

Systematic effects of mRNA secondary structure on gene expression and molecular function in budding yeast

Xia Wang^{1,2}

Email: xiawang@email.arizona.edu

Pidong Li³

Email: lipid@nwfufu.edu.cn

Ryan N. Gutenkunst^{1*}

*Corresponding author

Email: rgutenk@email.arizona.edu.

¹ Department of Molecular and Cellular Biology, University of Arizona, 1007 East Lowell Street, Tucson, Arizona 85721, USA

² School of Medical Engineering, Hefei University of Technology, Hefei, Anhui Province 230009, China

³ College of Life Sciences, Northwest A&F University, Yangling, Shaanxi Province 712100, China

Abstract

Dynamic control of gene expression is crucial for cellular adaptation to environmental challenges. mRNA secondary structure is known to be associated with mRNA and protein abundance, but little is known about how mRNA secondary structure affects gene expression dynamics. We report a genome-wide computational analysis of mRNA secondary structure, codon usage, and gene expression in budding yeast. We show that mRNA secondary structure combined with codon optimality regulates gene expression in multiple ways, from transcription to mRNA stability to translation. Moreover, we find that the effect of mRNA secondary structure on mRNA abundance is primarily mediated by transcription, not mRNA stability. Notably, genes with low mRNA secondary structure were substantially enriched for functions relevant to stress response, acting in the mitochondrion, endoplasmic reticulum, and ribosome. On the other hand, genes with high mRNA secondary structure were enriched for functions relevant to cellular maintenance, including macromolecular metabolism and biosynthesis. Our results suggest that mRNA secondary structure affects gene expression through coordination of multiple stages in protein biogenesis, with important consequences for stress response. The coupling of transcription to mRNA stability to translation makes concerted changes in mRNA and protein abundance possible and may amplify the effect of regulation to make quick responses to environmental variations.

Keywords: PARS; stress response; codon usage

Background

Gene expression is finely and dynamically regulated through control of each step in the Central Dogma [1]. For example, recent genome-wide studies have revealed that transcription and mRNA decay rates are coupled to regulate gene expression [2-4]. In steady state, the balance between mRNA transcription and degradation determines mRNA abundance, and translation and protein degradation then impact protein abundance [5, 6]. Gene expression must also be dynamically regulated to adapt to environmental challenges, and transcription rate, translation efficiency, and decay rate limit the speed at which gene expression can change [7]. Abundant evidence shows that multiple steps in gene expression are often coordinated to shape the dynamics of the transcriptome and proteome in response to environmental perturbations [2, 4, 6, 8]. The mechanism of this coordination is, however, elusive.

Increasing evidence suggests that folding into precise secondary structures is necessary for mRNA maturation. For example, RNA secondary structures have been proposed to affect splice-site recognition [9-13] and poly(A) site recognition [12, 14]. Specific structural elements also affect the stability of many eukaryotic mRNAs [15]. In addition, RNA secondary structure elements [16, 17], such as bulge loops, hairpin loops, and stems, mediate protein-RNA interactions that play integral roles in various post-transcriptional regulatory processes, including the degradation and translation of RNA [18, 19]. RNA secondary structure thus influences nearly every step in gene expression [20-22]. Sequence features, such as codon usage, have also been proposed to affect transcription [23-25], mRNA decay [24, 26] and translation efficiency [24, 27, 28]. RNA structure is affected by codon usage [29], so RNA structure may mediate the effects of codon usage on gene expression. Recent work has shown that steady-state RNA and protein abundance are strongly correlated with RNA structure and sequence features [30, 31], and other work has delineated the effects of mRNA thermodynamic stability, sequence features, and tRNA availability on translation [32]. Nevertheless, the systematic effects of RNA structure on gene expression remain unexplored.

Recently, a high-throughput experimental approach has been developed to assay *in vitro* mRNA secondary structure. In Parallel Analysis of RNA Structure (PARS), the poly-adenylated transcript pool is treated with RNase V1 that preferentially cleaves 3' phosphodiester bonds in double-stranded RNA and, separately, with RNase S1 that preferentially cleaves 3' bonds in single-stranded RNA. The obtained fragments are then deep sequenced to measure the degree to which each nucleotide was in a single- or double-stranded conformation. The PARS score is defined as the \log_2 of the ratio between the number of times the nucleotide immediately downstream of a given nucleotide was observed as the first base when treated with RNase V1 and when treated with RNase S1 [33]. Techniques exist to measure *in vivo* mRNA secondary structure [12, 17, 34], which depends both on the sequence of the mRNA and on any bound proteins or RNAs. We used PARS to measure *in vitro* mRNA secondary structure to focus on the intrinsic folding of the mRNA, independent of the local cellular environment.

Large-scale manipulative experiments on RNA structure and sequence features are challenging, but the statistical technique of path analysis provides a means to disentangle the effects of multiple potential causal factors on multiple steps in gene expression from observational data [35]. Path analysis was developed to analyze complex systems in which it is challenging to isolate causes and effects, because each component potentially affects many others through a network of direct and indirect interactions. In path analysis the total correlation between any two variables is decomposed into direct effects of one on the other, indirect effects mediated by other variables, and spurious correlations due to common causes. The estimated path coefficients represent the amount of change expected in the dependent variable as a result of a unit change in the independent variable. Path analysis has been used to infer causal relationships from observational molecular data in a number of research settings, including studies of human disease [36], plant physiology [37], and protein evolution [38].

In this work, we used path analysis to decompose the correlations between mRNA structure and sequence features on transcriptional and translational processes. We then explored the functional consequences of the associations between mRNA structure and different steps of gene expression, identifying cellular functions that are enriched in genes with low or high mRNA secondary structure.

We found that mRNA structure and optimal codon usage are significantly and positively associated with transcription rate, mRNA half-life, and translation efficiency. Surprisingly, we found that mRNA secondary structure affects transcript abundance primarily by affecting transcription rather than mRNA decay. Moreover, this relationship is not affected by optimal codon usage, although codon usage plays important roles in regulation of transcription, mRNA decay, and translation. We found that genes with high mRNA secondary structure were enriched for functions in cell maintenance, whereas genes with low mRNA secondary structure were enriched for assembly of cellular components relevant to stress response. Together, our results suggest that cells may use mRNA secondary structure, coupled with sequence features, to quickly regulate protein homeostasis and function in responses to changes in environmental conditions.

Materials and Methods

Yeast genomic data

mRNA sequences of *Saccharomyces cerevisiae* were retrieved from the Saccharomyces Genome Database [52].

Experimentally determined mRNA secondary structure in yeast

Parallel analysis of RNA structure (PARS) scores for 3000 genes in yeast were taken from Kertesz et al. [33].

Yeast mRNA levels and mRNA half-life

mRNA expression levels for 5083 genes were obtained from Pelechano et al. [43] where mRNA amount were derived from the RNA-Seq data [53], which had been normalized to real units (mRNAs/cell) using the 2000 most abundant mRNAs from the GATC-PCR data [54]. mRNA half-life data for 3888 genes were obtained from

Presnyak et al. [26].

Estimation of transcription rates and RNA polymerase II density in yeast

Nascent transcription rate data for 4666 genes obtained from an averaged Genomic Run-on dataset and RNA polymerase II density data for the same genes were taken from Pelechano et al. [43].

Protein abundance and protein half-life in yeast

Protein abundance data for 6067 genes for *S. cerevisiae* were downloaded from PaxDb [55]. Protein half-life data for 3796 genes were obtained from Christiano et al. [56].

Translation efficiency calculation in yeast

Translation efficiency (TE) was calculated as the ratio of ribosome-protected fragments (RPF) count to mRNA abundance. TE data for 5218 genes in normal condition was retrieved from McManus et al. [57]. For our analysis of stress response, TE was calculated from the data of Gerashchenko et al. [58], which report changes in RPF and mRNA abundance for 5623 genes in response to 30 minutes of oxidative stress, compared to the normal condition.

Prediction of RNA folding energy

The minimum free energies of mRNA sequences were calculated using the Matlab rnafold function (Matlab Bioinformatics toolbox). To measure the mean local folding of a genomic sequence, we computed and averaged the folding energy of sliding windows with a length of 30 nucleotides (nt) and a step size of 10 nt; 30 is the approximate length of the ribosomes' footprint [59].

Evolutionary rates

Evolutionary rates (dN, dS, and dN/dS) of genes were taken from Wall et al. [60].

Codon usage and GC content calculation

CodonW (<http://codonw.sourceforge.net/>) was used to estimate Codon Adaptation Index (CAI) and GC content.

Gene ontology (GO) enrichment analysis

GO enrichment analysis was performed with DAVID [61]. Fisher's exact test was used to test for overrepresentation of each GO term among lowly or highly folded genes.

Statistical analysis

Spearman's rank correlation (or partial correlation) coefficients were used to quantify relationships between variables. Standard errors of correlation coefficients were estimated using 10,000 bootstrap data samples, and Z-tests were used to test for statistical significance of individual effects or differences between effects.

Path analysis

Path analysis was used to investigate the contribution of PARS (or folding free energy) and CAI to Central Dogma processes. We used this path analysis to estimate the total, direct, and indirect effects of each variable on each other. For example, the indirect effect between PARS and mRNA abundance is mediated through transcription rate, CAI, and mRNA half-life. The analysis was performed using the R package sem [62].

Results

mRNA structured is correlated with codon usage

It has been reported that mRNA secondary structure, as measured by PARS score, is strongly associated with optimal codon usage, averaging over the gene [29]. Consistent with this report, we also found that the average PARS score over a gene is strongly positively correlated with that gene's codon adaptation index (CAI) (Spearman rank correlation $r=0.49$, $N=2925$, $p<10^{-177}$), a widely used index for quantifying codon usage bias [39].

To infer more specifically which codons are most strongly associated with mRNA structure, we separately analyzed all 64 codons. We found that the frequencies of most codons (49/64) are significantly correlated with gene-average PARS score ($p<0.01$), even after controlling for mRNA abundance and GC content (Table S1). In particular, 6 codons (ACC, GCC, GCU, GGU, GUC, UCC) are strongly positively correlated with PARS score ($r>0.3$, $p<10^{-5}$), and 6 codons (AAA, AGG, AUA, CGA, GGA, GUA) are strongly negatively correlated with PARS score ($r<-0.3$, $p<10^{-5}$) (Figure 1A). Codons that are positively correlated with PARS tend to be common codons with high translation efficiency [39, 40], whereas codons that are negatively correlated tend to be rare codons with low translation efficiency (Table S1). For example, both AUA and AUC code for isoleucine, and the correlation between frequency of AUA (which has low relative synonymous codon usage [39]) and PARS is $r=-0.433$ and between frequency of AUC (which has high relative synonymous codon usage [39]) and PARS is $r=0.25$. The association of mRNA secondary structure with high translation efficiency codons may act to smooth rates of translation, reducing the risk of potentially deleterious slow or fast ribosome movement [41].

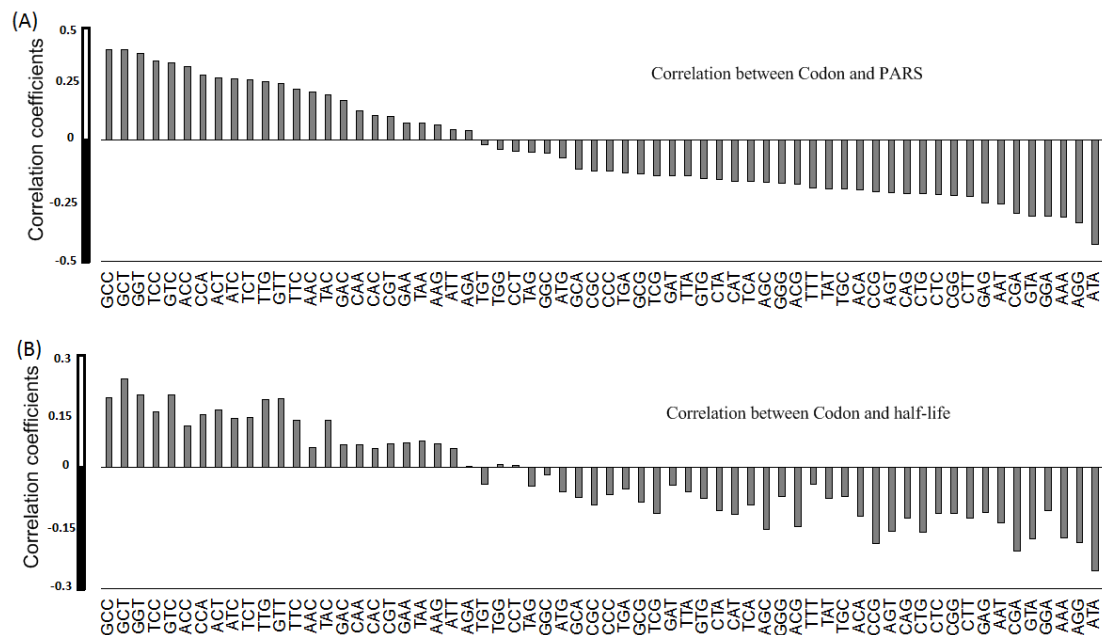


Figure 1. Rank correlations between frequencies within genes of all 64 codons and (A) mRNA secondary structure, measured by average PARS within each gene or (B) mRNA half-life.

In addition, the correlations between PARS score and evolutionary rates (i.e., dN, dS, and dN/dS) are negative and strong, even when we control for mRNA abundance ($r=-0.29$ for dS, $p<0.001$), suggesting that mRNAs with stronger secondary structure will evolve more slowly, potentially due to selection to preserve structure. This result is consistent with recent work of Park et al. [42], which demonstrated that mRNA folding is associated with slow evolution of highly expressed proteins.

Contributions of mRNA secondary structure and codon bias to mRNA abundance

Previous work [30] has shown that gene-average PARS score is strongly and positively associated with mRNA abundance, an observation we corroborate (Spearman rank correlation $r=0.45$, $N=2610$, $p<10^{-127}$; Figure S1). It is unknown, however, to what degree this correlation is driven by transcription rate and/or mRNA half-life.

To disentangle the effects of mRNA structure and other factors on gene expression, we used path analysis, a general technique for quantifying the directed acyclic dependencies among a set of variables. Path analysis begins with an assumed model, representing potential causal influences between the variables in the data. Multiple regression of the data on that model then reveals the strength of different causal links between variables. Within a given model, path analysis can thus identify the most important causal paths (those with the largest path coefficients), thus identifying the most important factors in determining the output of a complex system. Our model (Figure 2) began with the Central Dogma steps in gene expression. To those, we added potential direct influences from RNA structure (quantified as PARS) on transcription [9-12, 14], mRNA half-life [15] and translation [18, 19]. We also added potential direct influences from RNA sequence features (quantified by CAI) on transcription [23-25], mRNA half-life [24, 26], and translation [24, 27, 28]. We modeled PARS and CAI as exogenous variables, which may be correlated with each other [29] but are not causally determined by other variables in the model. The other variables, including mRNA half-life, translation, transcription, mRNA abundance, and protein abundance, are endogenous variables that depend directly or indirectly on PARS and CAI.

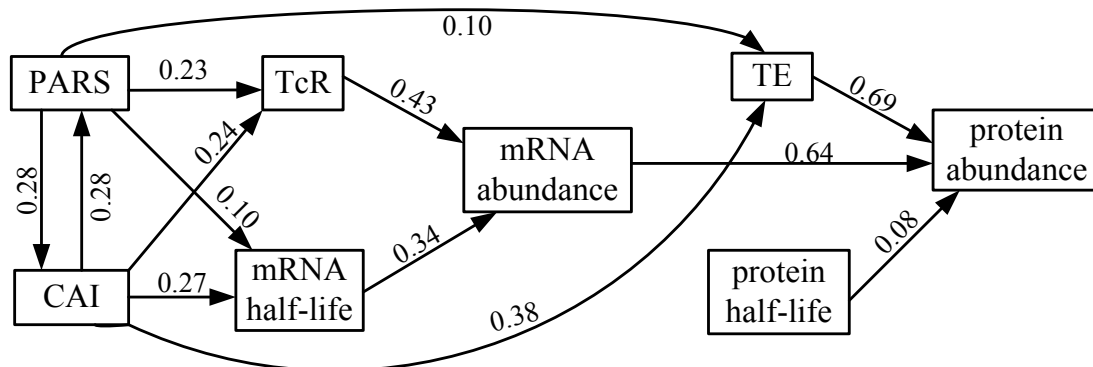


Figure 2. Path diagram presenting coefficients for all direct effects among mRNA secondary structure (PARS), codon adaptation index (CAI), transcription rate (TcR),

mRNA half-life, mRNA abundance, translation efficiency (TE), protein half-life, and protein abundance. The potential causal relations tested between the predictor variables are based on previous investigations of relationships among secondary structure, codon bias, and transcription [9-12, 14, 23, 25], mRNA decay [15, 24, 26] and translation [18, 19, 24, 27, 28].

We applied our model to 1,344 genes for which we had data on all variables considered. We found that PARS has a larger direct effect on transcription ($\beta=0.23$) than on mRNA half-life ($\beta=0.10$; Z-test, $p<0.01$) (Table S2). Notably, the effect of PARS on mRNA abundance mediated by transcription ($\beta = 0.23 \times 0.43 = 0.099$) is significantly larger than that mediated by mRNA half-life ($\beta = 0.10 \times 0.34 = 0.034$) (Z-test, $p<10^{-4}$). The effect of RNA structure on mRNA abundance is thus primarily driven by changes in transcription rather than mRNA half-life.

PARS is measured *in vitro*, and RNA may refold prior to analysis, which might bias our results. We thus repeated our analysis using an independent measure of RNA folding, by using a thermodynamic model to predict the minimum free energy (MFE) of each folded RNA [42]. Similar to our analysis with PARS, we found that the association between MFE and mRNA abundance mediated by transcription ($\beta = 0.21 \times 0.47 = 0.099$) is much stronger than the association mediated by mRNA half-life ($\beta = 0.12 \times 0.3 = 0.036$) (Z-test, $p<10^{-4}$) (Figure S2). This provides further evidence that RNA structure affects mRNA abundance primarily through transcription and not mRNA half-life.

In contrast to PARS, CAI has similar direct effects on transcription rate (TcR) and mRNA half-life (0.24 vs. 0.27; Z-test, $p=0.37$). Notably, the direct effects of PARS and CAI on TcR are similar (0.23 vs. 0.24; Z-test, $p=0.43$) and much stronger than any indirect effects, suggesting that both RNA structure and codon content are important to transcription rate. By contrast, the effect of CAI on mRNA half-life is significantly larger than that of mRNA structure (0.24 vs. 0.10; Z-test, $p<0.01$). The 12 codons (ACC, GCC, GCU, GGU, GUC, UCC, AAA, AGG, AUA, CGA, GGA, GUA) most strongly correlated with PARS also have moderate correlation with mRNA stability (Figure 1B). Codon usage may thus mediate the association between PARS score and mRNA half-life. Indeed, path analysis showed that the effect of PARS on mRNA half-life as mediated by CAI ($\beta = 0.28 \times 0.27 = 0.076$) is comparable to the direct effect of PARS on mRNA half-life ($\beta = 0.10$; Z-test, $p=0.36$).

Together, these analyses suggest that the correlation between mRNA structure (PARS) and mRNA abundance is driven primarily by effects of mRNA structure on transcription rate, rather than mRNA half-life. By contrast, codon usage has similar effects on transcription rate and mRNA half-life.

Contributions of mRNA secondary structure and codon bias to translation

PARS score ($r=0.51$, $N=2974$, $p<10^{-193}$) [30] and optimal codon usage (CAI) ($r=0.50$, $N=5077$, $p<10^{-178}$) [28] are known to be strongly positively associated with protein abundance. Because PARS and CAI are both properties of the mRNA, they are not expected to have direct causal effects on protein half-life, so we focus on the effects of PARS and CAI on translation efficiency. Our path analysis (Figure 2) shows that the direct effect of CAI on translation efficiency is much stronger than that of PARS

(0.38 vs. 0.10; Z-test, $p < 0.0001$). The indirect effect of PARS on translation efficiency as mediated by CAI ($\beta = 0.28 \times 0.38 = 0.11$) is similar to the direct effect ($\beta = 0.10$; Z-test, $p = 0.28$). Codon usage thus plays a substantial role in mediating the association between PARS score and translation efficiency, and it has a much larger direct effect than PARS.

mRNA secondary structure is associated with protein biological function

As shown above, mRNA secondary structure is significantly associated with transcriptional and translational processes. To examine the functional consequences of these associations, we asked whether there are relationships between mRNA secondary structure and biological function of the encoded proteins.

We first defined mRNAs to be highly-folded if they had an average PARS score greater than twice the average and lowly-folded if they had an average PARS score smaller than half the average. Gene Ontology (GO) enrichment analysis revealed that strongly folded transcripts were significantly ($p < 10^{-6}$) enriched for biological functions related to macromolecular (such as protein and lipid) metabolism and biosynthesis (Table 1). In contrast, weakly folded transcripts were significantly ($p < 10^{-6}$) enriched for cellular components such as mitochondrion, endoplasmic reticulum (ER) and ribosome assembly. Because PARS score is correlated with gene expression level [30], we also considered GO enrichment when genes were classified as highly- or lowly-folded while controlling for mRNA abundance (Supplemental Figure S1). Notably, we still found that the functions of lowly-folded transcripts were enriched in mitochondrion and ribosome assembly, while highly-folded transcripts were enriched in cell wall organization and metabolism-related enzyme activity (Supplemental Table S3). This suggests that highly-folded and lowly-folded transcripts contribute to different cellular processes (see Discussion).

[Table 1]

Discussion

RNA's ability to adopt complex secondary and tertiary folds has critical roles in processes ranging from ligand sensing to the regulation of translation, polyadenylation, and splicing [21, 22]. Here we investigated the relationships between mRNA secondary structure and codon content on multiple steps of gene expression (transcription, mRNA decay, translation) on a genome-wide scale.

We found that mRNA secondary structure is strongly positively correlated with transcription rate, and path analysis suggests that this is the dominant cause of the correlation between mRNA secondary structure and mRNA abundance. How does mRNA secondary structure affect transcription rate? Previous research has shown that the net transcription rate, whether limited by initiation or elongation, is strongly correlated with the density of RNA Polymerase II on the gene [43, 44]. We speculate that co-transcriptional folding of mRNA enhances transcription by reducing interaction between the nascent mRNA and adjacent polymerases, thereby enabling increased polymerase density. This model predicts that, for a given transcription rate, longer genes will need to fold more tightly, because their mRNAs are bulkier. Indeed, we find that gene length is significantly positively correlated with PARS score

($r=0.23$, $N=3001$, $p<10^{-35}$, Figure 3). Meanwhile, longer genes also probably have more regulating transcription factors that will also affect gene expression [45].

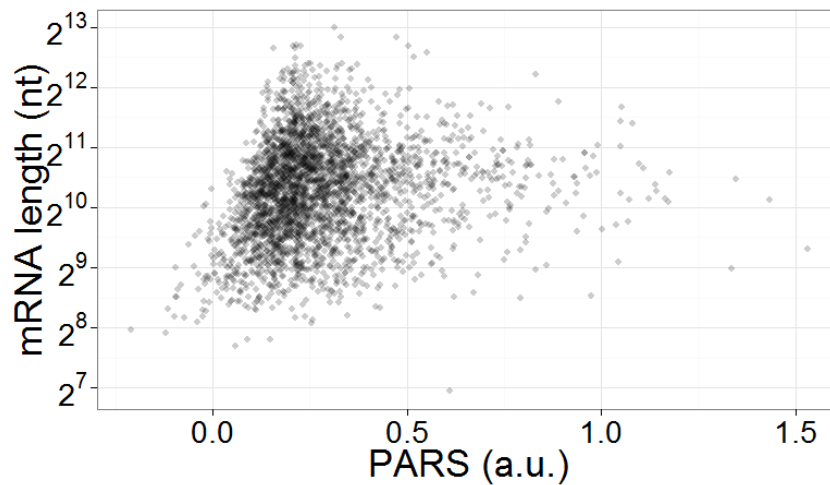


Figure 3. Strong correlation between mRNA length and mRNA secondary structure (PARS score).

We found that codon bias has much stronger direct effects on translation than mRNA structure, but they may still play complementary roles. Prior work suggests that local association within genes between mRNA secondary structure and high translation efficiency codons may act to smooth rates of translation [41]. We also found that strongly folded mRNAs have lower synonymous evolutionary rates dS , suggesting that selection is acting to conserve mRNA secondary structure [42, 46], thereby conserving its effects on gene expression. Together, these results suggest that a global coupled transcriptional and translational process controlling mRNA and protein abundance has strong association with mRNA secondary structure and other sequence information.

Growing evidence shows the importance of mRNA expression level in regulating transient responses to stress and stimulus such as temperature, oxidation, starvation, and metal ions [47, 48]. These stresses and stimuli can not only affect RNA structure and thermostability [47, 48], but also induce ribosomes, the endoplasmic reticulum, and mitochondria to undergo extensive structural and/or spatial reorganization [49, 50]. So are there any relationships between mRNA structure and biological functions of the encoded proteins? Our GO enrichment analysis revealed that lowly-folded mRNAs tend to function in the mitochondria and endoplasmic reticulum and in ribosome assembly. Recent *in vivo* genome-wide profiling of RNA secondary structure in *Arabidopsis thaliana* seedlings revealed that highly-folded mRNAs are related to cell function maintenance and lowly-folded mRNAs are related to stress and stimulus response [12]. Thus, lowly-folded mRNAs in yeast annotated for mitochondrion, ER, and ribosome function may also be relevant to stress response. In fact, ER ribosomes are the primary sites for synthesis and folding of proteins, which is sensitive to extracellular stresses and stimuli. Moreover, the ER and mitochondria interact both physiologically and functionally. Both are dynamic organelles capable of modifying their structure and function to maintain cellular homeostasis and determine

cell fate in response to changing environmental conditions [51]. Consistent with this model, in yeast RNA secondary structure (PARS) is negatively correlated ($r=-0.402$, $p<10^{-16}$) with fold change in translation efficiency in stress vs. normal conditions (Figure 4). Lowly-folded transcripts tend to have increased translation efficiency during stress condition, whereas highly-folded transcripts have reduced translation efficiency.

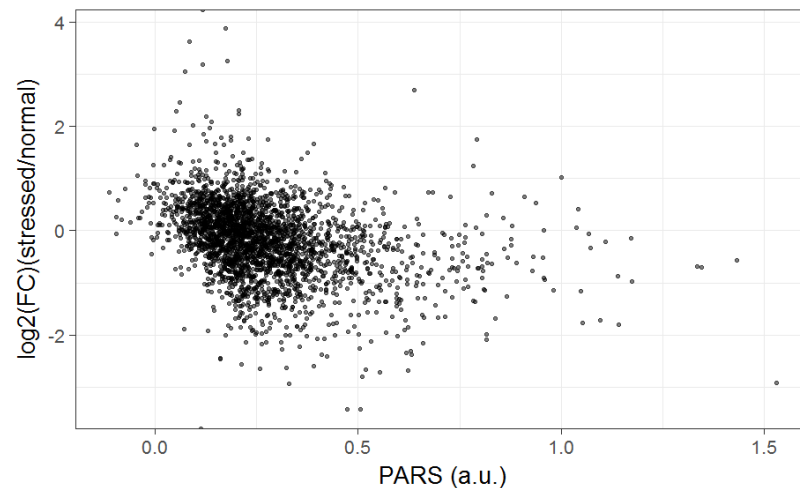


Figure 4. Strong correlation between fold change of translation efficiency in stress condition vs. normal condition and mRNA secondary structure (PARS score).

Combining our results, we found that lowly-folded mRNAs are related to mitochondrion, ER, and ribosome assembly and have short half-life and low transcription rates. After removal of stress, slow synthesis and rapid removal of stress-responsive mRNAs may thus facilitate rapid return of the cell to a normal state.

Additional material

Additional file 1: Supplementary Figures S1 and S2, Tables S1, S2, and S3.

Abbreviations

PARS: parallel analysis of RNA structure; RNAPII: RNA polymerase II; CAI: codon adaptation index; GO: Gene Ontology; ER: endoplasmic reticulum.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by DARPA contract WF911NF-14-1-0395 to R.G., and the NSF of China and Anhui Province (Grant No. 31500676 and 1508085SQC202, to X.W.). We thank Brian Mannakee and Travis Struck for helpful discussions.

Authors' contributions

XW and RG designed the analyses and wrote the manuscript. XW and PL conducted the data analysis. All authors have read and approved the manuscript for publication.

References

1. Pratt JM, Petty J, Riba-Garcia I, Robertson DH, Gaskell SJ, Oliver SG, Beynon RJ: **Dynamics of protein turnover, a missing dimension in proteomics.** *Molecular & Cellular Proteomics* 2002, **1**(8):579-591.
2. Marguerat S, Lawler K, Brazma A, Bahler J: **Contributions of transcription and mRNA decay to gene expression dynamics of fission yeast in response to oxidative stress.** *RNA Biol* 2014, **11**(6):702-714.
3. Dori-Bachash M, Shema E, Tirosh I: **Coupled evolution of transcription and mRNA degradation.** *PLoS Biol* 2011, **9**(7):e1001106.
4. Shalem O, Dahan O, Levo M, Martinez MR, Furman I, Segal E, Pilpel Y: **Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation.** *Mol Syst Biol* 2008, **4**:223.
5. Cajigas IJ, Will T, Schuman EM: **Protein homeostasis and synaptic plasticity.** *EMBO J* 2010, **29**(16):2746-2752.
6. Klein ME, Castillo PE, Jordan BA: **Coordination between Translation and Degradation Regulates Inducibility of mGluR-LTD.** *Cell Reports* 2015, **10**(9):1459-1466.
7. Voss TC, Hager GL: **Dynamic regulation of transcriptional states by chromatin and transcription factors.** *Nature Reviews Genetics* 2014, **15**(2):69-81.
8. Sidaway-Lee K, Costa MJ, Rand DA, Finkenzstadt B, Penfield S: **Direct measurement of transcription rates reveals multiple mechanisms for configuration of the Arabidopsis ambient temperature response.** *Genome Biol* 2014, **15**(3):R45.
9. Shepard PJ, Hertel KJ: **Conserved RNA secondary structures promote alternative splicing.** *RNA* 2008, **14**(8):1463-1469.
10. Buratti E, Baralle FE: **Influence of RNA secondary structure on the pre-mRNA splicing process.** *Mol Cell Biol* 2004, **24**(24):10505-10514.
11. Warf MB, Berglund JA: **Role of RNA structure in regulating pre-mRNA splicing.** *Trends in Biochemical Sciences* 2010, **35**(3):169-178.
12. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM: **In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features.** *Nature* 2014, **505**(7485):696-700.
13. Zafrir Z, Tuller T: **Nucleotide sequence composition adjacent to intronic splice sites improves splicing efficiency via its effect on pre-mRNA local folding in fungi.** *RNA* 2015, **21**(10):1704-1718.
14. Graveley BR, Fleming ES, Gilmartin GM: **RNA structure is a critical determinant of poly(A) site recognition by cleavage and polyadenylation specificity factor.** *Molecular and Cellular Biology* 1996, **16**(9):4942-4951.
15. Goodarzi H, Najafabadi HS, Oikonomou P, Greco TM, Fish L, Salavati R, Cristea IM, Tavazoie S: **Systematic discovery of structural elements governing stability of mammalian messenger RNAs.** *Nature* 2012, **485**(7397):264-268.
16. Fukunaga T, Ozaki H, Terai G, Asai K, Iwasaki W, Kiryu H: **CapR: revealing structural specificities of RNA-binding protein target recognition using CLIP-seq data.** *Genome Biology* 2014, **15**(1).
17. Incarnato D, Neri F, Anselmi F, Oliviero S: **Genome-wide profiling of mouse RNA**

- secondary structures reveals key features of the mammalian transcriptome.** *Genome Biol* 2014, **15**(10):491.
18. Keene JD: **RNA regulons: coordination of post-transcriptional events.** *Nature Reviews Genetics* 2007, **8**(7):533-543.
 19. Scharff LB, Childs L, Walther D, Bock R: **Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-Binding Sites.** *Plos Genetics* 2011, **7**(6).
 20. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY: **Understanding the transcriptome through RNA structure.** *Nature Reviews Genetics* 2011, **12**(9):641-655.
 21. Cruz JA, Westhof E: **The dynamic landscapes of RNA architecture.** *Cell* 2009, **136**(4):604-609.
 22. Sharp PA: **The centrality of RNA.** *Cell* 2009, **136**(4):577-580.
 23. Stergachis AB, Haugen E, Shafer A, Fu W, Vernot B, Reynolds A, Raubitschek A, Ziegler S, LeProust EM, Akey JM: **Exonic transcription factor binding directs codon choice and affects protein evolution.** *Science* 2013, **342**(6164):1367-1372.
 24. Plotkin JB, Kudla G: **Synonymous but not the same: the causes and consequences of codon bias.** *Nature Reviews Genetics* 2011, **12**(1):32-42.
 25. Trotta E: **The 3-Base Periodicity and Codon Usage of Coding Sequences Are Correlated with Gene Expression at the Level of Transcription Elongation.** *PLoS One* 2011, **6**(6).
 26. Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR *et al*: **Codon Optimality Is a Major Determinant of mRNA Stability.** *Cell* 2015, **160**(6):1111-1124.
 27. Pop C, Rouskin S, Ingolia NT, Han L, Phizicky EM, Weissman JS, Koller D: **Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation.** *Molecular systems biology* 2014, **10**(12).
 28. Tuller T, Waldman YY, Kupiec M, Ruppin E: **Translation efficiency is determined by both codon bias and folding energy.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(8):3645-3650.
 29. Trotta E: **Selection on codon bias in yeast: a transcriptional hypothesis.** *Nucleic Acids Res* 2013, **41**(20):9382-9395.
 30. Zur H, Tuller T: **Strong association between mRNA folding strength and protein abundance in *S. cerevisiae*.** *Embo Reports* 2012, **13**(3):272-277.
 31. Vogel C, de Sousa Abreu R, Ko D, Le SY, Shapiro BA, Burns SC, Sandhu D, Boutz DR, Marcotte EM, Penalva LO: **Sequence signatures and mRNA concentration can explain two - thirds of protein abundance variation in a human cell line.** *Molecular systems biology* 2010, **6**(1).
 32. Gingold H, Pilpel Y: **Determinants of translation efficiency and accuracy.** *Mol Syst Biol* 2011, **7**:481.
 33. Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E: **Genome-wide measurement of RNA secondary structure in yeast.** *Nature* 2010, **467**(7311):103-107.
 34. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS: **Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo.** *Nature* 2014, **505**(7485):701-705.
 35. Wright S: **The method of path coefficients.** *Annals of Mathematical Statistics* 1934, **5**:161-215.

36. Verdugo RA, Zeller T, Rotival M, Wild PS, Munzel T, Lackner KJ, Weidmann H, Ninio E, Tregouet DA, Cambien F *et al*: **Graphical modeling of gene expression in monocytes suggests molecular mechanisms explaining increased atherosclerosis in smokers.** *PLoS One* 2013, **8**(1):e50888.
37. Vogel H, Kroymann J, Mitchell-Olds T: **Different transcript patterns in response to specialist and generalist herbivores in the wild Arabidopsis relative Boechera divaricarpa.** *PLoS One* 2007, **2**(10):e1081.
38. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW: **Evolutionary rate in the protein interaction network.** *Science* 2002, **296**(5568):750-752.
39. Sharp PM, Li WH: **The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications.** *Nucleic Acids Res* 1987, **15**(3):1281-1295.
40. Zhang SP, Zubay G, Goldman E: **Low-usage codons in Escherichia coli, yeast, fruit fly and primates.** *Gene* 1991, **105**(1):61-72.
41. Gorochowski TE, Ignatova Z, Bovenberg RAL, Roubos JA: **Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate.** *Nucleic Acids Research* 2015, **43**(6):3022-3032.
42. Park C, Chen X, Yang JR, Zhang J: **Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly.** *Proc Natl Acad Sci U S A* 2013, **110**(8):E678-686.
43. Pelechano V, Chavez S, Perez-Ortin JE: **A complete set of nascent transcription rates for yeast genes.** *PLoS One* 2010, **5**(11):e15442.
44. Danko CG, Hah N, Luo X, Martins AL, Core L, Lis JT, Siepel A, Kraus WL: **Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells.** *Mol Cell* 2013, **50**(2):212-222.
45. Tuller T, Ruppin E, Kupiec M: **Properties of untranslated regions of the S. cerevisiae genome.** *BMC Genomics* 2009, **10**:391.
46. Chamary JV, Hurst LD: **Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals.** *Genome Biol* 2005, **6**(9):R75.
47. Grilley D, Soto AM, Draper DE: **Mg²⁺-RNA interaction free energies and their relationship to the folding of RNA tertiary structures.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(38):14003-14008.
48. Chursov A, Kopetzky SJ, Leshchiner I, Kondofersky I, Theis FJ, Frishman D, Shneider A: **Specific temperature-induced perturbations of secondary mRNA structures are associated with the cold-adapted temperature-sensitive phenotype of influenza A virus.** *RNA Biol* 2012, **9**(10):1266-1274.
49. Rainbolt TK, Saunders JM, Wiseman RL: **Stress-responsive regulation of mitochondria through the ER unfolded protein response.** *Trends in Endocrinology and Metabolism* 2014, **25**(10):528-537.
50. Liu B, Han Y, Qian SB: **Cotranslational response to proteotoxic stress by elongation pausing of ribosomes.** *Mol Cell* 2013, **49**(3):453-463.
51. Malhotra JD, Kaufman RJ: **ER stress and its functional link to mitochondria: role in cell survival and death.** *Cold Spring Harb Perspect Biol* 2011, **3**(9):a004424.
52. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR *et al*: **Saccharomyces Genome Database: the genomics**

- resource of budding yeast.** *Nucleic acids research* 2012, **40**(D1):D700-D705.
53. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M: **The transcriptional landscape of the yeast genome defined by RNA sequencing.** *Science* 2008, **320**(5881):1344-1349.
54. Miura F, Kawaguchi N, Yoshida M, Uematsu C, Kito K, Sakaki Y, Ito T: **Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs.** *BMC Genomics* 2008, **9**:574.
55. Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, von Mering C: **PaxDb, a Database of Protein Abundance Averages Across All Three Domains of Life.** *Molecular & Cellular Proteomics* 2012, **11**(8):492-500.
56. Christiano R, Nagaraj N, Frohlich F, Walther TC: **Global proteome turnover analyses of the Yeasts *S. cerevisiae* and *S. pombe*.** *Cell Rep* 2014, **9**(5):1959-1965.
57. McManus CJ, May GE, Spealman P, Shteyman A: **Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast.** *Genome Res* 2014, **24**(3):422-430.
58. Gerashchenko MV, Lobanov AV, Gladyshev VN: **Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress.** *Proc Natl Acad Sci U S A* 2012, **109**(43):17394-17399.
59. Guydosh NR, Green R: **Dom34 rescues ribosomes in 3' untranslated regions.** *Cell* 2014, **156**(5):950-962.
60. Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, Eisen MB, Feldman MW: **Functional genomic analysis of the rates of protein evolution.** *Proc Natl Acad Sci U S A* 2005, **102**(15):5483-5488.
61. Huang dW, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res* 2009, **37**(1):1-13.
62. Fox J: **Structural equation modeling with the sem package in R.** *Structural Equation Modeling-a Multidisciplinary Journal* 2006, **13**(3):465-486.

Table 1. Top 30 overrepresented GO terms in sets of highly and lowly folded mRNA.

highly folded mRNA			lowly folded mRNA		
GO-ID	Term	p-value	GO-ID	Term	p-value
GO:0006417	regulation of translation	2.89E-33	GO:0044429	mitochondrial part	2.05E-16
GO:0010608	posttranscriptional regulation of gene expression	1.06E-32	GO:0003735	structural constituent of ribosome	1.44E-14
GO:0032268	regulation of cellular protein metabolic process	7.70E-31	GO:0005761	mitochondrial ribosome	5.86E-14
GO:0044445	cytosolic part	9.11E-31	GO:0000313	organellar ribosome	5.86E-14
GO:0005829	cytosol	1.61E-24	GO:0033279	ribosomal subunit	9.42E-14
GO:0022626	cytosolic ribosome	3.49E-20	GO:0031090	organelle membrane	7.21E-13
GO:0006412	translation	5.65E-20	GO:0031974	membrane-enclosed lumen	3.97E-12
GO:0005783	endoplasmic reticulum	7.88E-18	GO:0030529	ribonucleoprotein complex	1.29E-11
GO:0001950	plasma membrane enriched fraction	2.01E-16	GO:0005840	ribosome	1.36E-11
GO:0005626	insoluble fraction	1.55E-14	GO:0070013	intracellular organelle lumen	2.82E-10
GO:0005624	membrane fraction	1.55E-14	GO:0043233	organelle lumen	2.82E-10
GO:0005576	extracellular region	2.89E-14	GO:0031975	envelope	1.26E-09
GO:0022625	cytosolic large ribosomal subunit	1.88E-13	GO:0031967	organelle envelope	1.26E-09
GO:0000267	cell fraction	2.06E-13	GO:0034470	ncRNA processing	1.81E-09
GO:0009277	fungal-type cell wall	1.59E-12	GO:0009451	RNA modification	5.20E-09
GO:0008610	lipid biosynthetic process	8.30E-12	GO:0005198	structural molecule activity	6.52E-09
GO:0042802	identical protein binding	9.13E-12	GO:0005759	mitochondrial matrix	8.40E-09
GO:0005618	cell wall	2.02E-11	GO:0031980	mitochondrial lumen	8.40E-09
GO:0030312	external encapsulating structure	2.02E-11	GO:0005762	mitochondrial large ribosomal subunit	9.42E-09
GO:0043094	cellular metabolic compound salvage ergosterol	4.88E-11	GO:0000315	organellar large ribosomal subunit	9.42E-09
GO:0006696	biosynthetic process	1.86E-10	GO:0008033	tRNA processing	1.07E-08

GO:0016129	phytosteroid biosynthetic process	1.86E-10	GO:0042254	ribosome biogenesis	1.56E-08
GO:0000502	proteasome complex	3.04E-10	GO:0012505	endomembrane system	2.26E-08
GO:0046394	carboxylic acid biosynthetic process	3.38E-10	GO:0005740	mitochondrial envelope	2.81E-08
GO:0016053	organic acid biosynthetic process	3.38E-10	GO:0044432	endoplasmic reticulum part	3.01E-08
GO:0055114	oxidation reduction	4.37E-10	GO:0031966	mitochondrial membrane	5.33E-08
GO:0044271	nitrogen compound biosynthetic process	4.44E-10	GO:0032543	mitochondrial translation	1.42E-07
GO:0031597	cytosolic proteasome complex	1.03E-09	GO:0015935	small ribosomal subunit	3.05E-07
GO:0034515	proteasome storage granule	1.03E-09	GO:0019866	organelle inner membrane	3.14E-07
GO:0016126	sterol biosynthetic process	1.19E-09	GO:0005743	mitochondrial inner membrane	3.68E-07
