Cell

Supplemental Information

Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity

Danny D. Nedialkova and Sebastian A. Leidel

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Yeast and Nematode Strains

All yeast strains (Table S4) were in the S288c BY4741 background (Winzeler et al, 1999). Single-gene knockout strains were generated by sporulation of strains from the BY4743 heterozygous diploid gene deletion collection (Thermo Scientific) and tetrad dissection. Gene disruption cassettes conferring kanamycin resistance were generally exchanged with auxotrophic markers to circumvent cross-resistance to paromomycin. Double- and triple-knockout strains were created by mating and sporulation. Gene knockouts were verified by the absence of mRNA and ribosome footprint reads from targeted genes and/or PCR. *C. elegans* strains were in the Bristol N2 wild-type background (Brenner, 1974).

Culture Conditions for Stress Treatment in Yeast

Transcript levels change only transiently following environmental perturbations in yeast and the magnitude of abundance change is proportional to the severity of the stress (Gasch et al., 2000). Therefore, to maximally capture the translation of response transcripts upon exposure to diamide or rapamycin, we chose a 30-min treatment, which corresponds to the peak of transcript abundance changes (Gasch et al., 2000; Hardwick et al., 1999). The amount of each drug was titrated to fully preserve cell viability at the time of analyses and to elicit similar growth delays in cultures of wild-type yeast at later time points. Based on this analysis, 1.5 mM diamide and 12.5 nM rapamycin were selected for subsequent experiments. Paromomycin was added to exponentially growing cultures at 1 mg/ml for 30 min.

Ribosome Profiling of Yeast and Nematodes

Elongating ribosomes were stabilized by a brief (60-sec) treatment of yeast cultures with 100 µg/ml CHX, and cells were harvested by rapid vacuum filtration (~45 sec) and flash-freezing. Yeast or L4 stage nematodes were pulverized under cryogenic conditions in 20 mM Tris-HCl (pH=7.5), 100 mM NaCl, 10 mM $MgCl_2$, 1% Triton X100, 0.5 mM DTT, 100 µg/ml CHX at 5 cps in a SPEX 6750 Freezer/Mill (SPEX SamplePrep). Lysates were clarified by centrifugation at 4°C, 3,000 g for 3 min followed by an additional centrifugation step at 4°C, 10,000 g for 5 min. Extracts from yeast or nematodes were treated with 11.25 U or 7.5 U RNase I (Ambion) per OD_{260} unit of extract for 60 min at 22°C with continuous mixing. After addition of 100 U SUPERase In (Ambion), ribosomes were resolved on 7%-47% sucrose density gradients in 50 mM Tris-HCl (pH=7.5), 50 mM NH_4Cl, 12 mM $MgCl_2$, 0.5 mM DTT, 100 µg/ml CHX for 3 h at 35,000 rpm, 4°C in a TH-641 rotor (Thermo Scientific). Gradients were fractionated by upward displacement at 0.75 ml/min with continuous monitoring of OD_{254} values. Fractions were collected at 32-sec intervals, and SDS was immediately added to 1% in monosome fractions, which were flash-frozen and stored at -80°C. RNA was isolated from monosome samples by the hot acid phenol method. Ribosome footprints were purified by separation of monosome RNA on 15% polyacrylamide, 8M urea, 1×TBE gels, staining with SYBR Gold (Life Technologies) and excision of bands between RNA oligonucleotide markers of 28 nt and 32 nt (yeast) or 28 nt and 30 nt (nematode).

Total RNA Extraction from Yeast and Poly(A) RNA Selection

A fraction of yeast cultures used for ribosome profiling was kept for total RNA extraction. Cells collected by rapid vacuum filtration were resuspended in ice-cold 50 mM NaOAc (pH=5.5), 10 mM EDTA, followed by addition of SDS to 1% and homogenization in a Precellys 24 bead beater (Bertin Technologies) at power setting 6.5 for two cycles of 20 sec. Lysates were clarified by centrifugation at 4°C, 10,000 g for 5 min, treated with 100 µg/ml Proteinase K for 20 min at 60°C, and total RNA was extracted with acid phenol. One hundred µg of RNA were treated with TURBO DNase (Ambion) and poly(A) RNA was purified by two rounds of selection with Poly(A)Purist MAG Kit (Ambion) according to the manufacturer's instructions. Purified RNA was fragmented by alkaline hydrolysis in 50 mM sodium bicarbonate (pH=9.2), 1 mM EDTA for 20 min at 95°C. Fragmented RNA was purified by ethanol precipitation and fragments of 50-80 nt were recovered from 15% polyacrylamide, 8M urea, 1×TBE gels stained with SYBR Gold.

Total RNA Extraction from *C. elegans*

Age-synchronized wild-type N2 and *tut-1(tm1297), elpc-1(tm2149)* mutant animals grown at 20°C were collected by filtration at day one of adulthood, resuspended in a fourfold excess of Tri Reagent (Sigma Aldrich), and flash-frozen. After six cycles of thawing in a 42°C waterbath and freezing in liquid nitrogen, total RNA was extracted using standard protocols and treated with TURBO DNase.

Generation of Deep Sequencing Libraries

Sequencing libraries from ribosome-protected footprints or randomly fragmented poly(A)-selected RNA were generated essentially as described in (Ingolia et al., 2012). 3'-adapter ligation was performed for 4 h at 22°C in a reaction containing 200 U of T4 RNA ligase 2 (truncated, NEB), 25% PEG 8000, and 10 U SUPERase In. Because rRNA contaminants in footprint libraries frequently originate from A-rich sequences (Brar et al., 2012; Guydosh and Green, 2014), we chose not to subtract those by antisense oligonucleotide hybridization to avoid unspecific depletion of footprints that could potentially skew the occupancy measurements of A-rich codons. Sequencing libraries of poly(A)-enriched RNA from adult N2 and *tut-1(tm1297), elpc-1(tm2149)* nematodes were generated with the TruSeq RNA Sample Prep Kit v2 (Illumina).

Sequencing Data Analysis

Sequencing data from ribosome footprints was pre-processed by trimming linker sequences from the 3' end of reads and alignment to rRNA genes with Bowtie 1.0.0 (Langmead et al., 2009). Non-rRNA reads longer than 25 nt were mapped with Bowtie to reference sets consisting of ORFs in the sacCer3 genome annotated as verified or uncharacterized in the Saccharomyces Genome Database on Jan 07, 2013 (5,906 sequences) or UCSC canonical transcripts (Hsu et al., 2006) of the *C. elegans* WS190 (ce6) genome (20,051 sequences). In both cases, 18 nt of 5' UTR and 3' UTR sequences of each ORF were included to allow alignment of ribosome footprints from initiating and terminating ribosomes. One mismatch in the seed sequence was allowed during mapping (except for *C. elegans* mRNA-Seq data, where two mismatches were allowed). Only reads with unique matches to the reference were retained for further analysis. The majority of footprint reads aligning to coding sequences had a length between 29 and 31 nt, within the range of previous observations (Ingolia et al., 2009; 2011). The total reads in each ribosome footprint library were 10 – 36 million (yeast) and 23 – 44 million (nematode). Of those, 5 – 20 million and 2 - 14 million mapped uniquely to coding regions, with the exception of libraries from paromomycin-treated yeast, which had lower depth (1 – 2 million uniquely mapping reads).

For global codon occupancy measurements, only reads with a perfect match to the reference were used. The offset of P site codons from the 5’ ends of footprints was inferred by examining the cumulative distribution of 29-31 nt reads aligning at annotated CDS start codons, which are positioned in the P site of initiating ribosomes (Ingolia et al., 2009; Jackson et al., 2010). In agreement with previous reports (Ingolia et al., 2009; Stadler and Fire, 2011), we found that footprint coverage begins abruptly 12 to 13 nt upstream of start codons in both *S. cerevisiae* and *C. elegans* (Figure S1A). For high-precision assignment of P site codons, we therefore selected reads with a 5’-nt in the 0 and -1 frames of canonical coding sequences, discarded the first nucleotide of reads beginning in the -1 frame, and used an offset of 12 nt from the 5’ end of each read to infer the P site codon (positions 13-15, Figure S1B). Since ribosome density in the beginning of ORFs can be significantly affected by growth conditions, as well as ongoing translation initiation or ribosomal run-off during sample preparation (Gerashchenko and Gladyshev, 2015; Ingolia et al., 2009; 2011), we excluded footprint reads mapping within the first 15 codons of coding sequences from our analyses. The frequency of each codon in different sites within footprints was then calculated. To obtain the basal occurrence of each codon, we computed its average frequency in the three non-tRNA binding positions downstream of the A site (+1 to +3, Figure S1B). These positions have not yet been translated and codon identity, therefore, should have negligible effects on ribosome occupancy. Codon representation in E, P, and A sites was then computed by dividing the frequency of a codon in each of these sites by its basal occurrence, an approach similar to the one described in (Stadler and Fire, 2011). Using this metric, we were able to detect known ribosome pausing events, such as those associated with the presence of proline codons within the P site in both mouse ES cells and *S. cerevisiae* (Ingolia et al., 2011; Zinshteyn and Gilbert, 2013), as well as when codons translated via G:U wobble occupied the P site in *C. elegans* (Stadler and Fire, 2011) (data not shown).

For comparing A-site codon occupancy at individual positions, the single-codon occupancy metric described in (Zinshteyn and Gilbert, 2013) was used. Briefly, the A-site codon in reads used for global codon occupancy measurements was assigned according to the criteria outlined above, and read counts at a given codon were normalized to the average per-codon read density in the ORF containing it. The first 15 codons of each ORF and codons with no reads were excluded from this analysis.

For gene-level measurements, differential ribosome occupancy (excluding the first 15 codons) and mRNA abundance were tested with DESeq, which uses a model based on the negative binomial distribution (Anders and Huber, 2010). Statistical significance was defined by a Benjamini-corrected *p*-value of less than 0.01, and low coverage genes were filtered out by requiring a mean value of normalized reads between two samples being compared of at least 50. Gene Ontology (GO) term enrichment was analyzed with the functional annotation tool of the DAVID bioinformatics web server (http://david.abcc.ncifcrf.gov/). GO terms with a Benjamini-corrected *p*-value of less than 0.01 were summarized and visualized in semantic similarity-based scatterplots using the REViGO web server (http://revigo.irb.hr/, Supek et al., 2011).

Isolation of Endogenous Protein Aggregates from Yeast

Protein aggregates were isolated as in (Koplin et al., 2010). Wild-type and knockout yeast strains were grown to mid-log phase (OD_{600} ~0.5) at 30°C and cells from 100 ml of culture were harvested by rapid vacuum filtration and flash-frozen. After resuspension in extraction buffer (20 mM NaPi pH=6.8, 1 mM EDTA, 10 mM DTT, 0.1% Tween 20) containing a protease inhibitor cocktail (0.5 mM AEBSF, 10 µg/ml aprotinin, 0.5 mg/ml benzamidine, 20 µM leupeptin, 5 µM pepstatin A), 60 U of zymolyase T20 (Zymo Research) were added, and the cell suspension was incubated at 22°C for 20 min with continuous mixing. Extracts were chilled on ice and sonicated eight times at level 4, duty cycle 50% with a Branson tip sonicator. Cell debris was sedimented by centrifugation at 200 g, 4°C for 20 min and protein concentration in supernatants was determined with the Bio-Rad Protein Assay. Protein concentration was equalized across samples and aggregates were pelleted from equal amounts of total protein by centrifugation at 16,000 g, 4°C for 20 min. Pellets were resuspended in washing buffer (20 mM NaPi pH=6.8, 2% NP-40, protease inhibitor cocktail) by sonicating six times at level 4, duty cycle 50%, and sedimented again at 16,000 g, 4°C for 20 min. After a second wash in buffer B, aggregates were resuspended in detergent-free washing buffer by sonicating 4 times at level 2, duty cycle 65%. Following sedimentation, pellets were dissolved in Laemmli sample buffer with 100 mM DTT and 8M urea by boiling at 95°C for 10 min. Total protein extracts and aggregates were separated on 4 -12% NuPAGE Bis-Tris gels (Life Technologies) in MES-SDS running buffer and gels were stained with Colloidal Blue Staining Kit (Life Technologies).

Quantitative Protein Mass Spectrometry

Isolated aggregates and matched total protein samples were size-separated by SDS-PAGE and subjected to in-gel trypsin digestion (Shevchenko et al., 2007). Extracted peptides were desalted using Empore-C18 StageTips (Rappsilber et al., 2003) and stored at 4°C until further use. Prior to LC-MS/MS, peptides were eluted using 2 x 20 µl of 80% ACN, 0.5% acetic acid, dried in an Eppendorf concentrator to a volume of about 2 µl, and resuspended in 10 µl Buffer A (0.5% acetic acid). 6 µl of this peptide solution were subsequently analyzed by nanoscale reversed-phase chromatography using a EASY nLC 1000 UHPLC on a PepMap C18 EASY-Spray™ column (15 cm x 75 µm ID, 3 µm particle diameter; Thermo Scientific; column temperature set to 45°C) that was online coupled via an EASY-Spray™ ESI source (Thermo Scientific) to a Q Exactive mass spectrometer (Thermo Scientific). Peptides were separated at a flow rate of 250 nl/min using a multistep gradient (3 – 25% B in 60 min; 25 – 32% B in 10 min; 32 – 95% B in 5 min; hold at 95% B for 5 min) before re-equilibration at starting conditions. The mass spectrometer was operated in data-dependent mode, acquiring full scan spectra at a resolution of 70.000 and an AGC target value of 3e6 (scan range 300 - 1750 m/z). The ten most intense ions were chosen for higher energy collisional dissociation (HCD) with a resolution of 17.500 at m/z 400 and a target value of 1e6. Dynamic exclusion was allowed and set to 20 sec, uncharged as well as singly charged compounds were excluded from the analysis. Data were recorded with Xcalibur (Thermo Scientific).

Raw MS files were processed using the MaxQuant computional platform (version 1.4.1.2) (Cox and Mann, 2008). Identification of peptides and proteins was enabled by the built-in Andromeda search engine by querying the concatenated forward and reverse yeast Uniprot database, release 2013-02, including common contaminants. The allowed initial mass deviation was set to 7 ppm and 20 ppm in the search for precursor and fragment ions, respectively. Trypsin with full enzyme specificity and only peptides with a minimum length of 7 amino acids were selected, and a maximum of two missed cleavages was allowed. Carbamidomethylation (Cys) was set as fixed modification, while oxidation (Met) and N-acetylation were defined as variable modifications. Protein and peptide identification was performed at a false discovery rate (FDR) of 1%.

Relative label free quantification was based on the measurements of three independent biological replicates for each strain analyzed by MaxQuant with the ‘match between runs’ option turned on. Data processing and annotation were performed using the Perseus module of MaxQuant (version 1.4.0.6). Reverse and contaminant hits, as well as those identified by a modified site only, were filtered out. At least one unique peptide was required for protein identification. Intensity values were logarithmized and missing values were replaced by imputation, simulating signals of low abundant proteins within the distribution of measured values using a width of 0.3 SD and a downshift of 1.8 SD. Statistical significance was determined by two-sample *t* tests with a permutation-based FDR cutoff of 5%. For aggregate samples, proteins identified in *ncs2Δelp6Δ* and *ssb1Δssb2Δ* samples were considered aggregated only if enriched at least 2-fold over wild type. A small number of paralogous proteins that passed these criteria differ by as little as one residue and could not be distinguished in our MS data despite high sequence coverage. In such cases, we included both paralogs in the set of aggregated proteins and denoted their enrichment value as “undefined”.

C. elegans RNAi and Lifespan Analysis

For RNAi experiments, *E. coli* HT115(DE3) carrying the double-stranded-RNA-expressing vector L4440 were seeded onto NGM plates containing 2mM IPTG and 50 µg/ml carbenicillin, and dsRNA expression was induced overnight at room temperature. Nematodes expressing a CAG-encoded polyQ stretch fused to YFP (Q35-YFP, Morley et al., 2002) were synchronized by egg lay on RNAi plates and grown at 20°C. Animals were transferred to fresh plates every 24 to 48 h and Q35-YFP aggregates in body wall muscle cells was imaged in one-day-old adults (four days after egg lay) on a Zeiss Axio Imager 2 microscope. RNAi treatment was unknown to the researcher during aggregate quantification.

Lifespan analysis was performed at 20°C on *E. coli* OP50. Nematodes were age-synchronized by egg lay and the young adult stage was defined as day 0. One hundred animals were used per condition and were scored for viability every day or every second day. Animals that had crawled off the plate or undergone internal hatching were censored. Statistical analysis was performed with GraphPad Prism using the log-rank (Mantel–Cox) method.

SUPPLEMENTAL REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science *335*, 552–557.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics, *77*(1), 71–94.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Gerashchenko, M.V., and Gladyshev, V.N. (2015). Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res *42*, e134.

Guydosh, N.R., and Green, R. (2014). Dom34 rescues ribosomes in 3' untranslated regions. Cell *156*, 950–962.

Hardwick, J.S., Kuruvilla, F.G., Tong, J.K., Shamji, A.F., and Schreiber, S.L. (1999). Rapamycin-modulated transcription defines the subset of nutrient-sensitive signaling pathways directly controlled by the Tor proteins. Proc. Natl. Acad. Sci. USA *96*, 14866–14870.

Hsu, F., Kent, W.J., Clawson, H., Kuhn, R.M., Diekhans, M., and Haussler, D. (2006). The UCSC Known Genes. Bioinformatics *22*, 1036–1046.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell. Biol. *11*, 113–127.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. Anal. Chem. *75*, 663–670.

Shevchenko, A., Tomas, H., Havli sbreve, J., Olsen, J.V., and Mann, M. (2007). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature Prot. *1*, 2856–2860.

Supek, F., Bošnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE *6*, e21800.