

# A Nascent Peptide Code for Translational Control of mRNA Stability in Human Cells

Phillip C. Burke, Heungwon Park, Arvind Rasi Subramaniam\*

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

\* Corresponding author: [rasi@fredhutch.org](mailto:rasi@fredhutch.org)

## Abstract

Stability of eukaryotic mRNAs is associated with their codon, amino acid, and GC content. Yet, coding sequence motifs that predictably alter mRNA stability in human cells remain poorly defined. Here, we develop a massively parallel assay to measure mRNA effects of thousands of synthetic and endogenous coding sequence motifs in human cells. We identify several families of simple dipeptide repeats whose translation triggers acute mRNA instability. Rather than individual amino acids, specific combinations of bulky and positively charged amino acids are critical for the destabilizing effects of dipeptide repeats. Remarkably, dipeptide sequences that form extended  $\beta$  strands *in silico* and *in vitro* drive ribosome stalling and mRNA instability *in vivo*. The resulting nascent peptide code underlies ribosome stalling and mRNA-destabilizing effects of hundreds of endogenous peptide sequences in the human proteome. Our work reveals an intrinsic role for the ribosome as a selectivity filter against the synthesis of bulky and aggregation-prone peptides.

## Introduction

Protein expression is determined by a balance between the translation rate and stability of mRNAs. In human cells, mRNA stability is often regulated by sequence motifs in the 3' untranslated region such as microRNA-binding sites and AU-rich elements (1). Additionally, the protein coding region has been recently recognized as a critical determinant of eukaryotic mRNA stability (2, 3). The role of the coding sequence in mRNA stability is best understood in the budding yeast *S. cerevisiae* where poorly translated codons and nascent peptide motifs with positively charged residues can destabilize mRNAs (4–7). Poorly translated codons have also been implicated in regulation of mRNA stability in several other organisms (8–11).

Coding sequence features regulating mRNA stability in human cells are less clear. Several recent studies examined the coding sequence determinants of endogenous mRNA stability in human cells and arrived at differing conclusions. Two studies implicated synonymous codon choice as the primary determinant of mRNA stability in human cells (12, 13). Another found GC and GC3 (wobble base GC) content as major factors regulating mRNA stability (14). A fourth study identified amino acid content to be an important contributor (15). Extended amino acid motifs and G-quadruplexes in coding regions have also been implicated as triggers of specific mammalian mRNA decay pathways (16, 17). The associations reported in these studies relied on endogenous human coding sequences. Since human mRNAs differ from each other in codon, amino acid, and GC content as well as in their length and the presence of specific se-

quence motifs, it is challenging to identify the contribution of each factor to mRNA stability. Further, reporters used in the above studies for validation differ extensively in their nucleotide or amino acid content, which complicates their interpretation.

Here, we developed a massively parallel assay to measure the mRNA effects of thousands of coding sequence motifs in human cells. We designed our assay with the initial goal of systematically delineating the individual contribution of mRNA features implicated in previous studies. Instead, we unexpectedly uncovered a potent role for the sequence and structure of the nascent peptide in regulating mRNA stability and ribosome elongation rate. The resulting nascent peptide code regulates the ribosome stalling and mRNA destabilizing effects of hundreds of endogenous peptide sequences from the human proteome. Our results point to an unappreciated role for the ribosome as a selectivity filter against the synthesis of bulky and aggregation-prone peptide sequences.

## Results

### A massively parallel assay for mRNA levels in human cells

We reasoned that coding sequence motifs that alter mRNA stability should be identifiable through their effects on steady state mRNA levels. To study the effect of coding sequence motifs on mRNA levels in an unbiased manner, we designed a library of 4,096 oligonucleotides made of all possible codon pairs (Fig. 1A). We repeated each codon pair as a tandem 8x repeat with the rationale that their effects will be amplified and readily mea-

surable. We cloned the oligonucleotide library as a pool into a dual fluorescence reporter vector separated by 2A linkers – a design widely used for studying ribosome stalling motifs in human cells (18–23). We added multiple random 24nt barcodes without stop codons 3' of each oligonucleotide insert and linked the barcode sequences to the corresponding insert by high-throughput sequencing. Most studies of coding sequence motifs use transient transfection or lentiviral integration of reporters, which makes measurement of steady state effects on mRNA levels across a large pool difficult. To avoid this, we stably integrated the reporter pool at the AAVS1 locus of HEK293T cells using CRISPR Cas9-mediated homologous recombination. We extracted mRNA and genomic DNA from the pooled cells and counted each barcode by high-throughput sequencing. Normalization of the total barcode count in the mRNA by the corresponding count in the genomic DNA for each of the 4,096 inserts provides a relative measure of the steady-state mRNA level of that insert.

We examined whether our assay captured the effects of known mRNA-destabilizing motifs. Stop codons in either the first or second position of the codon pair repeat decrease mRNA levels (Fig. 1B), consistent with their mRNA destabilizing effect due to nonsense-mediated decay (24–26). However, the destabilizing effects of codon pairs are only weakly correlated with known effects of GC content, synonymous codon usage, or amino acid content on mRNA stability (Fig. S1A–C) (12–15). Among the twenty amino acids, the positively charged amino acids lysine and arginine cause the largest average decreases in mRNA levels (Fig. 1B). This is in line with the known association between positively charged residues in the nascent peptide and slow elongation (27–31).

### Specific dipeptide repeats trigger mRNA instability

We wondered whether the mild average effects of amino acids on mRNA levels (Fig. 1B) belie larger effects driven by specific amino acid combinations. Therefore, we assessed the effect of each pairwise amino acid combination on mRNA level (Fig. 1C). While lysine and arginine are destabilizing on average (Fig. 1B), unexpectedly, these amino acids have mild or no destabilizing effect on mRNA levels on their own (Fig. 1C: Lys-Lys, Arg-Arg, Arg-Lys). Rather, the destabilizing effects of lysine and arginine are primarily driven by co-occurrence with bulky amino acids (32) (ratio of side chain volume to length > 18Å<sup>2</sup>) such as valine, isoleucine, leucine, phenylalanine, and tyrosine (Fig. 1C). Likewise, most bulky amino acids are highly destabilizing in combination with lysine and arginine, but not on their own (Fig. 1C). Further, a few dipeptides that contain certain pos-

itively charged amino acids (Arg-His) or bulky amino acids (Phe-Ser) also have a strong mRNA destabilizing effect (Fig. 1C). The combinatorial effect of positively charged and bulky amino acids on mRNA level is captured by a linear statistical model (Fig. 1D): Isoelectric point (32) (pI, a measure of positive charge) and bulkiness (32) of amino acids are positive correlates of mRNA level, while an interaction term between these two physical properties is a negative correlate of mRNA level [mRNA = (0.31 × pI) + (0.20 × bulkiness) – (0.03 × pI × bulkiness), Adjusted R<sup>2</sup> = 0.25]. By contrast, ignoring the interaction between pI and bulkiness results in negative or no correlation of these properties with mRNA level (mRNA = – 0.18 × pI, Adjusted R<sup>2</sup> = 0.21), which is in line with Fig. 1B. The destabilizing effects of dipeptide repeats in the translated +0 frame strongly correlates with the codon-matched +3 frame, but only weakly with the codon-mismatched +1 and +2 frames (Fig. 1E). The high correlation between the +0 and +3 frames is also seen from the diagonal symmetry of Fig. 1C and arises from similarity of the encoded peptides (for eg. (XY)<sub>8</sub> and (YX)<sub>8</sub> are identical except at their termini). These frame correlations are consistent with the mRNA destabilizing effects arising at the translational level as opposed to transcriptional or RNA processing differences. Together, our results show that translation of bulky and positively charged amino acids is critical for their negative effect on mRNA level.

### Nascent peptide primary sequence modulates mRNA stability

Several observations suggest that translation of specific dipeptide repeats is a general trigger of mRNA instability in human cells. First, mRNA levels of destabilizing repeats are lower than their frameshifted controls in multiple human cell lines (HEK293T, HeLa, HCT116, and K562; Fig. 2A), pointing to the generality of the observed mRNA effects. Second, upon actinomycin D treatment to inhibit transcription, mRNA levels of reporters with destabilizing dipeptides decay faster than their frameshifted controls (Fig. 2B). This confirms that the decrease in steady-state mRNA levels caused by dipeptide repeats arises from reduction in mRNA stability. Third, to decipher the effect of dipeptide repetition on mRNA instability, we systematically varied the number of several destabilizing dipeptides identified in our initial assay (Fig. 2C). As the number of dipeptide repeats increases from 1 to 8, each dipeptide becomes destabilizing at a distinct repeat number between 4 and 7 (Fig. 2C). Fourth, we altered the periodicity of dipeptide repeats by intermixing dipeptides with their reversed counterparts such that the overall amino acid composition remains unchanged (Fig. 2D). Even minor

perturbations of RH repeats completely abrogate their mRNA destabilizing effect (Fig. 2D, upper panel). VK repeats become gradually less destabilizing as their periodicity is decreased, while SF repeats show an intermediate trend (Fig. 2D, middle and lower panels). This last experiment reveals that the primary sequences of dipeptide repeats encode critical regulatory information beyond the identity of the amino acid pairs forming the repeats.

## Secondary structure of dipeptide repeats mediates mRNA stability effects

Since dipeptide sequences are known to form distinct secondary structures based on their periodicity (33, 34), we asked whether mRNA-destabilizing dipeptide repeats adopt specific secondary structures. Using a deep neural network model for secondary structure prediction (35), we assigned dipeptide repeats to either  $\alpha$  helices or  $\beta$  strands if their respective prediction probabilities are greater than 0.5. We find that dipeptide repeats predicted to form  $\beta$  strands have a significantly lower mRNA level on average than those predicted to form  $\alpha$  helices (Fig. 3A,  $P < 0.001$ , two-sided Mann-Whitney test). Further, many dipeptide repeats that strongly destabilize mRNAs *in vivo* are also computationally predicted to form  $\beta$  strands with a high probability (Fig. 3B). Notably, among dipeptides containing the positively charged amino acids lysine or arginine, the measured propensity of the second amino acid to occur in a  $\beta$  strand (36) ('Chou-Fasman propensity') is highly correlated with mRNA instability (Fig. 3C, left; Fig. S2A). This correlation is not observed with  $\alpha$  helix propensities of the same amino acids (Fig. 3C, right; Fig. S2A). mRNA levels of dipeptide repeats containing the negatively charged glutamate, which are also predicted to form  $\beta$  strands with high probability when combined with bulky amino acids (Fig. S2B), do not show significant correlation with  $\beta$  strand or  $\alpha$  helix propensities (Fig. S2C). Thus, a combination of bulky and positively charged amino acids in the primary sequence and  $\beta$  strand in the secondary structure are strong and significant predictors of the mRNA-destabilizing effects of dipeptide repeats [mRNA =  $(0.30 \times \text{pl}) + (0.23 \times \text{bulkiness}) - (0.03 \times \text{pl} \times \text{bulkiness}) - (0.52 \times \beta\text{-strand-propensity})$ , Adjusted  $R^2 = 0.27$ ].

## Extended $\beta$ strands drive ribosome stalling and mRNA instability

To test the causal role of  $\beta$  strands in nascent peptide-mediated translational control, we combined the mRNA-destabilizing dipeptides VK, KV, SF, and FS into 16 amino acid-long peptides. Even though the four constituent dipeptides are strongly predicted to form  $\beta$  strands on their own (Fig. 3A), their combinations can

form either  $\beta$  strands or  $\alpha$  helices with high probability (Fig. 4A). Importantly, all combinations are encoded by the same set of four amino acids to control for amino acid composition. We commercially synthesized two 16 amino acid peptides and used circular dichroism to confirm their secondary structure *in vitro* (Fig. 4B, left panel). As predicted (Fig. 4A), 4xSVKF primarily forms  $\beta$  strands in aqueous solution, while 4xSKVF forms  $\alpha$  helices in the presence of trifluoroethanol (TFE) as a co-solvent (37–39) (Fig. 4B, right panel). We then measured the transit time of ribosomes on mRNAs encoding 16 amino acid inserts preceding a Nanoluciferase reporter in a rabbit reticulocyte lysate (RRL) *in vitro* translation system (Fig. 4C, left). The  $\beta$  strand-forming 4xSVKF and 4xVKFS inserts slow ribosome elongation relative to the  $\alpha$  helix-forming 4xSKVF and 4xKVFS inserts (Fig. 4C, middle), with a 200 s difference in *in vitro* transit time (Fig. 4C, right). Strikingly, all  $\beta$  strand peptides decrease mRNA levels over 8-fold relative to  $\alpha$  helix controls when tested *in vivo* using our reporter assay (Fig. 4D). We observe similar destabilizing effects due to  $\beta$  strand formation in HeLa, HCT116, and K562 cells (Fig. S3). We conclude that the propensity of nascent peptides to form  $\beta$  strands is a critical determinant of dipeptide-mediated ribosome stalling and mRNA instability in human cells.

## Dipeptide motifs in the human genome trigger ribosome stalling and mRNA instability

We sought to identify endogenous sequences in the human genome that regulate mRNA translation based on the dipeptide code identified above. To do this, we scanned all annotated human protein coding sequences for destabilizing dipeptide combinations of bulky and positively charged amino acids (Fig. 5A). Using a heuristic peptide score (Fig. 5A, top), we identified the 16 amino acid long peptide within each coding sequence that has the maximum density of destabilizing dipeptides. mRNA regions encoding these dipeptide motifs exhibit increased ribosome occupancy relative to control motifs from the same mRNAs in HEK293T cells (23) (Fig. 5B). A similar increase in ribosome occupancy is also seen in other human and mouse cell lines (40, 41) (Fig. S4A,B), while total RNA density does not show such an increase (Fig. S4A,B). The peak in P-site density at about 25nt from the center of the motif (Fig. 5B) suggests that the 16 amino acid peptides are almost fully synthesized before they stall the ribosome.

To test whether the endogenous motifs identified above can destabilize mRNA, we cloned 1,201 such motifs into our reporter and measured their mRNA levels by high throughput sequencing (Fig. 5C, top). Motifs with high destabilizing dipeptide content result in lower mRNA lev-



els than control motifs ( $P < 0.01$ , Fig. 5C, left panel). Among destabilizing motifs, those predicted to form  $\beta$  strands result in lower mRNA levels than the remaining motifs ( $P < 0.05$ , Fig. 5C, right panel). To confirm the destabilizing role of the specific dipeptides identified in our study, we disrupted them by moving the bulky and positively charged amino acids to opposite ends without changing the amino acid composition in 1,079 endogenous motifs (Fig. 5D, top). As predicted, the resulting mutations increase mRNA levels (median  $\log_2 \Delta \text{mRNA} = 0.38$ ) with 783 mutated motifs having significantly higher mRNA levels ( $P < 0.05$ ) than their wild-type counterparts (Fig. 5D, bottom). Examination of destabilizing motifs with annotated  $\beta$  strand structures in the Protein Data Bank (PDB) shows that these  $\beta$  strands are part of anti-parallel  $\beta$  sheets and are significantly longer than the 5-6 residue length of typical  $\beta$  strands (42) (Fig. 5E). Together, these results show that  $\beta$ -stranded endogenous motifs containing bulky and positively charged dipeptides can stall ribosomes and trigger mRNA instability.

## Discussion

Here, we identify a combinatorial code composed of bulky, positively charged, and extended  $\beta$  strand nascent peptides that regulates translation and mRNA stability in human cells. We demonstrate that a minimal combination of these sequence and structural elements is sufficient to induce ribosome stalling and cause changes in gene expression, and is widespread in the human proteome. As discussed below, elements of the code uncovered here allow us to synthesize a large body of observations on nascent peptide-mediated stalling of ribosomes and regulation of mRNA stability in human cells. Our results also point to a role for the ribosome as a post-synthesis filter against nascent peptide sequences that are bulky and aggregation prone.

While positive charge in the nascent peptide can slow ribosomes (27, 28), our results show that positive charge by itself is insufficient to induce changes in gene expression in human cells. The importance of bulky amino acids for mRNA stability effects is in line with the role of side chain bulk in ribosome-associated quality control in *S. cerevisiae* (43). Further, bulky synthetic amino acid analogs in the nascent peptide and small molecules that add bulk to the exit tunnel can both reduce ribosome elongation rate (44–47). Ribosome profiling in *S. cerevisiae* and human cells shows that tripeptide combinations of bulky and positively charged amino acids are enriched at sites of increased ribosome density (23, 48). Bulky and positively charged amino acids also play criti-

cal roles in many known ribosome-arresting peptides (7, 49–52). Structural studies of arrest peptides suggest that the bulky and positively charged amino acids might stall ribosomes by altering the geometry of the peptidyl-transferase center (PTC) and/or by steric interactions with the constriction point in the exit tunnel formed by the uL4 and uL22 proteins (7, 51, 53, 54).

Our work shows that extended  $\beta$  strand motifs in nascent peptides contribute to ribosome stalling and mRNA instability in human cells. This role of a simple secondary structural motif like  $\beta$  strand is surprising given that cryo-EM studies of stalled ribosome nascent chain complexes reveal a diverse range of extended conformations, turns, and helices that are specific to each nascent arrest peptide (7, 47, 52, 55, 56). This comparison is complicated by the fact that cryo-EM studies are performed on post-arrest complexes where the nascent chain might have already undergone extensive conformational rearrangements. Further, while several destabilizing motifs uncovered here form  $\beta$  strands *in silico* and *in vitro* in isolation, they might have a significantly different structure within the confined geometry of the ribosome exit tunnel (39, 57–59). At the molecular level,  $\beta$  strands in nascent chains could contribute to stalling as an allosteric relay that communicates steric interactions between the nascent chain and the distal portions of the ribosome exit tunnel such as the uL4/uL22 constriction to the PTC (45, 54, 55, 60). This possibility is supported by our observation that destabilizing dipeptide repeats are at least 10-12 acids long (Fig. 2C), which is consistent with the distance between the uL4/uL22 constriction and the PTC.

The nascent peptide code for mRNA stability described here is significantly more complex and localized along the mRNA than previously associated factors such as codon, amino acid, and GC content (12–15). Thus, ribosome stalling nascent peptides might exert their effects on mRNA stability through distinct cellular pathways compared to the ones sensing codon, amino acid, and GC content of mRNAs (22, 61–63). Until our work, the UL4 CMV leader peptide, poly-lysine sequences encoded by poly-A, and the *Xbp1* arrest sequence were the only known nascent peptide motifs with ribosome stalling ability in human cells (64–66). However, the effect of these stall sequences on mRNA stability is not well characterized, with mild or no effects on steady state mRNA levels seen in transiently transfected reporter assays (3, 67, 68). Several other nascent peptide sequences stall human ribosomes in the presence of small molecule metabolites or drugs in the ribosome exit tunnel (69, 70). It is likely that the effect of nascent peptide motifs on ribosome stalling and mRNA stability

is modulated by other co-translational events such as nascent protein folding outside the ribosome (71, 72), membrane insertion (73, 74), and multiprotein assembly (75, 76).

In addition to the sequence and structural determinants of nascent peptide-mediated ribosome stalling studied here, several classes of nascent peptide sequences that stall ribosomes might not be revealed by our assay. For example, poly-prolines do not emerge as destabilizing motifs in our assay even though they are known to stall ribosomes (23). This is likely because poly-proline stalls are resolved without triggering quality control or mRNA instability (17). While extended  $\beta$  strands are the primary structural motif associated with stalling here, we also find motifs with unstructured regions that are nevertheless destabilizing (Figs. 3A, 5B). This might be in part due to limitations of existing computational methods (35) to predict secondary structures or their limited relevance to secondary structures forming inside the ribosome. It is also likely that the combinatorial code of positively charged, bulky, and  $\beta$  strand sequences uncovered here underlies some, but not all, classes of nascent peptides that have the potential to stall ribosomes and effect changes in gene expression.

The nature of the nascent peptide code uncovered here has important implications for cellular homeostasis and disease. Ribosome stalling and mRNA destabilization induced by bulky and extended  $\beta$  strands, which are highly aggregation prone (77, 78), implies that the ribosome has an intrinsic ability to throttle the synthesis of such proteins. Such throttling might serve as a quality control mechanism to test the ability of  $\beta$  strands to fold into anti-parallel  $\beta$  sheets and thus avoid aggregation. This ribosomal selectivity filter acts before other co-translational mechanisms such as codon optimality that help avoid aggregation after  $\beta$  strands emerge from the ribosome (79, 80). Finally, the gene regulatory potential of the destabilizing dipeptide motifs uncovered here suggests that disease-causing missense mutations occurring at these motifs might exert their phenotype by altering protein expression *in cis* rather than protein activity.

## Author Contributions

P.C.B. 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.

## Acknowledgements

We thank members of the Subramaniam lab, the Zid lab, the Basic Sciences Division, and the Computational Biology program at Fred Hutch for discussions and feedback on the manuscript. This research was funded by NIH R35 GM119835, NSF MCB 1846521, and the Sidney Kimmel Scholarship received by ARS. This research was supported by the Genomics Shared Resource of the Fred Hutch/University of Washington Cancer Consortium (P30 CA015704) and Fred Hutch Scientific Computing (NIH grants S10-OD-020069 and S10-OD-028685). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Competing interests

None

## Materials and Methods

### Plasmid construction

Plasmids, oligonucleotides, and cell lines used in this study are listed in Supplementary Tables S1-S3.

### Parent vector construction

The AAVS1-targeting parent vector pPBHS285 used for this study was constructed using Addgene plasmid #68375 (81) as a backbone. The PGK1 promoter was replaced with the CMV promoter and the native pCMV 5' UTR region. The coding sequence was replaced by a codon-optimized mKate2 and eYFP fusion cassette, linked with two 2A linker sequences. These 2A sequences surround a cassette encoding an EcoRV restriction site, Illumina R1 sequencing primer binding site, and a T7 promoter. The R1 primer binding and T7 sequences are in reverse orientation (3' - 5') for *in vitro* transcription and sequencing of inserts and barcode sequences at the EcoRV site.

### Variable oligo pool design

Four oligo pools were designed for this study.

Pool 1 (Fig. 1B-E, Fig. 3A,C) encodes all possible dicodon (6 nt) combinations, for a total of 4096 codon pairs. These 6nt dicodon inserts were repeated eight times to create 8x dicodon repeat inserts, each 48nt in length.

Pool 2 (Fig. 2C,D, Fig. 4D) encodes several dipeptide combinations identified in Library 1 as causing mRNA instability. For Fig. 2C, the number of dipeptide repeats was systematically reduced from 8 to 1. Repeats were replaced with a Ser-Gly linker, shown to be not destabilizing in Library 1, to maintain a constant 48nt insert length. For Fig. 2D, periodicity of dipeptides was by in-

terspersing 1, 2, or 4 tandem repeats of each dipeptide with an equal number of its sequence-reversed counterpart. For Fig. 4D, destabilizing dipeptides KV and SF were combined and rearranged to form either alpha helices or beta strands, as predicted by S4PRED.

Pool 3 (Fig. 5C-E) encodes the 16 amino acid nascent peptide motifs from the human proteome identified as potentially destabilizing by the scoring method described in Fig. 5A along with 4 flanking codons on either side. The library encodes the top 1079 predicted stalling motifs with a peptide score > 9, and 122 control motifs with a peptide score < 3. The library also includes the mutants with reordered amino acids from the 1,079 endogenous destabilizing dipeptide motifs, which were designed as described in Fig. 5D.

Pool 4 (Fig 2A,B, Fig S3) encodes 8 inserts: 3 destabilizing dipeptide repeats (RH)<sub>8</sub>, (VK)<sub>8</sub>, (SF)<sub>8</sub>, their respective frameshift controls (PS)<sub>8</sub>, (QS)<sub>8</sub>, (FQ)<sub>8</sub>, the  $\beta$  strand peptide (SVKF)<sub>4</sub>, and the  $\alpha$  helix peptide (SKVF)<sub>4</sub>.

Oligo pools 1-3 were synthesized by Twist Biosciences with flanking sequences for PCR and cloning into the EcoRV site of the parent pPBHS285 vector. Oligo pool 4 was cloned by PCRing individual inserts and pooling them before cloning.

### Plasmid library construction

Parent vector pPBHS285 was digested with EcoRV. The oligo pools described above were PCR amplified using primers oHJ01 and either oPB348 (Library 1) or oPB409 (Libraries 2–4). oPB348 and oPB409 both encode a 24 nt random barcode region, comprised of 8×VNN repeats to exclude in-frame stop codons (where V is any nucleotide except T). Barcoded oligo pools were cloned into pPBHS285 by Gibson assembly. Assembled plasmid pools were transformed with high efficiency into NEB10Beta *E.coli*. For pools 1-3, the transformed plasmid pools were extracted from 15-50 *E.coli* colonies per insert in the library, thus bottlenecking the number of unique barcodes present in each plasmid pool. Resulting plasmid pools contained between 60,000–400,000 unique barcode sequences for pools 1-3. For pool 4, the transformed library was bottlenecked to around 150 barcodes per insert, and 6 such pools with distinct barcodes were extracted for multiplexed library preparation of different cell lines.

The plasmid libraries corresponding to pools 1-4 are pPBHS286, pPBHS309, pPHS296, and pPHS406, respectively. Variable insert and barcode sequences for each plasmid library are provided as part of the data analysis code.

### CRISPR vectors

The CLYBL-targeted Cas9-BFP expression vector pPHS15 was constructed by Golden Gate assembly of either entry plasmids or PCR products with pPHS11 (MTK0\_047 (82) Addgene #123977) as backbone, pPHS3 (MTK2\_007 (82) Addgene #123702) for the pEF1a promoter, pADHS5 (83) (pU6-(BbsI)\_CBh-Cas9-T2A-BFP (84) Addgene #64323) for the Cas9-2A-BFP insert cassette, and pPHS6 (MTK4b\_003 (82) Addgene #123842) for the rabbit  $\beta$ -globin terminator. sgRNA vectors pPBHS320 (gRNA\_AAVS1-T1 Addgene #41817) and pADHS4 (83) (eSpCas9(1.1)\_No\_FLAG\_AAVS1\_T2 Addgene #79888) were used for insertion at the AAVS1 locus. pASHS16 (MTK234\_030 spCas9-sgRNA1-hCLYBL (82) Addgene #123910) was used for insertion at the CLYBL locus.

### Cell line maintenance and generation

HEK293T cells (RRID:CVCL\_0063, ATCC CRL-3216), HCT116 cells (RRID:CVCL\_0291, NCI60 cancer line panel), and HeLa cells (RRID:CVCL\_0030, ATCC CCL-2) were grown in DMEM (Thermo 11965084). K562 cells (RRID:CVCL\_0004, ATCC CCL-243) were grown in IMDM (Thermo 12440053). Media for all cells was supplemented with 10% FBS (Thermo 26140079). Cells were grown at 37C in 5% CO<sub>2</sub>. All transfections into HEK293T, HCT116, and HeLa cells were performed using Lipofectamine 3000 (Thermo L3000015). Transfections into K562 cells were performed using an Amaxa Nucleofector V kit (Lonza VCA-1003). HEK293T cells that stably express Cas9 (hsPB80) were generated by transfecting the CLYBL::Cas9-BFP vector pPHS15 and spCas9 sgRNA1 hCLYBL vector, and selecting with 200  $\mu$ g/mL hygromycin.

### CRISPR integration of plasmid libraries

hsPB80 CLYBL::Cas9-BFP HEK293T cells were seeded to 50% confluency on 15 cm dishes for all library transfections. 10  $\mu$ g of library plasmid (pPBHS286, pPBHS309, or pPHS296) and 1.5  $\mu$ g of each AAVS1 targeting CRISPR vector were transfected per 15 cm dish. pPBHS286, and pPBHS309 were each transfected into a single 15 cm dish. pPHS296 was transfected into three 15 cm dishes. pPHS406 pools with different barcodes were transfected into single 10 cm dishes of hsPB80, HeLa, HCT116 and 2 million cells of K562. Cells were selected with 2  $\mu$ g/mL puromycin, added 48 hours post-transfection. Cells from the three pPHS296 transfections were combined at the start of selection. Puromycin selection was removed after 6-10 days, once cells were growing robustly in selection. 24 hours after removing puromycin selection, stable li-



brary cells were plated into two separate 15cm dishes, to reach 75% confluency the next day, for matched mRNA and gDNA harvests. For pHPHS406, libraries were maintained in two 10 cm dishes or T75 flasks (for K562).

### mRNA stability measurement

hsPB80 cells containing the stably integrated pHPHS406 library were seeded to 50% confluence in a 6-well plate. Actinomycin D (ActD) power was dissolved in DMSO at 1 mM (1.25 mg/mL) and added to each well of the 6-well plate to a final concentration of 5  $\mu$ g/mL. Before harvesting, 1 million HeLa cells containing the pHPHS406 library were lysed in 6 mL of Trizol reagent, to create a Trizol lysis solution containing a set number of mRNAs with different barcodes than those in the hsPB80 pHPHS406 pool, for barcode count normalization across samples. ActD treated hsPB80 wells were harvested at 0, 0.5, 1, 2, 4, and 6 hours after the addition of ActD by adding 0.75 mL of the Trizol lysis solution above to wells at each timepoint, then following the manufacturer's Trizol mRNA extraction protocol.

### Library Genomic DNA extraction

Reporter library genomic DNA was harvested from one 75% confluent 15 cm or 10 cm dish of stably expressing library cells. Genomic DNA was harvested using Quick-DNA kit (Zymo D3024), following the manufacturer's instructions, with 3 mL of genomic DNA lysis buffer per 15 cm plate, and 1 mL of the same buffer per 10 cm plate. Between 0.5-10  $\mu$ g of purified genomic DNA from each library sample was sheared into ~350 nucleotide length fragments by sonication for 10 min on ice using a Diagenode Bioruptor. Sheared 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 (85, 86), using a HiScribe T7 High Yield RNA Synthesis Kit (NEB E2040S). Transcribed gRNA was treated with DNase I (NEB M0303S) and cleaned using an RNA Clean and Concentrator kit (Zymo R1013).

### Library mRNA extraction

Reporter library mRNA was harvested from one 75% confluent 15 cm or 10 cm dish of stably expressing library cells. mRNA was harvested by using 3 mL of Trizol reagent (Thermo) to lyse cells directly on the plate, and then following the manufacturer's mRNA extraction protocol. Purified mRNA was then DNaseI (NEB M0303S) treated and cleaned using an RNA Clean and Concentrator kit (Zymo R1013).

### mRNA and genomic DNA barcode sequencing

Between 0.5-10  $\mu$ g of DNaseI-treated mRNA and gRNA for each library was reverse transcribed into cDNA using Maxima H Minus Reverse Transcriptase (Thermo EP0752) and a primer annealing to the Illumina R1 primer binding site (oPB354). A 170-nucleotide region surrounding the 24-nucleotide barcode was PCR amplified from the resulting cDNA in two rounds, using Phusion Flash High-Fidelity PCR Master Mix master-mix (Thermo F548L). Round 1 PCR was carried out for 10 cycles, with cDNA template comprising 1/10th of the PCR reaction volume, using primers oPB361 and oPB354. Round 1 PCRs were cleaned using a 2x volume of Agencourt Ampure XP beads (Beckman Coulter A63880) to remove primers. Cleaned samples were then used as template for Round 2 PCR, carried out for 5-15 cycles, using a common reverse primer (oAS111) and indexed forward primers for pooled high-throughput sequencing of different samples (oAS112-135 and oHP281-290). Amplified samples were run on a 1.5% agarose gel and fragments of the correct size were purified using ADB Agarose Dissolving Buffer (Zymo D4001-1-100) and UPrep Micro Spin Columns (Gene-see 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 HiSeq 2500 or Illumina NextSeq 2000 in 1x50, 2x50, or 1x100 mode (depending on other samples pooled with the sequencing library).

### Insert-barcode linkage sequencing

Plasmid library pools 1-4 (pPBHS286, pPBHS309, pHPHS296, and pHPHS406) were diluted to 10 ng/ $\mu$ L. A 240-nucleotide region surrounding the 48-nucleotide variable insert sequence and the 24-nucleotide barcode was PCR amplified from these pools in two rounds, using Phusion Flash High-Fidelity PCR Master Mix master-mix (Thermo F548L). Round 1 PCR was carried out for 10 cycles, with 10 ng/ $\mu$ L plasmid pool template comprising 1/10th of the PCR reaction volume, using primers oPB29 and oPB354. Round 1 PCRs were digested with DpnI (Thermo FD1704) at 37°C for 30 minutes to remove template plasmid and cleaned using a 2x volume of Agencourt Ampure XP beads (Beckman Coulter A63880) to remove primers and enzyme. Cleaned samples were used as template for Round 2 PCR, for 5 cycles, using oAS111 and indexed forward primers (oAS112-135 and oHP281-290). Amplified Round 2 PCR products were purified after size selection and quantified as described above for barcode sequencing. Samples were sequenced using an Illumina MiSeq or Illumina NextSeq 2000 in 2x50 or 1x100 mode.

## Rabbit reticulocyte nanoluciferase transit time assay

DNA fragments encoding 4xKVFS and 4xSKVF ( $\alpha$  helix) and 4xVKFS and 4xSVKF ( $\beta$  strand) peptides were generated by PCR-amplifying overlapping oligos that encode each sequence in the forward and reverse direction (oPB470-473 and oPB488-491). Nanoluciferase cassette was amplified from an IDT gBlock (oPN204) using oAS1287 and oPB465. Insert sequences and the Nanoluciferase cassette were combined by overlap PCR using oPB464 and oPB462, which add a 5' T7 promoter site and a 3' polyA tail to the amplified reporter template, with oAS1545 used to bridge oPB462 annealing. Resulting insert-Nanoluciferase cassette sequences were confirmed by Sanger sequencing. The PCR products were transcribed into mRNA using a HiScribe T7 High Yield RNA Synthesis Kit (NEB E2040S). mRNA was cleaned using an RNA Clean and Concentrator kit (Zymo R1013). *In vitro* Nanoluciferase reporter translation reactions were performed as described in Susorov et al. 2020 (87). Reaction mixture containing 50% of nuclease-treated rabbit reticulocyte lysate (RRL) (PRL4960, Promega) was supplemented with 30 mM Hepes-KOH (pH = 7.5), 50 mM KOAc, 1.0 mM Mg(OAc)<sub>2</sub>, 0.2 mM ATP and GTP, 0.04 mM of 20 amino acids (PRL4960, Promega), and 2 mM DTT. Nanoluciferase substrate furimazine (PRN1620, Promega) was added to the mixture at 1%. 15  $\mu$ L aliquots of the mixture were placed in a 384-well plate and incubated at 30°C for 5 min in a microplate reader (Tecan INFINITE M1000 PRO). Translation reactions were started by simultaneous addition of 3  $\mu$ L mRNA, to a final concentration of 10 ng/ $\mu$ L, and luminescence signal was recorded every 10 seconds over a period of 25 minutes.

## Circular dichroism

4xSKVF ( $\alpha$  helix) and 4xSVKF ( $\beta$  strand) peptides were commercially synthesized (Genscript) at >90% purity level. Peptides were dissolved in water to 400  $\mu$ M concentration, then diluted into 10 mM sodium-phosphate buffer (pH = 7.5) and 0, 20, or 40 volumetric percent of 2,2,2-trifluoroethanol (TFE) to final concentrations ranging between 15-30  $\mu$ M. CD spectra were measured at 25°C using a Jasco J-815 Circular Dichroism Spectropolarimeter. The CD spectra were recorded between 180-260 nm with a resolution of 0.5 nm for both peptides and blank buffer solutions in 1 mm cuvettes.

## Computational analyses

Pre-processing steps for high-throughput sequencing were implemented as Snakemake workflows (88). Python (v3.7.4) and R (v3.6.2) programming languages were used for all analyses unless mentioned otherwise.

In the description below, files ending in .py refer to Python scripts and files ending in .Rmd or .R refer to R Markdown or R scripts.

## Barcode to insert assignment

The raw data from insert-barcode linkage sequencing are in .fastq format. If the inserts and barcodes were on paired-end reads instead of single-end reads, the reads were renamed in increasing numerical order starting at 0 to enable easy matching of insert and barcode reads. This was done in `rename_fastq_paired_reads.py`. The oligo pools were used to create a reference fasta file in `create_reference_for_aligning_library.R`. A bowtie2 (89) (v2.4.2) reference was created from the fasta file using the bowtie2-build command with default options. The insert read was aligned to the bowtie2 reference using bowtie2 command with options -N 1 -L 22 --end-to-end with the --trim5 and --trim3 options set to include only the region corresponding to the insert. The alignments were sorted and indexed using samtools (90) (v1.11) commands sort and index with default options. The alignments were filtered to include only reads with simple cigar strings and a MAPQ score greater than 20 in `filter_alignments.R`. The barcodes corresponding to each filtered alignment were parsed and tallied in `count_barcode_insert_pairs.py`. Depending on the sequencing depth, only barcodes that were observed at least 4-10 times were included in the tally. The tallied barcodes were aligned against themselves using bowtie2-build with default options and bowtie2 with options -L 24 -N 1 --all --norc. The self-alignment was used to exclude barcodes that are linked to distinct inserts or ones 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. The final list of insert barcode pairs is written as a tab-delimited .tsv.gz 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. The filtered barcodes .tsv.gz file from the insert-barcode linkage sequencing is used to create a reference fasta file in `create_bowtie_reference.R`. A bowtie2 (v2.4.2) reference was created from the fasta file using the bowtie2-build command with default options. The barcodes were aligned to the bowtie2 reference using bowtie2 command with options -N 1 -L 20 --norc with the --trim5 and --trim3 options set to include only the region corresponding to the barcode. The alignments were sorted, indexed, and tallied using the



samtools commands sort, index, idxstats with default options. GNU awk (v4.1.4) was used for miscellaneous processing of tab-delimited data between preprocessing steps. The final list of counts per barcode in each sample of genomic DNA or mRNA is written as a tab-delimited .tsv.gz file for calculating mRNA levels below.

### mRNA quantification

All barcode counts corresponding to each insert in each sample were summed. Only inserts with a minimum of 200 reads and 6 barcodes summed across the mRNA and gRNA samples were included. Otherwise the data were designated as missing. mRNA levels were calculated as the log2 ratio of the summed mRNA barcode counts to the summed gRNA barcode counts. mRNA levels were median-normalized within each library. For mRNA stability measurements, the summed mRNA counts for each insert at each time point were normalized by the total barcode counts for the spiked-in HeLa cells at the same time point. Then, the spike-in normalized mRNA levels for each insert were further normalized to the time 0 value.

### Linear statistical modeling of mRNA levels

Amino acid scales for isoelectric point pI, bulkiness, and secondary structure propensity were taken from prior studies (32, 36, 91). The median-normalized mRNA levels for lysine, arginine, or glutamate dipeptides were modeled as a function of amino acid scales (as indicated in the figures) using the R function lm with default parameters. Only fit coefficients significantly different from zero ( $P < 0.05$ ) are reported for each linear model.

### Secondary structure prediction

Secondary structure was predicted solely from the amino acid sequence using the default single sequence model in S4PRED (35) (downloaded from <https://github.com/psipred/s4pred> on Apr 17, 2021) and the neural network was used without any modification in predict\_secondary\_structure.py.

Cartoons of 4xSVKF and 4xSKVF in Fig. 4A were predicted using the PEP-FOLD3 server (92) with default parameters and the resulting PDB files were visualized using PyMOL software (Schrodinger).

### Calculation of secondary structure content from circular dichroism

The raw circular dichroism data (Fig. 4B, left panel) were converted to the two-column spectrum file format as required for SESCA (93) (v095, downloaded from <https://www.mpibpc.mpg.de/sesca> on Jul 28, 2021). Secondary structure was estimated using the SESCA script SESCA\_deconv.py using the pre-computed ba-

sis set Map\_BB\_DS-dTSC3.dat and options @err 2 @rep 100. The output .txt file was parsed to extract the  $\alpha$  helix,  $\beta$  strand, and random coil content shown in Fig. 4B, right panel.

### Calculation of ribosome transit time

The raw luminescence vs. time data (Fig. 4C, middle panel) were fit to a straight line in the linear regimes ( $600s < t < 900s$  for 4xSKVF and 4xKVFS,  $900s < t < 1200s$  for 4xSVKF and 4xVKFS) using the R function lm. The intercept term from the fit was used as the transit time of ribosomes across the full transcript and its mean and standard error across technical replicates is shown in Fig. 4C, right panel.

### Calculation of ribosome density around endogenous destabilizing motifs

One canonical transcript for each protein coding gene from GENCODE v32 was identified by filtering for transcripts that are annotated as CCDS or that have a transcript\_support\_level of 1. Duplicate transcripts on the Y chromosome were discarded. After the above filtering, the transcript with the smallest ENSEMBL transcript id was picked for each gene, followed by the transcript with the longest CDS, and finally by the transcript with the smallest CCDS id. The same procedure was repeated for identifying canonical mouse transcripts from GENCODE vM25. The translated CDS of canonical transcripts were scanned for dipeptides formed by all 20 combinations of K, R and L, I, Y, V, F. The location of such dipeptides was given a score of 1, and locations of overlapping dipeptides were given a score of 2, and a score of 0 otherwise. The center of the 16 amino acid window with the highest peptide score was identified using the convolve function from the Python numpy package with the option mode set to same by convolution with a vector of length 16 filled with 1, followed by the argmax function to find the location of the maximum convolution score. A control motif in each transcript was randomly identified among the subset of 16 amino windows that had a peptide score less than 3. The above identification of destabilizing and control motifs is done in find\_destabilizing\_motifs.py.

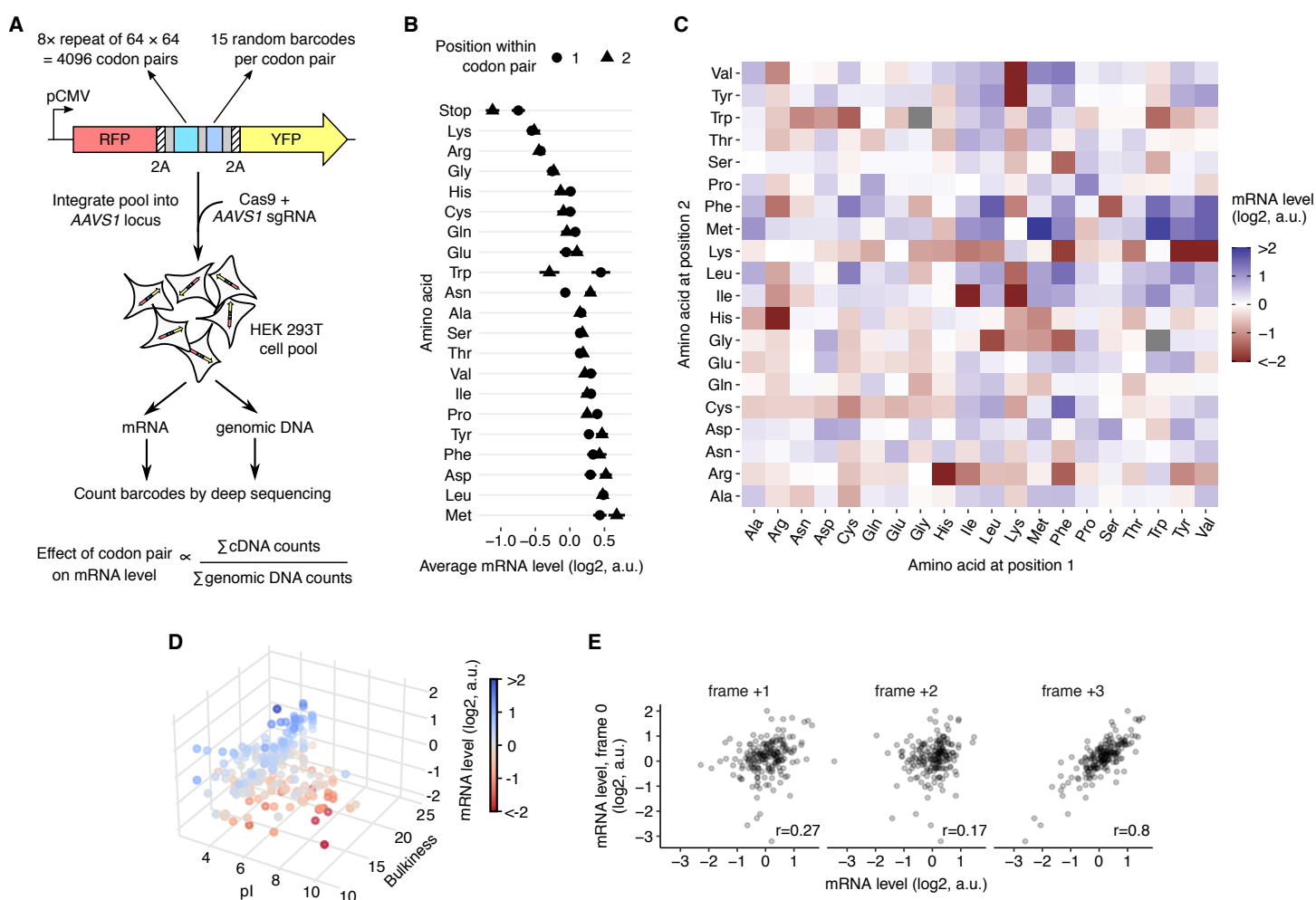
Ribo-seq data from the indicated studies (23, 40, 41) were downloaded from SRA based on GEO accession numbers (GSM39075{97,98}, GSM28281{65,66,67}, GSM16440{76,77,78,79,88,89}), trimmed to remove 5' and 3' adapters using cutadapt (94), filtered to remove contaminating ribosome RNA reads, aligned against the above canonical transcripts using bowtie2 with default parameters, sorted and indexed using samtools. The alignments were trimmed to the P-site as specified in the studies and tallied to calculate coverage at each tran-

script location (`calculate_transcripts_coverage.R`). The list of transcripts was filtered for transcripts that had an average coverage of 1 read per codon in the CDS region. The P-site density around each motif was calculated by normalizing the read counts at each location by the average read density on that transcript (`find_ribosome_density_around_motifs.py`). The normalized P-site densities were averaged at each location around the motif across all destabilizing or control motifs to obtain the meta-gene profiles shown in Figs. [5B](#), [S4A](#), [S4B](#) (`plot_destabilizing_motif_density.Rmd`).

## Statistical analyses

For barcode sequencing, error bars were calculated as the standard deviation of 100 bootstrap samples of barcodes across the gRNA and mRNA samples. The standard deviation was measured for the log2 mRNA levels calculated as described in the *mRNA quantification* section. 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 statistical significant differences were calculated using the `t.test` or `wilcox.test` R functions as appropriate for each figure (see figure captions).

**Figure 1**



### Dipeptide repeats trigger mRNA instability.

**(A)** Schematic of massively parallel array for measuring reporter mRNA levels. 8x repeats of all 4096 codon pairs are synthesized as pooled oligonucleotides, linked in-frame to 24nt random barcodes, and cloned between RFP and YFP reporters with intervening 2A sequences. Each insert has a median of 15 random barcodes without in-frame stop codons. Reporter cassettes are integrated as a pool at the AAVS1 locus in HEK293T cells by Cas9-mediated homologous recombination and constitutively expressed off the CMV promoter. Steady state mRNA level of each insert is determined by sequencing corresponding barcodes in the cDNA and the genomic DNA and normalizing the summed cDNA read counts by the genomic DNA read counts.

**(B)** mRNA level of reporters with codons encoding one of the twenty amino acids or a stop codon in position 1 (circles) or position 2 (triangles) of the 8x codon pair insert shown in A.

**(C)** mRNA level of reporters encoding 400 different dipeptide repeats. Amino acids encoded by the first or second position in the dipeptide are shown along the horizontal or vertical axis respectively. Two dipeptide repeats with missing values are shown in grey.

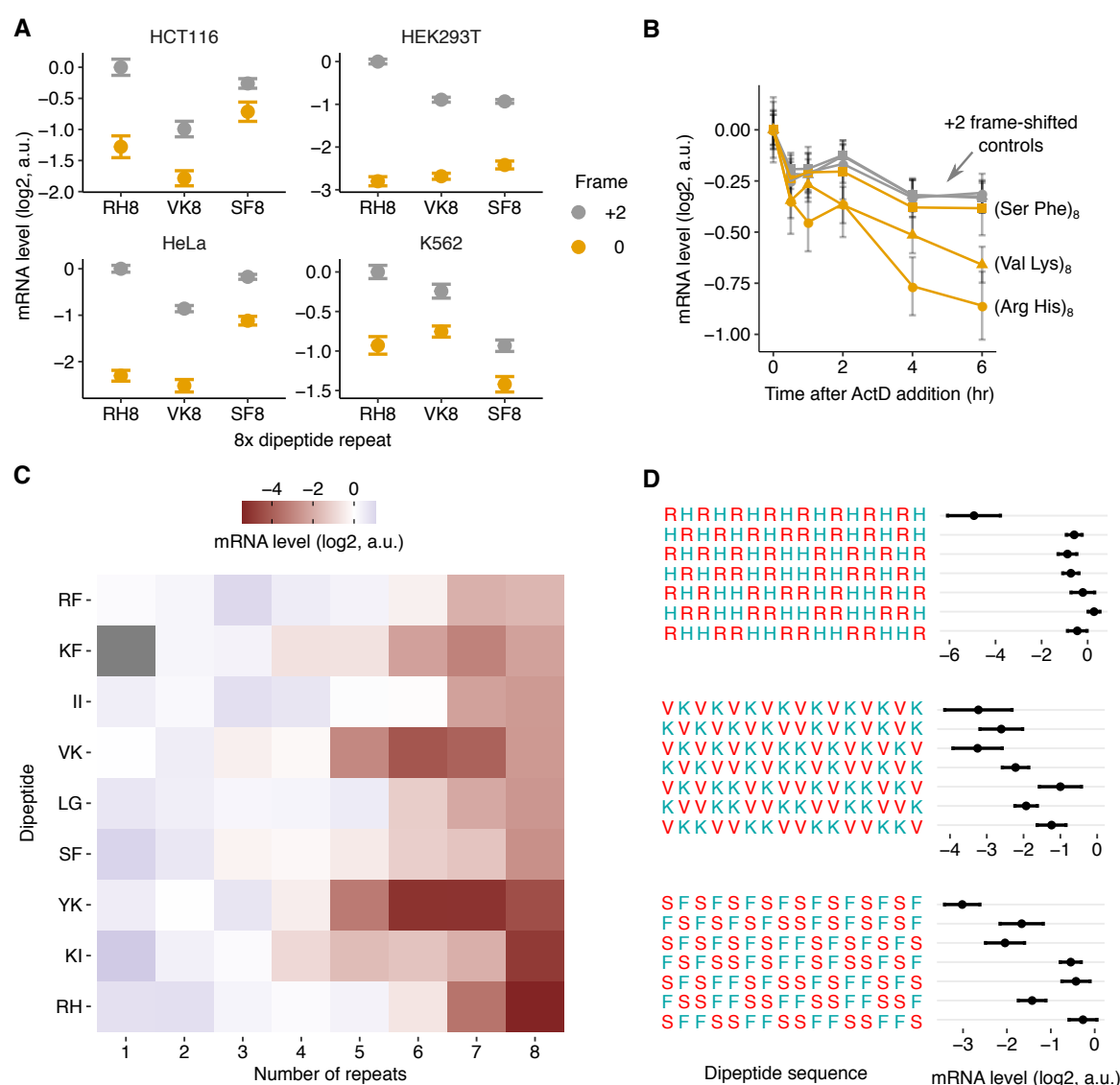
**(D)** mRNA level of dipeptide repeat-encoding reporters plotted as a function of the average isoelectric point (pI) and the bulkiness (32) of the two amino acids in the dipeptide.

**(E)** mRNA level of reporters encoding 190 different dipeptide repeats (excluding reversed repeats) in the correct reading frame (frame 0, vertical axis) or in reading frames shifted by +1, +2, or +3 nucleotides (horizontal axes).  $r$  is the Pearson correlation coefficient between frame 0 and the frame-shifted mRNA levels. mRNA levels in B–E are in arbitrary units (a.u.) and are normalized to the median value across all dipeptide repeats.

Error bars in B represent standard errors of measurement over barcoded replicates calculated by bootstrap. Most error bars in B are smaller than data markers. \*\* and \* indicate one-sided t-test P-values of <0.01 and <0.05 respectively for decrease in mRNA levels at indicated time points of the destabilizing dipeptide repeats with respect to their +2 frameshift controls.



**Figure 2**



# **Nascent peptide primary sequence modulates mRNA stability in human cells**

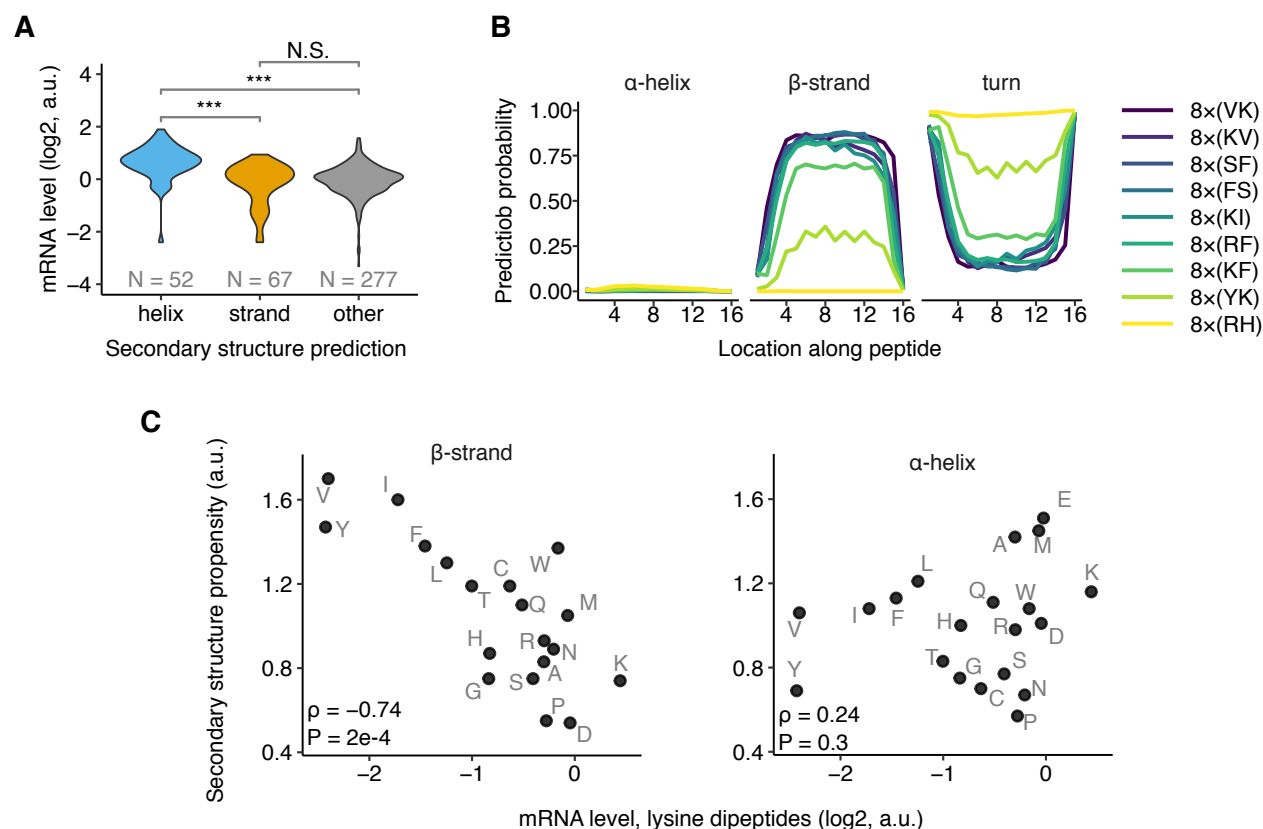
**(A)** mRNA levels of reporters with dipeptide repeats (orange): (Arg His)<sub>8</sub>, (Val Lys)<sub>8</sub>, (Ser Phe)<sub>8</sub>, or the three +2 frameshift controls (grey): (Pro Ser)<sub>8</sub>, (Gln Ser)<sub>8</sub>, (Phe Gln)<sub>8</sub> in 4 different cell lines: HCT116, HEK293T, HeLa, and K562.

**(B)** mRNA stability of reporters from **A** in HEK293T cells. Reporter mRNA levels are measured at indicated time points after Actinomycin D-induced transcriptional shut off. Most points for the three frameshift controls overlap with each other.

**(C)** mRNA levels of dipeptide-encoding reporters with different dipeptide repeat length. Missing value is shown in grey.

**(D)** mRNA levels of dipeptide-encoding reporters with different dipeptide repeat periodicity. Periodicity of dipeptides is systematically reduced by interspersing 1, 2, or 4 tandem repeats of each dipeptide with an equal number of its sequence-reversed counterpart. mRNA levels are measured using the pooled sequencing assay in Fig.1A and normalized by the median value across all inserts in the pool. Amino acids in A, C, and D are labeled by their one-letter codes. Error bars in A, B, and D represent standard errors of measurement over barcoded replicates calculated by bootstrap.

**Figure 3**



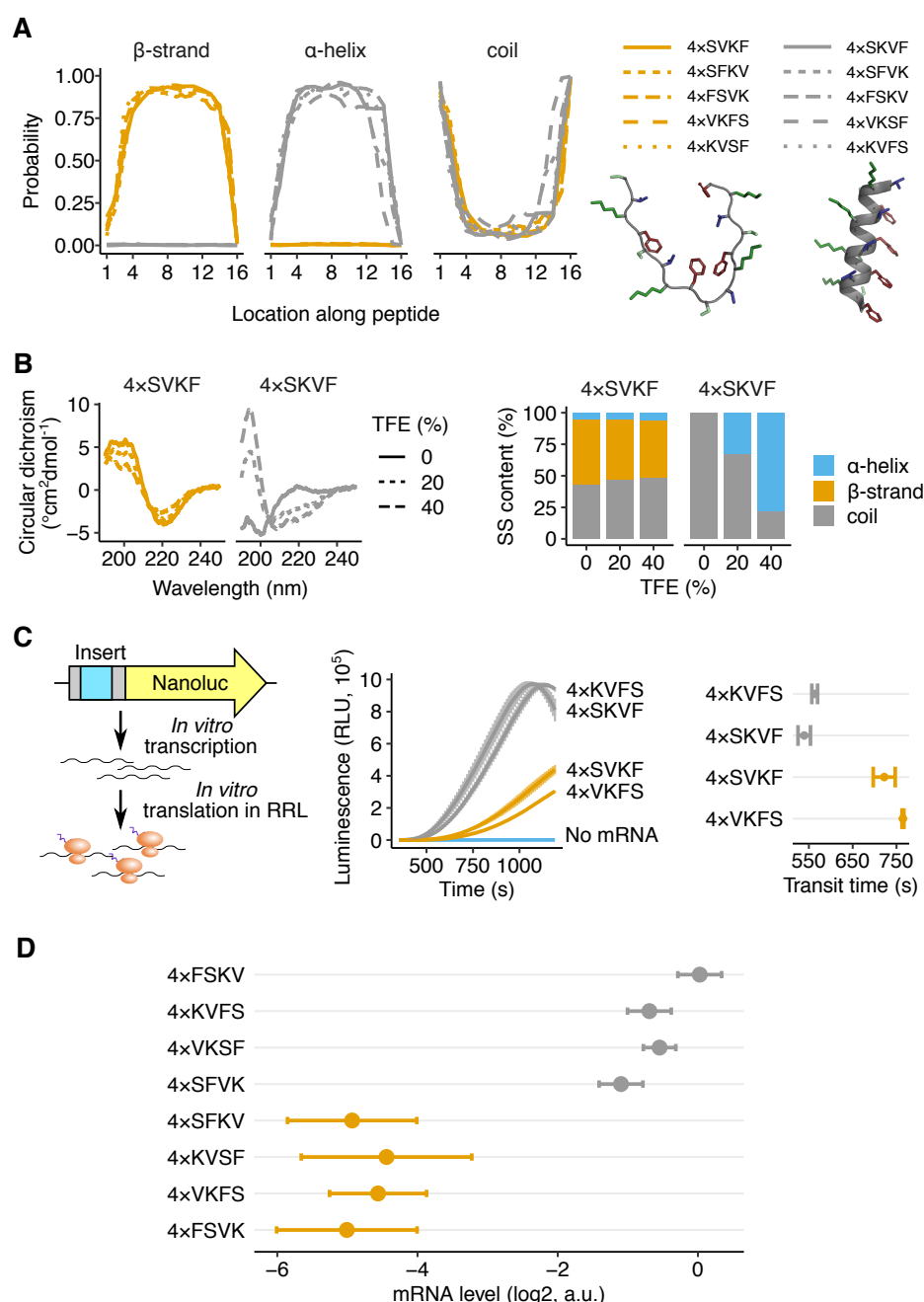
### Secondary structure of dipeptide repeats mediates mRNA instability effects.

**(A)** Distribution of mRNA levels of dipeptide repeat-encoding reporters from Fig. 1C partitioned by predicted protein secondary structure. Per residue probability of secondary structure formation are predicted using S4PRED (35). Inserts with >50% average prediction probability of forming  $\alpha$  helix or  $\beta$  strand are classified as such, or else grouped as 'other.'  $N$  is the number of dipeptide repeats predicted to be in each category. \*\*\*:  $P < 0.001$ , N.S.: not significant (two-sided Mann-Whitney U test).

**(B)** Computationally predicted secondary structure probability along 16 amino acid-long peptide sequences encoded by destabilizing dipeptides. Secondary structure probabilities are predicted using S4PRED.

**(C)** mRNA levels of dipeptide repeat-encoding reporters with lysines in one position and one of twenty amino acids in the other position of the repeat (labeled in grey) shown on horizontal axes. Propensity (36) of the second amino acid to occur in a  $\beta$  strand (left) or an  $\alpha$  helix (right) shown on vertical axes.  $\rho$  is the Spearman correlation coefficient between the two axes with the indicated  $P$  value.

**Figure 4**



#### Extended β strands drive ribosome stalling and mRNA instability.

**(A)** Computationally predicted secondary structure probability along 16 amino acid-long peptide sequences encoded by alternating VK or KV dipeptides with SF or FS dipeptides identified as destabilizing in Fig. 1. Secondary structure probabilities are predicted using S4PRED. The 10 different peptide sequences are 4x repeats of the dipeptide combinations shown in the legend (for example, 4xSVKF: SVKF SVKF SVKF SVKF). Predicted β strand and α helix structures of 4xSVKF and 4xSKVF respectively using PEP-FOLD3 (92) are shown below the legends. The peptide backbones are in grey and the side chains of amino acids are colored.

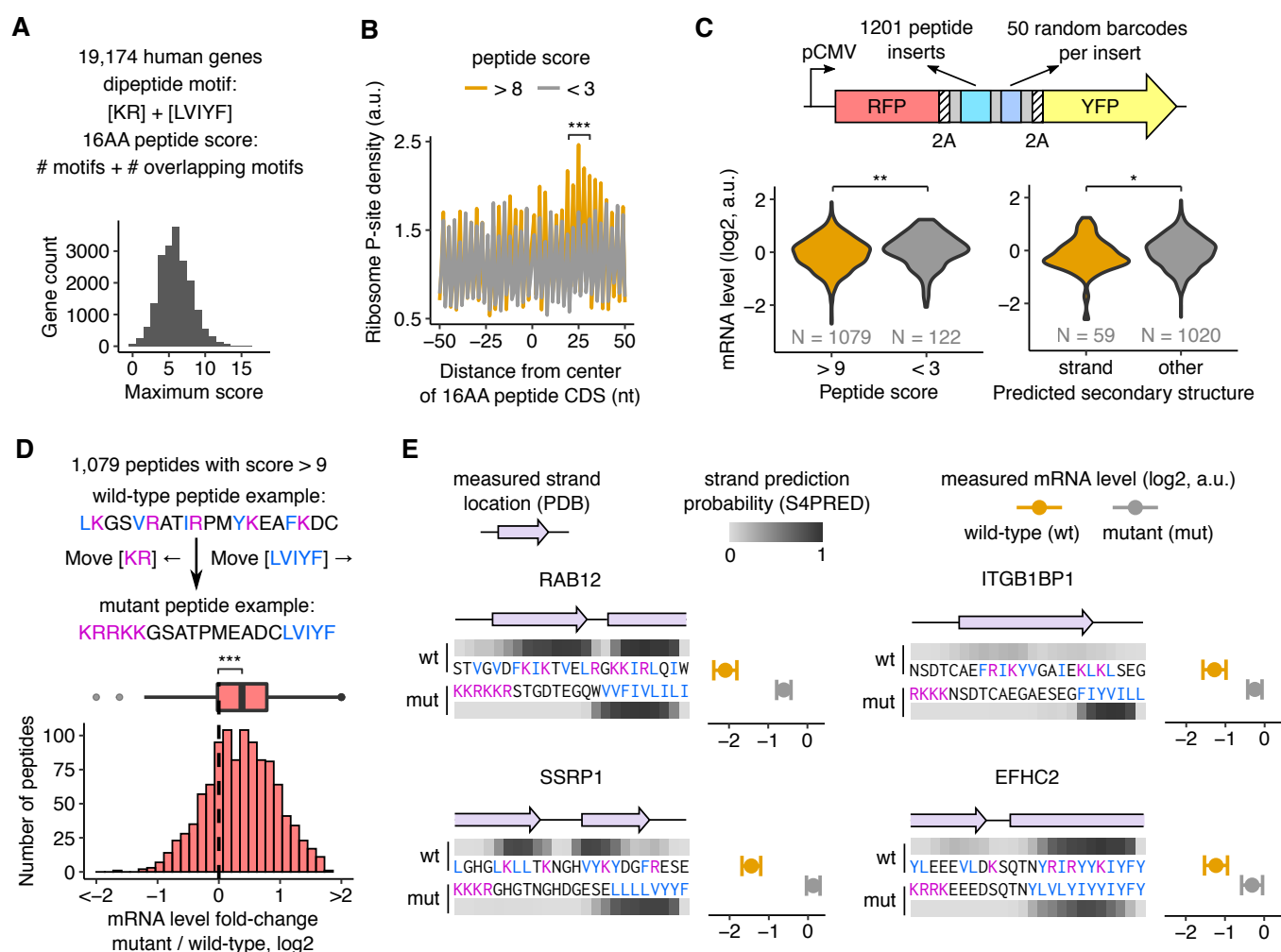
**(B)** Measured circular dichroism spectra of *in vitro* synthesized 4xSVKF or 4xSKVF peptides (left). Measurements are performed with 0, 20, 40% of Trifluoroethanol (TFE) as co-solvent in 10 mM sodium phosphate buffer (pH = 7.5). Relative content of different secondary structures is estimated by linear deconvolution of the measured spectra from a pre-computed basis set using SESCA (93).

**(C)** *In vitro* measurements of ribosome transit time on mRNAs encoding β strand- or α helix-forming peptides followed by Nanoluciferase (left panel). Luminescence is measured as a function of time after addition of *in vitro* transcribed mRNAs to rabbit reticulocyte lysate (RRL) at t=0s (middle panel). Standard error of measurement across three technical replicates is shown as a shaded area on either side of the mean. Ribosome transit times (right panel) are estimated by measuring the X-intercept of the linear portion of the raw luminescence signal in the middle panel.

**(D)** *In vivo* mRNA levels of reporters encoding one of eight different dipeptide combinations. mRNA levels are measured using the reporter constructs and pooled sequencing assay in Fig. 1A. Error bars represent standard errors of measurement over barcoded replicates calculated by bootstrap.



**Figure 5**



# **Dipeptide motifs in the human genome trigger ribosome stalling and mRNA instability.**

**(A)** Scoring methodology for dipeptide motifs in human CDS. Destabilizing dipeptides formed by lysine (K) or arginine (R) with an adjacent leucine (L), valine (V), isoleucine (I), tyrosine (Y), or phenylalanine (F) are given a score of 1. If two such dipeptides overlap, an additional score of 1 is given. The 16 amino acid peptide window with the maximum score is identified in each CDS, and the distribution of these peptide scores across all genes is shown in the lower panel.

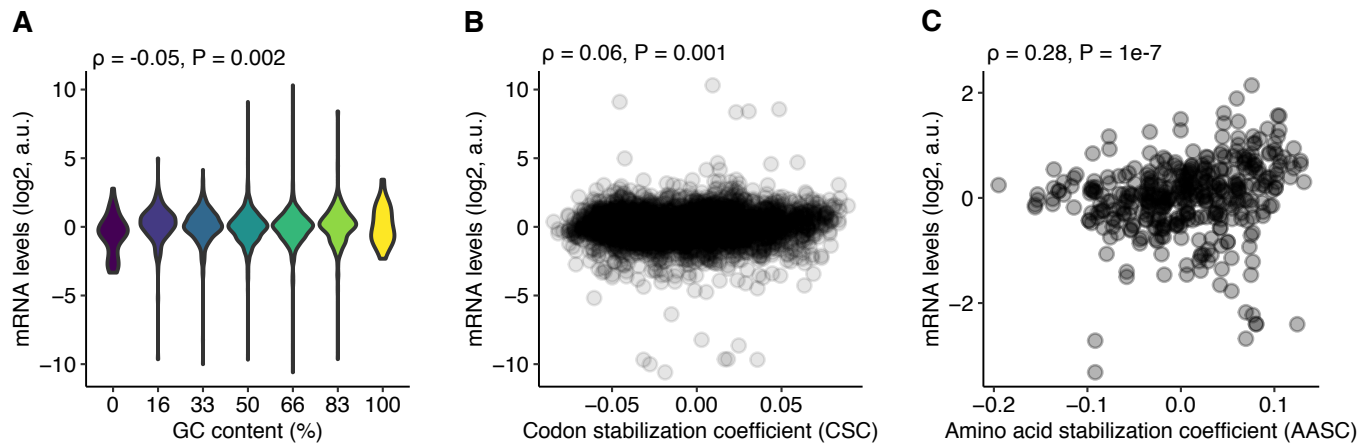
**(B)** Measured ribosome P-site density around destabilizing dipeptide motifs in HEK293T cells (23). The first nucleotide of the 9th codon encoding the 16AA peptide motif is designated as 0nt. \*\*\*:  $P = 1e-5$  (two-sided Mann-Whitney U test) for increase in ribosome P-site density in +20 to +30nt around the destabilizing peptide motif. Comparison is between motifs with peptide score > 8 and randomly chosen control motifs in the same set of genes with peptide score < 3. 485 motifs in genes with an average ribosome P-site density greater than 1 per codon are included in this metagene analysis. Metagene ribosome P-site densities are averaged across all destabilizing or control motifs after normalization by the mean density for each gene.

**(C)** mRNA levels of destabilizing (peptide score > 9) and control motifs (peptide score < 3) are measured by pooled cloning (1,201 total inserts) into a reporter construct followed by deep sequencing as in Fig. 1A. Left panel: Distribution of measured mRNA levels of destabilizing dipeptide motifs compared to control motifs. Right panel: Distribution of measured mRNA levels of destabilizing dipeptide motifs in the left panel partitioned by predicted secondary structure. 59 motifs with an average  $\beta$  strand prediction probability > 0.5 using S4PRED are classified as  $\beta$  strands.

**(D)** Increase in measured mRNA levels upon reordering amino acids in 1,079 endogenous destabilizing dipeptide motifs from C (median  $\Delta\log_2$  mRNA = 0.38). All codons encoding K or R are moved to the 5' end of the mutated motif and codons encoding L, V, I, Y, or F are moved to the 3' end. mRNA levels of motifs are measured using the pooled reporter assay in C. \*\*\*:  $P < 0.001$ , \*\*:  $P < 0.01$ , \*:  $P < 0.05$  using two-sided Mann-Whitney U test in C and D.

**(E)** Examples of endogenous destabilizing motifs with known  $\beta$ -stranded secondary structure. The measured secondary structure of each wild-type motif from PDB is shown as a purple ribbon diagram. Prediction probability for  $\beta$  strands using S4PRED is shown as a grayscale heatmap for wild-type and mutant motifs. Measured mRNA levels of wild-type (orange) and mutant (grey) motifs are shown on the right within each panel. mRNA levels of wild-type and mutant motifs are measured using the pooled reporter assay in C. Error bars represent standard errors of measurement over barcoded replicates calculated by bootstrap.

## Figure S1



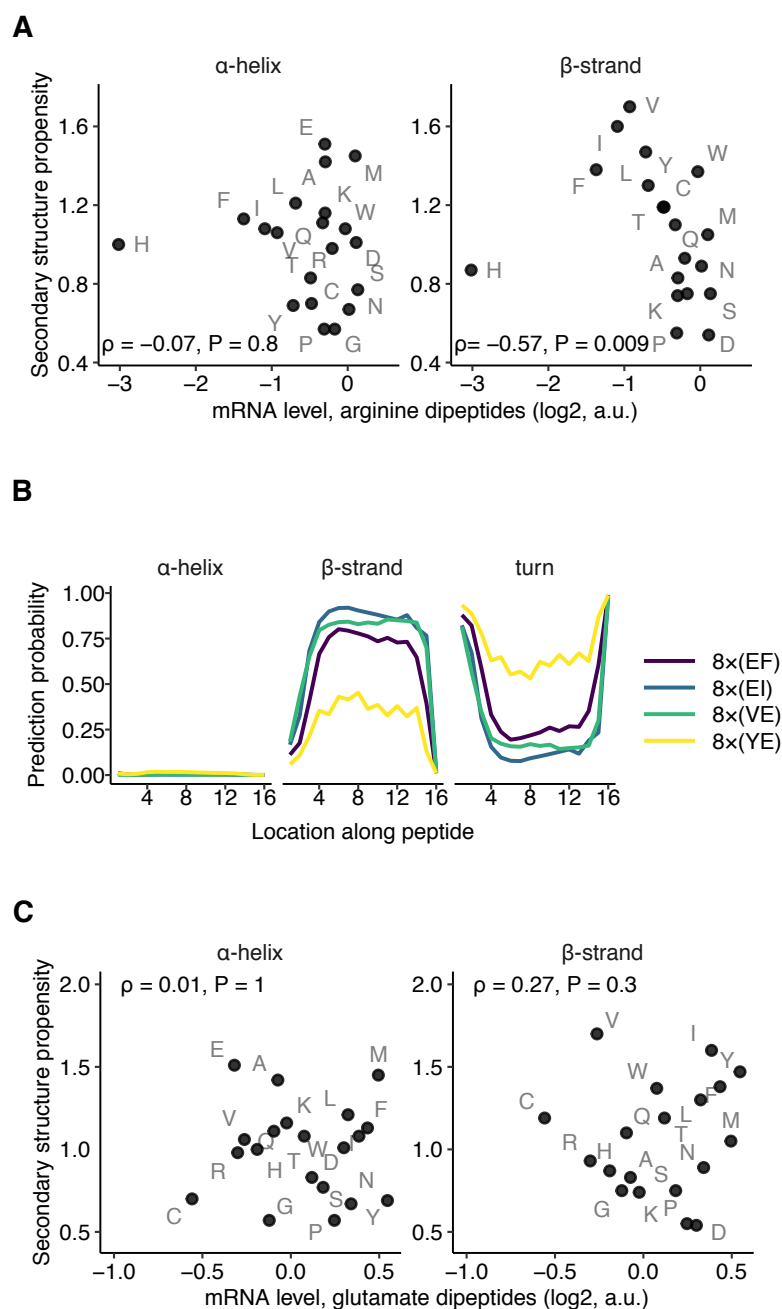
**(A)** mRNA levels of dicodon repeats as a function of their GC content.

**(B)** mRNA levels of dicodon repeats as a function of their mean codon stabilization coefficient (12).

**(C)** mRNA levels of dipeptide repeats as a function of their mean amino acid stabilization coefficient (15).

Spearman correlation coefficient  $\rho$  and its P-value are shown in A-C.

**Figure S2**



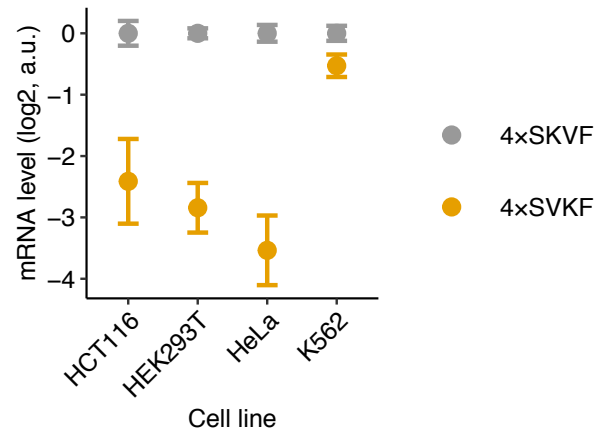
**(A,C)** mRNA levels of dipeptide repeat-encoding reporters with arginine (A) or glutamate (C) in one position and one of twenty amino acids in the other position of the repeat (labeled in grey) plotted as a function of the propensity (36) of the second amino acid to occur in a  $\beta$  strand (right) or an  $\alpha$  helix (left).  $\rho$  is the Spearman correlation coefficient between the two axes with the indicated  $P$  value.

**(B)** Computationally predicted secondary structure probability along 16 amino acid-long peptide sequences encoded by dipeptides with glutamate that form  $\beta$  strands. Secondary structure probabilities are predicted using S4PRED (35).

Amino acids are labeled by their one-letter codes.

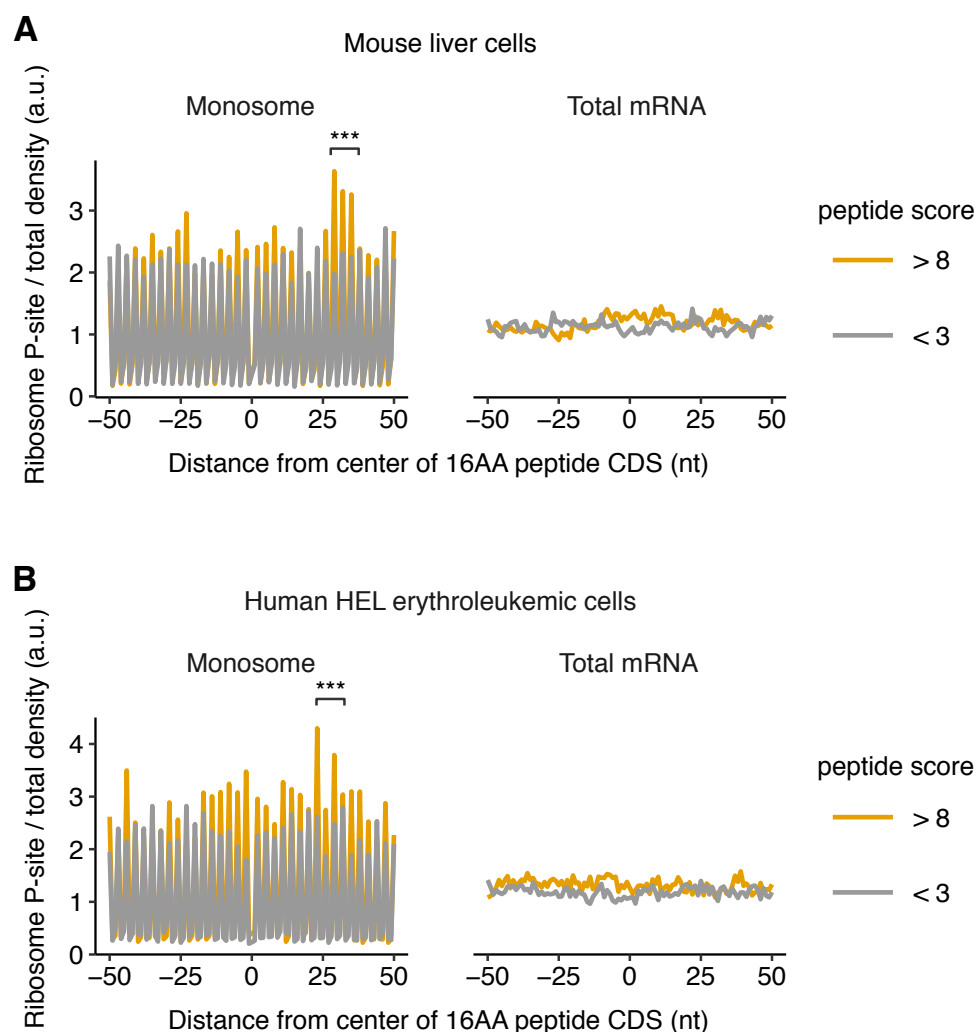


**Figure S3**



mRNA levels of reporters with  $\alpha$  helix (SKVF)<sub>4</sub> or  $\beta$  strand (SVKF)<sub>4</sub> forming peptides in 4 different cell lines HCT116, HEK293T, HeLa, and K562. Error bars represent standard errors of measurement over barcoded replicates calculated by bootstrap.

**Figure S4**



**(A,B)** Measured ribosome P-site density or total RNA density around destabilizing dipeptide motifs in mouse liver cells (41) and human HEL erythroleukemic cells (40). The first nucleotide of the 9th codon encoding the 16AA peptide motif is designated as 0nt. \*\*\*:  $P < 0.001$  (two-sided Mann-Whitney U test) for increase in ribosome P-site density in +25 to +35 (A) or +20 to +30nt (B) around the destabilizing peptide motif. Comparison is between motifs with peptide score  $> 8$  and randomly chosen control motifs in the same set of genes with peptide score  $< 3$ .

## References

1. A. M. Heck, J. Wilusz, The Interplay between the RNA Decay and Translation Machinery in Eukaryotes. *Cold Spring Harb Perspect Biol.* **10**, a032839 (2018).
2. G. Hanson, J. Collier, Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology.* **19**, 20–30 (2018).
3. K. N. D'Orazio, R. Green, Ribosome states signal RNA quality control. *Molecular Cell* (2021), doi:[10.1016/j.molcel.2021.02.022](https://doi.org/10.1016/j.molcel.2021.02.022).
4. M. K. Doma, R. Parker, Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature.* **440**, 561–564 (2006).
5. V. Presnyak, N. Alhusaini, Y.-H. Chen, S. Martin, N. Morris, N. Kline, S. Olson, D. Weinberg, Kristian E. Baker, Brenton R. Graveley, J. Collier, Codon Optimality Is a Major Determinant of mRNA Stability. *Cell.* **160**, 1111–1124 (2015).
6. H. Park, A. R. Subramaniam, Inverted translational control of eukaryotic gene expression by ribosome collisions. *PLOS Biology.* **17**, e3000396 (2019).
7. Y. Matsuo, P. Tesina, S. Nakajima, M. Mizuno, A. Endo, R. Buschauer, J. Cheng, O. Shounai, K. Ikeuchi, Y. Saeki, T. Becker, R. Beckmann, T. Inada, RQT complex dissociates ribosomes collided on endogenous RQC substrate SDD1. *Nature Structural & Molecular Biology*, 1–10 (2020).
8. Y. Mishima, Y. Tomari, Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Molecular Cell.* **61**, 874–885 (2016).
9. A. A. Bazzini, F. del Viso, M. A. Moreno-Mateos, T. G. Johnstone, C. E. Vejnar, Y. Qin, J. Yao, M. K. Khokha, A. J. Giraldez, Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *The EMBO Journal*, e201694699 (2016).
10. J. de Freitas Nascimento, S. Kelly, J. Sunter, M. Carrington, Codon choice directs constitutive mRNA levels in trypanosomes. *eLife.* **7**, e32467 (2018).
11. Y. Harigaya, R. Parker, Analysis of the association between codon optimality and mRNA stability in *Schizosaccharomyces pombe*. *BMC Genomics.* **17**, 895 (2016).
12. Q. Wu, S. G. Medina, G. Kushawah, M. L. DeVore, L. A. Castellano, J. M. Hand, M. Wright, A. A. Bazzini, Translation affects mRNA stability in a codon-dependent manner in human cells. *eLife.* **8**, e45396 (2019).
13. A. Narula, J. Ellis, J. M. Taliaferro, O. S. Rissland, Coding regions affect mRNA stability in human cells. *RNA*, rna.073239.119 (2019).
14. F. Hia, S. F. Yang, Y. Shichino, M. Yoshinaga, Y. Murakawa, A. Vandenbon, A. Fukao, T. Fujiwara, M. Landthaler, T. Natsume, S. Adachi, S. Iwasaki, O. Takeuchi, Codon bias confers stability to human mRNAs. *EMBO reports.* **0**, e48220 (2019).
15. M. E. Forrest, O. Pinkard, S. Martin, T. J. Sweet, G. Hanson, J. Collier, Codon and amino acid content are associated with mRNA stability in mammalian cells. *PLOS ONE.* **15**, e0228730 (2020).
16. R. Weber, M.-Y. Chung, C. Keskeny, U. Zinnall, M. Landthaler, E. Valkov, E. Izaurralde, C. Igreja, 4EHP and GIGYF1/2 Mediate Translation-Coupled Messenger RNA Decay. *Cell Reports.* **33**, 108262 (2020).
17. A. C. Tuck, A. Rankova, A. B. Arpat, L. A. Liechti, D. Hess, V. Ilesmantavicius, V. Castelo-Szekely, D. Gatfield, M. Bühler, Mammalian RNA Decay Pathways Are Highly Specialized and Widely Linked to Translation. *Molecular Cell.* **77**, 1222–1236.e13 (2020).
18. E. Sundaramoorthy, M. Leonard, R. Mak, J. Liao, A. Fulzele, E. J. Bennett, ZNF598 and RACK1 Regulate Mammalian Ribosome-Associated Quality Control Function by Mediating Regulatory 40S Ribosomal Ubiquitylation. *Molecular Cell.* **65**, 751–760.e4 (2017).
19. S. Juszkievicz, R. S. Hegde, Initiation of Quality Control during Poly(A) Translation Requires Site-Specific Ribosome Ubiquitination. *Molecular Cell* (2017), doi:[10.1016/j.molcel.2016.11.039](https://doi.org/10.1016/j.molcel.2016.11.039).
20. S. Juszkievicz, V. Chandrasekaran, Z. Lin, S. Kraatz, V. Ramakrishnan, R. S. Hegde, ZNF598 Is a Quality Control Sensor of Collided Ribosomes. *Mol. Cell.* **72**, 469–481 (2018).
21. S. Juszkievicz, G. Slodkiewicz, Z. Lin, P. Freire-Pritchett, S.-Y. Peak-Chew, R. S. Hegde, Ribosome collisions trigger cis-acting feedback inhibition of translation initiation. *eLife.* **9**, e60038 (2020).
22. N. K. Sinha, A. Ordureau, K. Best, J. A. Saba, B. Zinshteyn, E. Sundaramoorthy, A. Fulzele, D. M. Garshott, T. Denk, M. Thoms, J. A. Paulo, J. W. Harper, E. J. Bennett, R. Beckmann, R. Green, EDF1 coordinates cellular responses to ribosome collisions. *eLife.* **9**, e58828 (2020).
23. P. Han, Y. Shichino, T. Schneider-Poetsch, M. Mito, S. Hashimoto, T. Udagawa, K. Kohno, M. Yoshida, Y. Mishima, T. Inada, S. Iwasaki, Genome-wide Survey of Ribosome Collision. *Cell Reports.* **31**, 107610 (2020).
24. N. Amrani, R. Ganesan, S. Kervestin, D. A. Mangus, S. Ghosh, A. Jacobson, A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature.* **432**, 112–118 (2004).
25. A. B. Eberle, L. Stalder, H. Mathys, R. Z. Orozco, O. Mühlemann, Posttranscriptional Gene Regulation by Spatial Rearrangement of the 3' Untranslated Region. *PLOS Biology.* **6**, e92 (2008).



26. G. Singh, I. Rebbapragada, J. Lykke-Andersen, A Competition between Stimulators and Antagonists of Upf Complex Recruitment Governs Human Nonsense-Mediated mRNA Decay. *PLOS Biology*. **6**, e111 (2008).
27. J. Lu, C. Deutsch, Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. *Journal of Molecular Biology*. **384**, 73–86 (2008).
28. C. A. Charneski, L. D. Hurst, Positively Charged Residues Are the Major Determinants of Ribosomal Velocity. *PLoS Biol.* **11**, e1001508 (2013).
29. R. D. Requião, H. J. A. de Souza, S. Rossetto, T. Domitrovic, F. L. Palhano, Increased ribosome density associated to positively charged residues is evident in ribosome profiling experiments performed in the absence of translation inhibitors. *RNA Biol.* **13**, 561–568 (2016).
30. J. Lu, W. R. Kobertz, C. Deutsch, Mapping the Electrostatic Potential within the Ribosomal Exit Tunnel. *Journal of Molecular Biology*. **371**, 1378–1391 (2007).
31. D. A. Nissley, Q. V. Vu, F. Trovato, N. Ahmed, Y. Jiang, M. Suan Li, E. P. O'Brien, Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling. *J Am Chem Soc.* **142**, 6103–6110 (2020).
32. J. M. Zimmerman, N. Eliezer, R. Simha, The characterization of amino acid sequences in proteins by statistical methods. *Journal of Theoretical Biology*. **21**, 170–201 (1968).
33. S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, M. H. Hecht, Protein design by binary patterning of polar and nonpolar amino acids. *Science*. **262**, 1680–1685 (1993).
34. H. Xiong, B. L. Buckwalter, H. M. Shieh, M. H. Hecht, Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *PNAS*. **92**, 6349–6353 (1995).
35. L. Moffat, D. T. Jones, Increasing the Accuracy of Single Sequence Prediction Methods Using a Deep Semi-Supervised Learning Framework. *Bioinformatics* (2021), doi:10.1093/bioinformatics/btab491.
36. P. Y. Chou, G. D. Fasman, Empirical Predictions of Protein Conformation. *Annu. Rev. Biochem.* **47**, 251–276 (1978).
37. P. Luo, R. L. Baldwin, Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry*. **36**, 8413–8421 (1997).
38. A. Jasanoff, A. R. Fersht, Quantitative determination of helical propensities from trifluoroethanol titration curves. *Biochemistry*. **33**, 2129–2135 (1994).
39. M. H. Kolar, G. Nagy, J. Kunkel, S. M. Vaiana, L. V. Bock, H. Grubmüller, Folding of VemP into translation-arresting secondary structure is driven by the ribosome exit tunnel. *bioRxiv*, 2021.04.15.440051 (2021).
40. D. A. Kuppers, S. Arora, Y. Lim, A. R. Lim, L. M. Carter, P. D. Corrin, C. L. Plaisier, R. Basom, J. J. Delrow, S. Wang, H. H. He, B. Torok-Storb, A. C. Hsieh, P. J. Paddison, N 6 -methyladenosine mRNA marking promotes selective translation of regulons required for human erythropoiesis. *Nat Commun.* **10**, 1–17 (2019).
41. P. Janich, A. B. Arpat, V. Castelo-Szekely, M. Lopes, D. Gattfield, Ribosome profiling reveals the rhythmic liver transcriptome and circadian clock regulation by upstream open reading frames. *Genome Res.* **25**, 1848–1859 (2015).
42. A. M. Watkins, P. S. Arora, Anatomy of  $\beta$ -Strands at Protein-Protein Interfaces. *ACS Chem Biol.* **9**, 1747–1754 (2014).
43. M. Mizuno, S. Ebine, O. Shounai, S. Nakajima, S. Tomomatsu, K. Ikeuchi, Y. Matsuo, T. Inada, The nascent polypeptide in the 60S subunit determines the Rqc2-dependency of ribosomal quality control. *Nucleic Acids Research* (2021), doi:10.1093/nar/gkab005.
44. H. Ramu, N. Vázquez-Laslop, D. Klepacki, Q. Dai, J. Piccirilli, R. Micura, A. S. Mankin, Nascent Peptide in the Ribosome Exit Tunnel Affects Functional Properties of the A-Site of the Peptidyl Transferase Center. *Molecular Cell*. **41**, 321–330 (2011).
45. J. Lu, Z. Hua, W. R. Kobertz, C. Deutsch, Nascent Peptide Side Chains Induce Rearrangements in Distinct Locations of the Ribosomal Tunnel. *Journal of Molecular Biology*. **411**, 499–510 (2011).
46. P. Po, E. Delaney, H. Gamper, D. M. Szantai-Kis, L. Speight, L. Tu, A. Kosolapov, E. J. Petersson, Y.-M. Hou, C. Deutsch, Effect of Nascent Peptide Steric Bulk on Elongation Kinetics in the Ribosome Exit Tunnel. *Journal of Molecular Biology*. **429**, 1873–1888 (2017).
47. W. Li, F. R. Ward, K. F. McClure, S. T.-L. Chang, E. Montabana, S. Liras, R. G. Dullea, J. H. D. Cate, Structural basis for selective stalling of human ribosome nascent chain complexes by a drug-like molecule. *Nature Structural & Molecular Biology*. **26**, 501–509 (2019).
48. R. Sabi, T. Tuller, Computational analysis of nascent peptides that induce ribosome stalling and their proteomic distribution in *Saccharomyces cerevisiae*. *RNA*. **23**, 983–994 (2017).
49. A. L. Parola, B. K. Kobilka, The peptide product of a 5' leader cistron in the beta 2 adrenergic receptor mRNA inhibits receptor synthesis. *Journal of Biological Chemistry*. **269**, 4497–4505 (1994).

50. K. Reynolds, A. M. Zimmer, A. Zimmer, Regulation of RAR beta 2 mRNA expression: evidence for an inhibitory peptide encoded in the 5'-untranslated region. *Journal of Cell Biology*. **134**, 827–835 (1996).
51. V. Shanmuganathan, N. Schiller, A. Magoulopoulou, J. Cheng, K. Braunger, F. Cymer, O. Berninghausen, B. Beatrix, K. Kohno, G. von Heijne, R. Beckmann, Structural and mutational analysis of the ribosome-arresting human XBP1u. *eLife*. **8**, e46267 (2019).
52. S. Matheisl, O. Berninghausen, T. Becker, R. Beckmann, Structure of a human translation termination complex. *Nucleic Acids Res*. **43**, 8615–8626 (2015).
53. S. Bhushan, T. Hoffmann, B. Seidelt, J. Frauenfeld, T. Mielke, O. Berninghausen, D. N. Wilson, R. Beckmann, SecM-Stalled Ribosomes Adopt an Altered Geometry at the Peptidyl Transferase Center. *PLOS Biology*. **9**, e1000581 (2011).
54. B. Seidelt, C. A. Innis, D. N. Wilson, M. Gartmann, J.-P. Armache, E. Villa, L. G. Trabuco, T. Becker, T. Mielke, K. Schulten, T. A. Steitz, R. Beckmann, Structural Insight into Nascent Polypeptide Chain-Mediated Translational Stalling. *Science*. **326**, 1412–1415 (2009).
55. D. N. Wilson, S. Arenz, R. Beckmann, Translation regulation via nascent polypeptide-mediated ribosome stalling. *Current Opinion in Structural Biology*. **37**, 123–133 (2016).
56. T. Su, J. Cheng, D. Sohmen, R. Hedman, O. Berninghausen, G. von Heijne, D. N. Wilson, R. Beckmann, The force-sensing peptide VemP employs extreme compaction and secondary structure formation to induce ribosomal stalling. *eLife*. **6**, e25642 (2017).
57. B. Hardesty, G. Kramer, in *Progress in Nucleic Acid Research and Molecular Biology* (Academic Press, 2000; <https://www.sciencedirect.com/science/article/pii/S0079660300660269>), vol. 66, pp. 41–66.
58. C. A. Woolhead, A. E. Johnson, H. D. Bernstein, Translation Arrest Requires Two-Way Communication between a Nascent Polypeptide and the Ribosome. *Molecular Cell*. **22**, 587–598 (2006).
59. J. Lu, C. Deutsch, Secondary Structure Formation of a Transmembrane Segment in Kv Channels. *Biochemistry*. **44**, 8230–8243 (2005).
60. M.-N. Yap, H. D. Bernstein, The Plasticity of a Translation Arrest Motif Yields Insights into Nascent Polypeptide Recognition inside the Ribosome Tunnel. *Molecular Cell*. **34**, 201–211 (2009).
61. K. N. D'Orazio, C. C.-C. Wu, N. Sinha, R. Loll-Kripplleber, G. W. Brown, R. Green, The endonuclease Cue2 cleaves mRNAs at stalled ribosomes during No Go Decay. *eLife*. **8**, e49117 (2019).
62. M. L. Glover, A. M. Burroughs, P. C. Monem, T. A. Egelhofer, M. N. Pule, L. Aravind, J. A. Arribere, NONU-1 Encodes a Conserved Endonuclease Required for mRNA Translation Surveillance. *Cell Rep*. **30**, 4321–4331.e4 (2020).
63. R. Buschauer, Y. Matsuo, T. Sugiyama, Y.-H. Chen, N. Alhusaini, T. Sweet, K. Ikeuchi, J. Cheng, Y. Matsuki, R. Nobuta, A. Gilmozzi, O. Berninghausen, P. Tesina, T. Becker, J. Collier, T. Inada, R. Beckmann, The Ccr4-Not complex monitors the translating ribosome for codon optimality. *Science*. **368** (2020), doi:10.1126/science.aay6912.
64. J. Cao, A. P. Geballe, Mutational Analysis of the Translational Signal in the Human Cytomegalovirus gpUL4 (gp48) Transcript Leader by Retroviral Infection. *Virology*. **205**, 151–160 (1994).
65. K. Yanagitani, Y. Kimata, H. Kadokura, K. Kohno, Translational Pausing Ensures Membrane Targeting and Cytoplasmic Splicing of XBP1u mRNA. *Science*. **331**, 586–589 (2011).
66. V. Chandrasekaran, S. Juszkievicz, J. Choi, J. D. Puglisi, A. Brown, S. Shao, V. Ramakrishnan, R. S. Hegde, Mechanism of ribosome stalling during translation of a poly(A) tail. *Nature Structural & Molecular Biology*. **26**, 1132–1140 (2019).
67. D. H. Goldman, N. M. Livingston, J. Movsik, B. Wu, R. Green, Live-cell imaging reveals kinetic determinants of quality control triggered by ribosome stalling. *Molecular Cell* (2021), doi:10.1016/j.molcel.2021.01.029.
68. K. L. Hickey, K. Dickson, J. Z. Cogan, J. M. Replogle, M. Schoof, K. N. D'Orazio, N. K. Sinha, J. A. Hussmann, M. Jost, A. Frost, R. Green, J. S. Weissman, K. K. Kostova, GIGYF2 and 4EHP Inhibit Translation Initiation of Defective Messenger RNAs to Assist Ribosome-Associated Quality Control. *Molecular Cell*. **79**, 950–962.e6 (2020).
69. N. G. Lintner, K. F. McClure, D. Petersen, A. T. Londregan, D. W. Piotrowski, L. Wei, J. Xiao, M. Bolt, P. M. Loria, B. Maguire, K. F. Geoghegan, A. Huang, T. Rolph, S. Liras, J. A. Doudna, R. G. Dullea, J. H. D. Cate, Selective stalling of human translation through small-molecule engagement of the ribosome nascent chain. *PLOS Biology*. **15**, e2001882 (2017).
70. I. P. Ivanov, B.-S. Shin, G. Loughran, I. Tzani, S. K. Young-Baird, C. Cao, J. F. Atkins, T. E. Dever, Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. *Molecular Cell*. **70**, 254–264 (2018).
71. F. Cymer, G. von Heijne, Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *PNAS*. **110**, 14640–14645 (2013).

72. Ola B. Nilsson, R. Hedman, J. Marino, S. Wickles, L. Bischoff, M. Johansson, A. Müller-Lucks, F. Trovato, Joseph D. Puglisi, Edward P. O'Brien, R. Beckmann, G. von Heijne, Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Reports*. **12**, 1533–1540 (2015).
73. N. Ismail, R. Hedman, N. Schiller, G. von Heijne, A biphasic pulling force acts on transmembrane helices during translocation-mediated membrane integration. *Nat Struct Mol Biol*. **19**, 1018–1022 (2012).
74. Andrey L. Karamyshev, Anna E. Patrick, Zemfira N. Karamysheva, Dustin S. Griesemer, H. Hudson, S. Tjon-Kon-Sang, I. Nilsson, H. Otto, Q. Liu, S. Rospert, G. von Heijne, Arthur E. Johnson, Philip J. Thomas, Inefficient SRP Interaction with a Nascent Chain Triggers a mRNA Quality Control Pathway. *Cell*. **156**, 146–157 (2014).
75. A. Shiber, K. Döring, U. Friedrich, K. Klann, D. Merker, M. Zedan, F. Tippmann, G. Kramer, B. Bukau, Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature*, 1 (2018).
76. M. Bertolini, K. Fenzl, I. Kats, F. Wruck, F. Tippmann, J. Schmitt, J. J. Auburger, S. Tans, B. Bukau, G. Kramer, Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science*. **371**, 57–64 (2021).
77. H. E. Stanger, F. A. Syud, J. F. Espinosa, I. Girit, T. Muir, S. H. Gellman, Length-dependent stability and strand length limits in antiparallel  $\beta$ -sheet secondary structure. *PNAS*. **98**, 12015–12020 (2001).
78. J. S. Richardson, D. C. Richardson, Natural  $\beta$ -sheet proteins use negative design to avoid edge-to-edge aggregation. *PNAS*. **99**, 2754–2759 (2002).
79. S. Pechmann, J. Frydman, Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nature Structural & Molecular Biology* (2012), doi:10.1038/nsmb.2466.
80. J. L. Chaney, A. Steele, R. Carmichael, A. Rodriguez, A. T. Specht, K. Ngo, J. Li, S. Emrich, P. L. Clark, Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLOS Computational Biology*. **13**, e1005531 (2017).
81. M. Dalvai, J. Loehr, K. Jacquet, C. C. Huard, C. Roques, P. Herst, J. Côté, Y. Doyon, A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. *Cell Rep*. **13**, 621–633 (2015).
82. J. P. Fonseca, A. R. Bonny, G. R. Kumar, A. H. Ng, J. Town, Q. C. Wu, E. Aslankoochi, S. Y. Chen, G. Dods, P. Harrigan, L. C. Osimiri, A. L. Kistler, H. El-Samad, A Toolkit for Rapid Modular Construction of Biological Circuits in Mammalian Cells. *ACS Synth. Biol*. **8**, 2593–2606 (2019).
83. A. M. Darnell, A. R. Subramaniam, E. K. O'Shea, Translational Control through Differential Ribosome Pausing during Amino Acid Limitation in Mammalian Cells. *Molecular Cell*. **71**, 229–243.e11 (2018).
84. V. T. Chu, T. Weber, B. Wefers, W. Wurst, S. Sander, K. Rajewsky, R. Kühn, Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol*. **33**, 543–548 (2015).
85. R. Muller, Z. A. Meacham, L. Ferguson, N. T. Ingolia, CIBER-seq dissects genetic networks by quantitative CRISPRi profiling of expression phenotypes. *Science*. **370** (2020), doi:10.1126/science.abb9662.
86. N. J. McGlincy, Z. A. Meacham, K. K. Reynaud, R. Muller, R. Baum, N. T. Ingolia, A genome-scale CRISPR interference guide library enables comprehensive phenotypic profiling in yeast. *BMC Genomics*. **22**, 205 (2021).
87. D. Susorov, S. Egri, A. A. Korostelev, Termi-Luc: a versatile assay to monitor full-protein release from ribosomes. *RNA*. **26**, 2044–2050 (2020).
88. J. Köster, S. Rahmann, Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. **28**, 2520–2522 (2012).
89. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods*. **9**, 357–359 (2012).
90. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
91. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch, in *The Proteomics Protocols Handbook*, J. M. Walker, Ed. (Humana Press, Totowa, NJ, 2005; <https://doi.org/10.1385/1-59259-890-0:571>), *Springer Protocols Handbooks*, pp. 571–607.
92. A. Lamiab, P. Thévenet, J. Rey, M. Vavrusa, P. Derreumaux, P. Tufféry, PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Research*. **44**, W449–W454 (2016).
93. G. Nagy, M. Igaev, N. C. Jones, S. V. Hoffmann, H. Grubmüller, SESCA: Predicting Circular Dichroism Spectra from Protein Molecular Structures. *J. Chem. Theory Comput*. **15**, 5087–5102 (2019).
94. M. Martin, *EMBnet.journal*, in press, doi:10.14806/ej.17.1.200.