation of CAT tails blocks their degradation and causes proteotoxicity in *S. cerevisiae*. *PLoS One*, **15**, e0227841.
46. Brandman,O. and Hegde,R.S. (2016) Ribosome-associated protein quality control. *Nat. Struct. Mol. Biol.*, **23**, 7–15.
47. Bengtson,M.H. and Joazeiro,C.A.P. (2010) Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. *Nature*, **467**, 470–473.
48. D'Orazio,K.N. and Green,R. (2021) Ribosome states signal RNA quality control. *Mol. Cell*, **81**, 1372–1383.
49. 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.
50. Gietz,R.D. and Schiestl,R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, **2**, 31–34.
51. Muller,R., Meacham,Z.A., Ferguson,L. and Ingolia,N.T. (2020) CiBER-seq dissects genetic networks by quantitative CRISPRi profiling of expression phenotypes. *Science*, **370**, eabb9662.
52. 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.
53. Reis,M. dos, Savva,R. and Wernisch,L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**, 5036–5044.
54. Wallace,E.W.J., Airoidi,E.M. and Drummond,D.A. (2013) Estimating selection on synonymous codon usage from noisy experimental data. *Mol. Biol. Evol.*, **30**, 1438–1453.
55. Pechmann,S. and Frydman,J. (2013) Evolutionary conservation of codon optimality reveals hidden signatures of co-translational folding. *Nat. Struct. Mol. Biol.*, **20**, 237–243.
56. Gamble,C.E., Brule,C.E., Dean,K.M., Fields,S. and Grayhack,E.J. (2016) Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell*, **166**, 679–690.
57. Dimitrova,L.N., Kuroha,K., Tatematsu,T. and Inada,T. (2009) Nascent peptide-dependent translation arrest leads to Not4p-mediated protein degradation by the proteasome *. *J. Biol. Chem.*, **284**, 10343–10352.
58. Kuroha,K., Akamatsu,M., Dimitrova,L., Ito,T., Kato,Y., Shirahige,K. and Inada,T. (2010) Receptor for activated C kinase 1 stimulates nascent polypeptide-dependent translation arrest. *EMBO Rep.*, **11**, 956–961.
59. Brandman,O., Stewart-Ornstein,J., Wong,D., Larson,A., Williams,C.C., Li,G.-W., Zhou,S., King,D., Shen,P.S., Weibezahn,J., *et al.* (2012) A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell*, **151**, 1042–1054.

60. Sitron, C.S., Park, J.H. and Brandman, O. (2017) Asc1, Hel2, and Slh1 couple translation arrest to nascent chain degradation. *RNA*, **23**, 798–810.
61. Ashe, M.P., De Long, S.K. and Sachs, A.B. (2000) Glucose depletion rapidly inhibits translation initiation in yeast. *Mol. Biol. Cell*, **11**, 833–848.
62. Gancedo, J.M. (2008) The early steps of glucose signalling in yeast. *FEMS Microbiol. Rev.*, **32**, 673–704.
63. Hoek, T.A., Khuperkar, D., Lindeboom, R.G.H., Sonneveld, S., Verhagen, B.M.P., Boersma, S., Vermeulen, M. and Tanenbaum, M.E. (2019) Single-molecule imaging uncovers rules governing nonsense-mediated mRNA decay. *Mol. Cell*, **75**, 324–339.
64. Juszkiewicz, S. and Hegde, R.S. (2017) Initiation of quality control during poly(A) translation requires site-specific ribosome ubiquitination. *Mol. Cell*, **65**, 743–750.
65. Sundaramoorthy, E., Leonard, M., Mak, R., Liao, J., Fulzele, A. and Bennett, E.J. (2017) ZNF598 and RACK1 regulate mammalian ribosome-associated quality control function by mediating regulatory 40S ribosomal ubiquitylation. *Mol. Cell*, **65**, 751–760.
66. Matsuo, Y., Ikeuchi, K., Saeki, Y., Iwasaki, S., Schmidt, C., Udagawa, T., Sato, F., Tsuchiya, H., Becker, T., Tanaka, K., et al. (2017) Ubiquitination of stalled ribosome triggers ribosome-associated quality control. *Nat. Commun.*, **8**, 159.
67. Veltri, A.J., D’Orazio, K.N., Lessen, L.N., Loll-Kripplleber, R., Brown, G.W. and Green, R. (2022) Distinct elongation stalls during translation are linked with distinct pathways for mRNA degradation. *eLife*, **11**, e76038.
68. Hickey, K.L., Dickson, K., Cogan, J.Z., Replogle, J.M., Schoof, M., D’Orazio, K.N., Sinha, N.K., Hussmann, J.A., Jost, M., Frost, A., et al. (2020) GIGYF2 and 4EHP inhibit translation initiation of defective messenger RNAs to assist ribosome-associated quality control. *Mol. Cell*, **79**, 950–962.
69. Juszkiewicz, S., Slodkiewicz, G., Lin, Z., Freire-Pritchett, P., Peak-Chew, S.-Y. and Hegde, R.S. (2020) Ribosome collisions trigger cis-acting feedback inhibition of translation initiation. *eLife*, **9**, e60038.
70. MacArthur, M.W. and Thornton, J.M. (1991) Influence of proline residues on protein conformation. *J. Mol. Biol.*, **218**, 397–412.
71. Richardson, J.S. and Richardson, D.C. (1989) The de novo design of protein structures. *Trends Biochem. Sci.*, **14**, 304–309.
72. Li, S.C., Goto, N.K., Williams, K.A. and Deber, C.M. (1996) Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 6676–6681.
73. Sinha, N.K., Ordureau, A., Best, K., Saba, J.A., Zinshteyn, B., Sundaramoorthy, E., Fulzele, A., Garshott, D.M., Denk, T., Thoms, M., et al. (2020) EDF1 coordinates cellular responses to ribosome collisions. *eLife*, **9**, e58828.
74. Barros, G.C., Requião, R.D., Carneiro, R.L., Masuda, C.A., Moreira, M.H., Rossetto, S., Domitrovic, T. and Palhano, F.L. (2021) Rqc1 and other yeast proteins containing highly positively charged sequences are not targets of the RQC complex. *J. Biol. Chem.*, **296**, 100586.
75. Radhakrishnan, A., Chen, Y.-H., Martin, S., Alhusaini, N., Green, R. and Collier, J. (2016) The DEAD-box protein Dhh1p couples mRNA decay and translation by monitoring codon optimality. *Cell*, **167**, 122–132.
76. Hanson, G. and Collier, J. (2018) Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.*, **19**, 20–30.
77. Hoekema, A., Kastelein, R.A., Vasser, M. and de Boer, H.A. (1987) Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. *Mol. Cell. Biol.*, **7**, 2914–2924.
78. Brule, C.E. and Grayhack, E.J. (2017) Synonymous codons: choose wisely for expression. *Trends Genet.*, **33**, 283–297.
79. Sabi, R. and Tuller, T. (2015) A comparative genomics study on the effect of individual amino acids on ribosome stalling. *Bmc Genomics [Electronic Resource]*, **16**, S5.
80. Artieri, C.G. and Fraser, H.B. (2014) Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res.*, **24**, 2011–2021.
81. Peil, L., Starosta, A.L., Lassak, J., Atkinson, G.C., Virumäe, K., Spitzer, M., Tenson, T., Jung, K., Remme, J. and Wilson, D.N. (2013) Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 15265–15270.
82. Tanner, D.R., Cariello, D.A., Woolstenhulme, C.J., Broadbent, M.A. and Buskirk, A.R. (2009) Genetic identification of nascent peptides that induce ribosome stalling. *J. Biol. Chem.*, **284**, 34809–34818.
83. Cao, J. and Geballe, A.P. (1995) Translational inhibition by a human cytomegalovirus upstream open reading frame despite inefficient utilization of its AUG codon. *J. Virol.*, **69**, 1030–1036.
84. Bottorff, T.A., Park, H., Geballe, A.P. and Subramaniam, A.R. (2022) Translational buffering by ribosome stalling in upstream open reading frames. *PLoS Genet.*, **18**, e1010460.
85. Gendron, T.F., Chew, J., Stankowski, J.N., Hayes, L.R., Zhang, Y.-J., Prudencio, M., Carlomagno, Y., Daugherty, L.M., Jansen-West, K., Perkerson, E.A., et al. (2017) Poly(GP) proteins are a useful pharmacodynamic marker for C9ORF72-associated amyotrophic lateral sclerosis. *Sci. Transl. Med.*, **9**, eaai7866.
86. Al-Turki, T.M. and Griffith, J.D. (2023) Mammalian telomeric RNA (TERRA) can be translated to produce valine–arginine and glycine–leucine dipeptide repeat proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **120**, e2221529120.