Translation Initiation: A Regulatory Role for Poly(A) Tracts in Front of the AUG Codon in *Saccharomyces cerevisiae*

Xuhua Xia,*[,1] Vivian MacKay,[†] Xiaoquan Yao,* Jianhua Wu,[‡] Fumihito Miura,[§] Takashi Ito,[§] and David R. Morris[†]

*Department of Biology, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada, [†]Department of Biochemistry, University of Washington, Seattle, Washington 98195-7350, [‡]Institute of Biomechanics and Department of Biomedical Engineering, School of Bioscience and Bioengineering, South China University of Technology, 510641 Guangzhou, China, and [§]Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Bunkyo-Ku, Tokyo 113-0032, Japan

ABSTRACT The 5′-UTR serves as the loading dock for ribosomes during translation initiation and is the key site for translation regulation. Many genes in the yeast *Saccharomyces cerevisiae* contain poly(A) tracts in their 5′-UTRs. We studied these pre-AUG poly(A) tracts in a set of 3274 recently identified 5′-UTRs in the yeast to characterize their effect on *in vivo* protein abundance, ribosomal density, and protein synthesis rate in the yeast. The protein abundance and the protein synthesis rate increase with the length of the poly(A), but exhibit a dramatic decrease when the poly(A) length is ≥12. The ribosomal density also reaches the lowest level when the poly(A) length is ≥12. This supports the hypothesis that a pre-AUG poly(A) tract can bind to translation initiation factors to enhance translation initiation, but a long (≥12) pre-AUG poly(A) tract will bind to Pab1p, whose binding size is 12 consecutive A residues in yeast, resulting in repression of translation. The hypothesis explains why a long pre-AUG poly(A) leads to more efficient translation initiation than a short one when PABP is absent, and why pre-AUG poly(A) is short in the early genes but long in the late genes of vaccinia virus.

POLY(A) tracts in 5′-UTR have been recognized recently as important sites for translation regulation. These poly (A) tracts, referred to hereafter as pre-AUG A_N, where N stands for the number of consecutive A nucleotides, can interact with translation initiation factors or poly(A) binding proteins (PABP) to either increase or decrease translation efficiency. Pre-AUG A_N can enhance internal ribosomal entry both in the presence of PABP and eIF-4G in the yeast, *Saccharomyces cerevisiae* (Gilbert *et al.* 2007), and in the complete absence of PABP and eIF-4G (Shirokikh and Spirin 2008). Translation initiation factors eIF-4B and eIF-4F can bind to poly(A) tracts (Gallie and Tanguay 1994), and exogenous poly(A) added to an *in vitro* translation system can inhibit translation initiation (Lodish and Nathan 1972; Jacobson and Favreau 1983; Grossi De Sa *et al.* 1988), most probably by sequestering the translation initiation factors (Gallie and Tanguay 1994) and PABP (Gilbert *et al.* 2007). The inhibiting effect of exogenous poly(A) on translation can be removed by addition of either translation initiation factors eIF-4B and eIF-4F (with eIF-4A) in combination (Gallie and Tanguay 1994) or PABP (Grossi De Sa *et al.* 1988; Gilbert *et al.* 2007) which presumably would bind to the exogenous poly(A) and free translation initiation factors sequestered by the exogenous poly(A).

While pre-AUG A_N may improve translation efficiency, a few studies (Wu and Bag 1998; Bag 2001; Melo *et al.* 2003a,b; Patel *et al.* 2005; Ma *et al.* 2006; Patel and Bag 2006; Bag and Bhattacharjee 2010) suggest an inhibitory effect of PABP when it binds to a long pre-AUG A_N and presumably interferes with the scanning mechanism of translation initiation. The binding site of Pab1p in the yeast is about 12 consecutive A nucleotides, with the binding affinity decreasing rapidly with shorter poly(A) until 8, below which there is little affinity (Sachs *et al.* 1987). This suggests that mRNAs with pre-AUG A_N of different lengths may

doi: 10.1534/genetics.111.132068
Manuscript received June 25, 2011; accepted for publication July 24, 2011
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/content/suppl/2011/08/12/genetics.111.132068.DC1.
[1]Corresponding author: Department of Biology, University of Ottawa, 30 Marie Curie, P. O. Box 450, Station A, Ottawa, ON, K1N 6N5, Canada. E-mail: xxia@uottawa.ca

interact differently with yeast Pab1p and have different translation efficiencies.

Several recent technological breakthroughs have eliminated two fundamental difficulties in characterizing the relationship between pre-AUG A_N and translation efficiency in *S. cerevisiae*. First, the transcription start site (TSS) necessary for the accurate delineation of 5′-UTR and pre-AUG A_N has been characterized for thousands of yeast genes by direct mapping of the capped yeast mRNA sequences (Miura *et al.* 2006). This approach to delineate TSSs is conceptually more direct and technologically more accurate than the serial analysis of gene expression (SAGE) approach (Zhang and Dietrich 2006). A merged set of TSSs from these two studies has recently been compiled (Lawless *et al.* 2009).

The second difficulty in quantifying the relationship between pre-AUG A_N and translation efficiency is the lack of large-scale characterization of protein production and ribosomal density on mRNA. This difficulty is alleviated by a recent proteomic study in yeast (Ghaemmaghami *et al.* 2003) and several studies characterizing the ribosomal density and the protein synthesis rate for thousands of yeast genes (Arava *et al.* 2003; MacKay *et al.* 2004; Ingolia *et al.* 2009).

Through a detailed analysis of sequence features in characterized 5′-UTRs and their relationship to *in vivo* protein production and ribosomal loading, two interesting patterns were revealed. First, the frequency of nucleotide A increased dramatically toward the initiation codon, with a concurrent decrease in nucleotide U. This trend is particularly pronounced in highly expressed genes relative to lowly expressed genes when gene expression is measured by either the protein abundance, by the predicted rate of protein synthesis based on mRNA abundance, by the ribosomal density, or by the codon adaptation index, which is highly correlated with transcript and protein abundance in the yeast (Duret and Mouchiroud 1999; Coghlan and Wolfe 2000). Second, the protein abundance and the protein synthesis rate both increase with the length of pre-AUG A_N up to $N = 11$ and decrease dramatically for genes having a pre-AUG A_N with $N \geq 12$. The ribosomal density also reaches the lowest level for genes having a pre-AUG A_N with $N \geq 12$.

Materials and Methods

The 5′-UTR sequences were extracted by using the specification in table 4 of Miura *et al.* (2006) and the yeast genomic sequences retrieved from University of Tokyo Genome Browser (http://yeast.utgenome.org/). A total of 3274 genes have their TSS characterized, with many having multiple TSSs and consequently multiple 5′-UTRs. To avoid overrepresentation of the sequence patterns of genes with multiple 5′-UTRs, only the longest 5′-UTR for each gene is used. The 5′-UTR sequences are available as a FASTA file (MiuraTSS.FAS in Supporting Information, File S5).

Three measures of translation efficiency were used for quantifying the relationship between pre-AUG A_N and translation efficiency. The first is codon adaptation index (CAI) (Sharp and Li 1987) with its improved version (Xia 2007) implemented in DAMBE (Xia 2001; Xia and Xie 2001). We retrieved the genomic sequences of the 16 chromosomes of *S. cerevisiae* from the genome database at National Center for Biotechnology Information, extracted coding sequences (CDSs) and calculated their corresponding CAI values by using DAMBE with the Eysc_h.cut codon usage table. The benefit of using CAI is that it can be computed for any gene with a codon frequency distribution, whereas experimentally measured protein expression or ribosomal loading data are often limited to relatively highly expressed proteins or transcripts.

The second measure of translation efficiency is the protein abundance, in units of molecules/cell, measured experimentally in a large-scale quantification of yeast protein abundance (Ghaemmaghami *et al.* 2003). The dataset contains protein abundance data for 3850 yeast genes after excluding 18 genes that do not have a matched name in the current yeast database. For characterizing the relationship between the protein abundance and the length of pre-AUG A_N, the protein abundance was log transformed to stabilize the variance and to linearize the relationship. The protein abundance data used in this article is available as GhaemmaghamiProtein.xls in File S4.

The third measure of translation efficiency is based on large-scale characterizations of ribosomal loading on gene-specific mRNA sequences (Arava *et al.* 2003; MacKay *et al.* 2004; Ingolia *et al.* 2009). The experimental data for the ribosomal density and the predicted protein synthesis rate in two experimental conditions (mating pheromone treatment and control) were reliably measured for 3916 genes (supplemental table II in MacKay *et al.* 2004). Another data set characterizing ribosomal density, with a similar method but in a different laboratory (Arava *et al.* 2003), as well as a data set characterizing ribosomal loading with the quite different deep sequencing method (Ingolia *et al.* 2009), were also analyzed and compared against results from MacKay *et al.* (2004). These three data sets, referred to hereafter as MacKayData, AravaData, and IngoliaData are noteworthy in that they are highly concordant (Concordance.pdf in File S1 and File S2). All other data used this paper are in AllData.xls (File S3).

Results

Overrepresentation of A in 5′-UTRs

The nucleotide frequency distribution over 150 5′-UTR sites immediately upstream of the translation initiation codon shows a dramatic increase of nucleotide A toward the translation initiation codon starting from site −40, with a concurrent decrease in nucleotide U (Figure 1). The nucleotide frequency distribution before and after site −40 are highly significantly different ($\chi^2 = 2815.84$, DF = 3, $P < 0.0001$). A similar pattern was also observed in a previous study

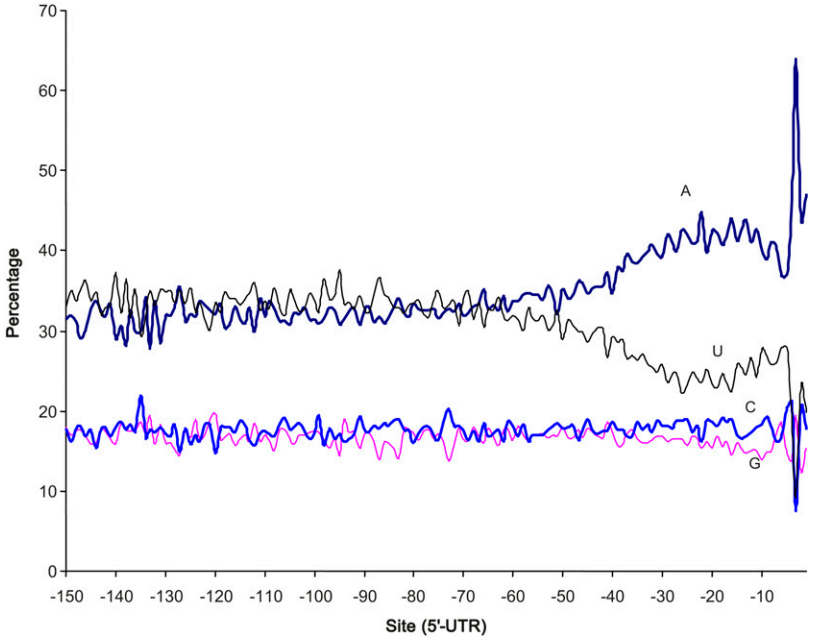

Figure 1 The site-specific nucleotide frequencies of 150 nucleotide sites (*x*-axis), immediately upstream from the start codon AUG in yeast 5′-UTR sequences. Site −1 is represented by 3274 5′-UTR sequences and site −150 is represented by 747 5′-UTR sequences.

(Shabalina *et al.* 2004), which suggested that the 30-nt segment immediately upstream of the initiation codon was functionally important in eukaryotic genes, although that study suffered from the fact that sequences upstream of the initiation codon in the genome are not necessarily part of the 5′-UTR because many yeast genes have 5′-UTRs shorter than 30 nt.

If the pattern of increasing A and decreasing U toward the initiation codon (Figure 1) is related to translation efficiency, then the pattern should be more pronounced in highly expressed genes than in lowly expressed genes. To test this prediction, we compared the nucleotide distribution between the 500 genes with the highest CAI values (high-expression group, HEG) and the 500 genes with the lowest CAI values (low-expression group, LEG). Genes in HEG exhibit the pattern much more dramatically than those in the LEG (Figure 2). The nucleotide frequency distribution between HEG and LEG are highly significantly different (χ^2 = 592.74, DF = 3, $P < 0.0001$). The same pattern is observed when the predicted protein synthesis rate (MacKay *et al.* 2004) is used as a proxy of translation efficiency (Figure 3).

Genes having a pre-AUG A_N with $N < 12$ produce more proteins than those With $N \geq 12$

Genes with a pre-AUG $A_{<12}$ have protein abundance higher than those with a pre-AUG $A_{\geq 12}$ (Table 1), with the two groups differing from each other significantly (F = 4.96, DF_{model} = 1, DF_{error} = 2218, P = 0.0261). There is no significant difference either among poly(A) length groups 2, 3, . . . , 11 or among the poly(A) length groups with the length ≥12.

One peculiar data point in Table 1 is the protein abundance for genes with a pre-AUG A_{10}. These genes have a singularly low protein abundance and include a gene with the lowest protein abundance (*YJL084C/ALY2*, alpha arrestin) of only 3.8950. The A_{10} in this gene precedes "GA," *i.e.*, in the configuration of AAAAAAAAAAGA. There are two other yeast genes (YPL271W and YBL101C) sharing this AAAAAAAAAABA configuration (where B means "not A," *i.e.*, C, G, or U) with protein abundance values equal to 8.3610 and 6.6839 (which is also low), respectively. Also, one of the A_{10} genes (YMR181C) is starvation-induced (Gilbert *et al.* 2007) and has a relatively low protein abundance as well (= 6.5800). Without these four genes, the mean protein abundance for A_{10} genes is 8.1528.

The observation above led us to ask whether the pre-AUG A_N needs to be contiguous for its effect on protein abundance. We compiled genes with a pre-AUG 12mer that has 11 A residues but with a non-A breaking the contiguity, *i.e.*, in the configuration of A**B**AAAAAAAAAA, AA**B**AAAAAAAAA, AAA**B**AAAAAAAA, ... , AAAAAAAAA**B**A and compared their protein abundance with those of genes containing a pre-AUG A_{12}. Protein abundance is significantly higher in the former (mean = 8.5970, n = 43) than in the latter (mean = 7.2013, n = 12), with t = 2.3453, DF = 53, and P = 0.0228 (two-tailed test). Thus, contiguity in poly(A) appears important for decreasing protein production. However, protein abundance is higher for genes with a non-A close to the 5′ end than with a non-A at the 3′ end, *i.e.*, A**B**AAAAAAAAAA and AA**B**AAAAAAAAA genes produce more proteins (mean = 8.6399, n = 9) than AAAAAAA AAA**B**A and AAAAAAAAA**B**AA genes (mean = 6.9207, n = 6), with t = 2.2314, DF = 13, P = 0.0439 (two-tailed test). This suggests that where contiguity is broken may be important.

Genes with a pre-AUG A_{11} produce more proteins than those with a pre-AUG A_{12}

The binding affinity of Pab1p approaches maximum when the poly(A) length reaches 12 but decreases quickly when

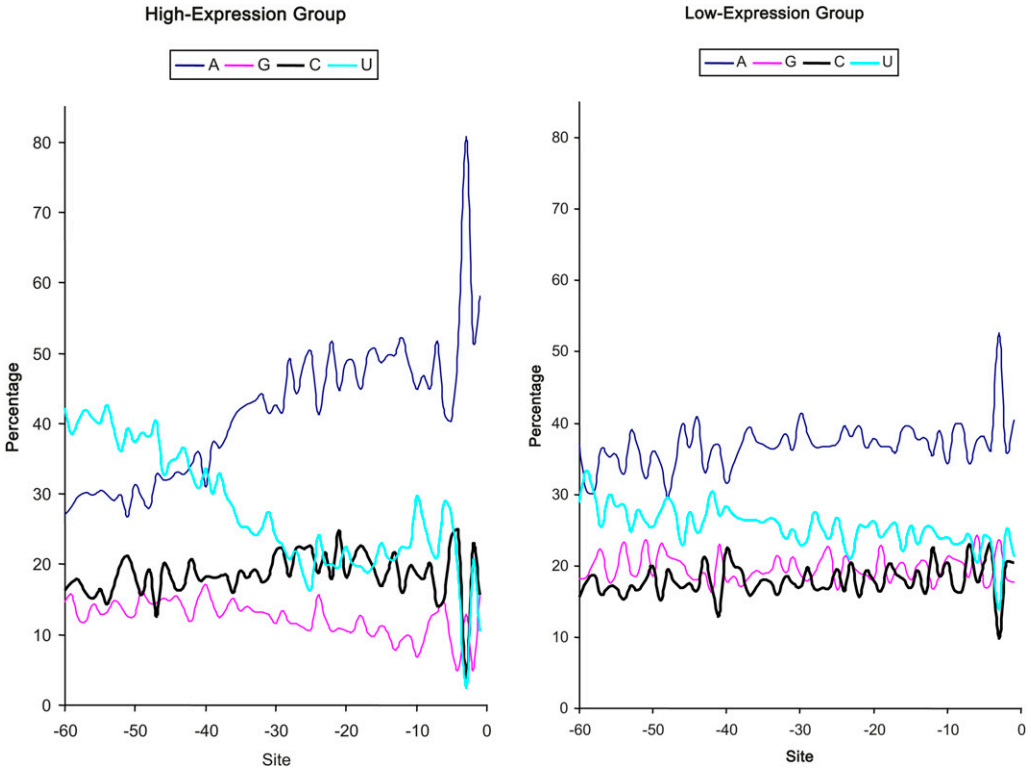

Figure 2 Contrasting site-specific nucleotide frequencies between highly expressed genes (500 genes with the highest CAI values) and lowly expressed genes (500 genes with the lowest CAI values). The increase in A usage toward the initiation AUG is visually more dramatic in highly expressed genes than lowly expressed genes. The nucleotide frequency distribution between HEG and LEG are highly significantly different ($\chi^2 = 592.74$, DF = 3, $P < 0.0001$).

the poly(A) length is 11 or shorter, with hardly any affinity when the poly(A) length is 8 (Sachs *et al.* 1987). So we examined whether genes with a pre-AUG A_{11}, on average, produce more proteins than genes with a pre-AUG A_{12}.

Among the yeast genes with characterized protein abundance (Ghaemmaghami *et al.* 2003), 12 have a pre-AUG A_{12} and 18 have a pre-AUG A_{11} (including the *Pab1/YER165W* mRNA, Table 2). Genes with a pre-AUG A_{11} have higher protein abundance than genes with a pre-AUG A_{12} ($t = 3.5827$, DF = 28, $P = 0.0013$, two-tailed test, Table 2). This is consistent with the interpretation that mRNAs with a pre-AUG $A_{<12}$ may enhance translation, but mRNAs with a pre-AUG $A_{\geq 12}$ may be subject to translation repression mediated by PABP binding to the pre-AUG poly(A) and interfering with the ribosomal scanning. The mean protein abundance values are 7.2013 for genes with a pre-AUG A_{12} and 9.2160

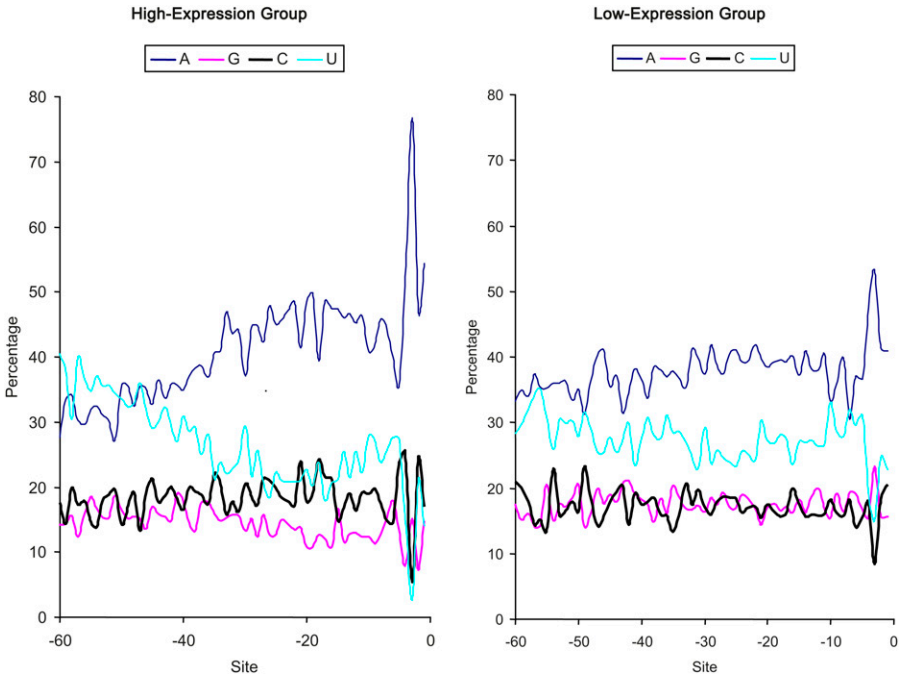

Figure 3 Contrasting site-specific nucleotide frequencies between highly expressed genes (500 genes with the highest predicted protein synthesis rates) and lowly expressed genes (500 genes with the lowest predicted protein synthesis rates). Protein synthesis rates are from supplemental table II in MacKay *et al.* (2004).

Table 1 Comparison of protein abundance among genes with different lengths (*L*) of pre-AUG poly(A)

L	*N*	Mean	SD
2	275	7.9081	1.4264
3	523	8.0553	1.6128
4	599	8.1663	1.5917
5	355	8.2830	1.7301
6	187	8.1552	1.7907
7	114	8.1861	1.6532
8	66	8.6858	1.7825
9	34	8.3668	1.5573
10	20	7.7982	1.6347
11	18	9.2160	1.5686
12	12	7.2013	1.4117
13	4	7.6616	0.4500
14	4	7.4732	1.8465
15	5	7.6702	1.6804
16	4	7.8449	1.1042

The differences are statistically significant ($F = 2.32$, $DF_{model} = 14$, $DF_{error} = 2205$, $P = 0.0036$). The *P*-value would be reduced to 0.0004 if the genes with poly(A) length ≥12 were lumped into one group. A gene (YJR113C) with an A_{21} in its 5′-UTR was included into the group of genes with a A_{16}.

for genes with a pre-AUG A_{11} (Table 2). Converted back to the original scale, the former is 1341.12 molecules/cell and the latter is 10056.42 molecules/cell.

The analysis above based on protein abundance has the problem that protein abundance is affected by a number of other factors such as translation elongation and mRNA abundance. For example, the protein abundance is highly correlated with CAI, which is a measure of translation elongation efficiency (Figure 4). It is therefore possible that the difference in protein abundance between the A_{11} genes and the A_{12} genes may simply be due to the mRNAs from the A_{11} genes having higher translation elongation efficiency than those from the A_{12} genes. It is known that different genes can differ much in translation elongation efficiency because of their differential codon adaptation to the tRNA pool (Xia 1998, 2005; van Weringh *et al.* 2011).

To control for the potential difference in translation elongation efficiency between the A_{11} and A_{12} genes, we have performed an analysis of covariance (ANCOVA) to test the difference in protein abundance between the A_{11} and A_{12} genes with CAI as a covariate. The relationship between protein production and CAI (Figure 4) is highly significant (slope = 9.8479, $P < 0.0001$), and the intercept for A_{11} genes is highly significantly greater than that for A_{12} genes by 1.4108 ($P = 0.0015$). Thus, the A_{11} genes still have significantly higher protein abundance than the A_{12} genes when translation elongation efficiency has been controlled for by using CAI as a covariate.

We have also performed a similar ANCOVA by using characterized mRNA abundance (Holstege *et al.* 1998) as a covariate. Protein abundance is significantly correlated with the mRNA abundance (Figure 5) for yeast genes with either an A_{11} or an A_{12} in their 5′-UTR, with a slope of 0.9455 ($P < 0.0001$). The intercept for A_{11} genes is significantly greater than that for A_{12} genes by 1.1096 ($P = 0.0274$). Thus, the A_{11} genes produce more proteins than the A_{12} genes, when mRNA abundance has been controlled for. A similar pattern is observed when the mRNA abundance characterized by Miura *et al.* (2008) is used.

Protein synthesis rate and ribosomal density depend on the length of pre-AUG A_N

One problem with the analysis above is that protein abundance depends on both protein production and protein degradation. For example, cyclins typically have a half-life in minutes (Aviram *et al.* 2008) and are consequently

Table 2 Protein abundance for genes with a poly(A) tract of length 11 or 12 (A_{11} or A_{12}) in their 5′-UTR

Gene	$L_{5'\text{-UTR}}$[a]	$MaxL_{Poly(A)}$[b]	lnProt[c]	CAI
YHR082C	867	12	7.1448	0.4409
YBR077C	68	12	8.1566	0.3919
YDR295C	266	12	7.3121	0.3714
YFR047C	43	12	7.8878	0.4590
YGL006W	282	12	4.5945	0.4082
YGR238C	133	12	6.1493	0.3607
YHR043C	112	12	7.7817	0.3977
YKR097W	175	12	10.2470	0.5721
YLR190W	461	12	6.9747	0.3952
YLR192C	109	12	9.7912	0.5625
YMR016C	684	12	5.7506	0.3615
YMR251W-A	142	12	8.7225	0.6739
YNL051W	203	12	6.1493	0.4228
YNL128W	265	12	6.9881	0.3633
YBR056W-A	47	11	7.5809	0.4077
YDL140C	519	11	10.1271	0.4864
YDL246C	321	11	7.1664	0.4695
YDR033W	555	11	12.1143	0.7725
YDR055W	82	11	9.3654	0.5406
YER115C	294	11	8.1442	0.3465
YER159C	180	11	8.5424	0.2948
YER165W	145	11	12.1955	0.7079
YGL037C	373	11	8.9517	0.5606
YHL012W	77	11	7.4276	0.3696
YJR070C	72	11	10.5931	0.6736
YKL080W	130	11	9.9577	0.4954
YKL084W	52	11	11.5340	0.4050
YKL186C	377	11	7.4939	0.3297
YKR092C	318	11	9.4679	0.5404
YLL013C	298	11	6.7403	0.4263
YLL029W	155	11	8.0823	0.3828
YLR328W	90	11	8.5424	0.4664
YLR341W	98	11	7.9929	0.4145
YML091C	316	11	7.0521	0.4622
YMR070W	213	11	7.4353	0.3715
YMR291W	735	11	9.2261	0.4517
YMR296C	211	11	10.0159	0.4094
YOL155W-A	195	11	8.6049	0.3095
YPL154C	279	11	7.7964	0.6100
YPL204W	189	11	9.2396	0.4254
YPR041W	121	11	10.7845	0.5831

The protein abundance is significantly higher in A_{11} genes (mean = 8.9694) than in A_{12} genes (mean = 7.4036), on the basis of a two-sample *t*-test ($t = 3.0879$, DF = 39, $P = 0.0037$, two-tailed test). Codon adaptation index (CAI) is included as a covariate for analysis of covariance (see text for details).

[a] 5′-UTR length.

[b] Maximum length of poly(A) tracts in 5′-UTR.

[c] Log-transformed protein abundance.

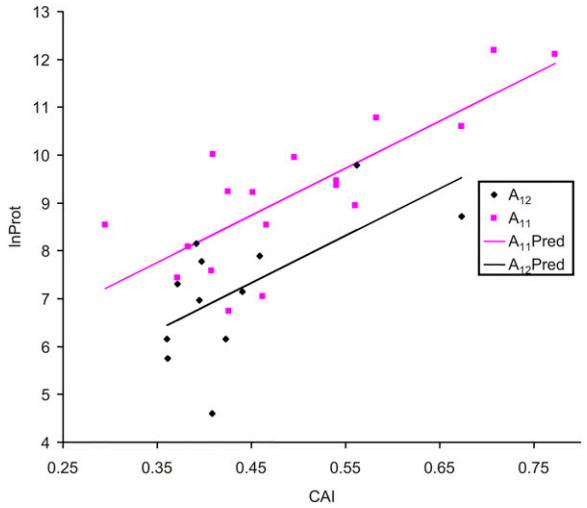

Figure 4 Visualization of the difference in protein abundance between A_{11} and A_{12} genes, when codon adaptation index (CAI) is use as a covariate to control for differences in translation elongation efficiency. A_{11}Pred and A_{12}Pred are fitted regression lines from an analysis of covariance, and lnProt is the log-transformed protein abundance from Ghaemmaghami *et al.* (2003). The difference in intercept is 1.4108 ($P = 0.0015$).

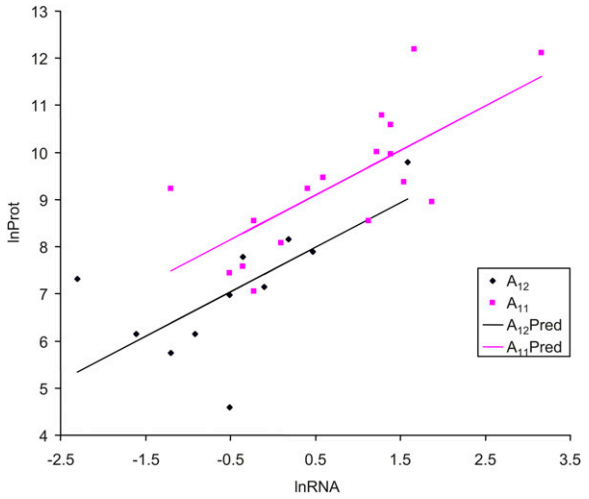

Figure 5 Visualization of the difference in protein abundance between the A_{11} and the A_{12} genes, when mRNA abundance (lnRNA) is use as a covariate to control for the potential difference in mRNA abundance between the A_{11} and the A_{12} genes. A_{11}Pred and A_{12}Pred are fitted regression lines from an analysis of covariance. lnProt is the log-transformed protein abundance from Ghaemmaghami *et al.* (2003). The difference in intercept is 1.1096 ($P = 0.0274$). lnRNA is the log-transformed mRNA abundance from Holstege *et al.* (1998). A similar pattern is observed when mRNA abundance from Miura *et al.* (2008) is used.

expected to have relatively low abundance even if translation initiation and elongation of their mRNAs are efficient. Thus, the low protein abundance in the A_{12} genes relative to the A_{11} genes could be due to more rapid degradation of proteins from the A_{12} genes than those from the A_{11} genes.

The ultimate solution to avoid the shortcomings above is to use a more direct measure of translation efficiency than protein abundance to avoid the confounding effect of mRNA abundance or protein degradation. Such measures, expressed as translation efficiency on the basis of ribosomal density, have become available recently for yeast genes (Arava *et al.* 2003; Serikawa *et al.* 2003; MacKay *et al.* 2004; Ingolia *et al.* 2009).

Among the 3916 genes in MacKayData, 2139 genes have their transcription start sites experimentally determined by Miura *et al.* (2006). The predicted protein synthesis rate was measured under two experimental conditions (control yeast cells and those treated with a mating pheromone). The two measured rates, however, are similar, with the Pearson correlation being 0.9317 for log-transformed data. We simply took the average of the two rates, referred to hereafter as MeanRate (of protein synthesis), and studied its relationship to the length of pre-AUG A_N.

MeanRate increases with the length of pre-AUG A_N, but decreases dramatically when poly(A) length reaches 12 (Figure 6). The regression line in Figure 6 is highly significant ($P < 0.0001$) for genes whose pre-AUG A_N is no longer than 11. Thus, a pre-AUG A_N, up to 11, tends to increase the protein synthesis rate, but a long ($\geq$12) pre-AUG A_N tends to decrease the protein synthesis rate.

We have also examined the relationship between the experimentally measured ribosomal density and the length of the pre-AUG A_N. Ingolia *et al.* (2009) characterized ribosomal density for yeast genes in both rich and starvation conditions. Here we use their ribosomal density for 5311 yeast genes characterized under two rich conditions, log transformed to stabilize variance. While ribosomal density fluctuate for genes with the pre-AUG A_N length (Figure 7), the mean ribosomal density for genes with a pre-AUG $A_{<12}$ (= 3.6879) is significantly greater than that (=3.0136) for genes with a pre-AUG $A_{\geq 12}$ ($t = 2.1932$, DF = 2911, $P = 0.0284$, two-tailed test).

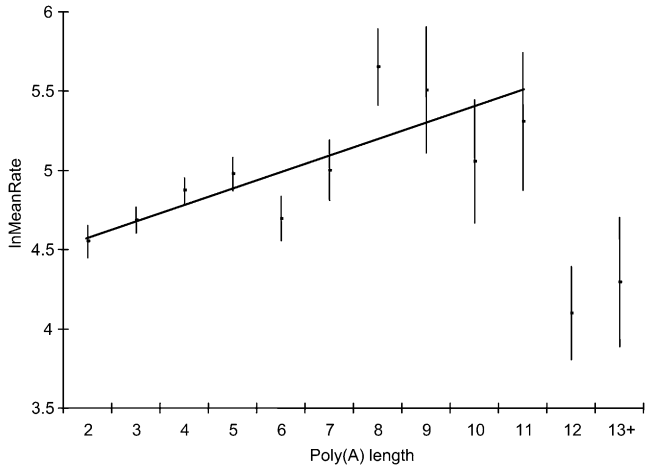

Figure 6 Protein synthesis rate (lnMeanRate) increases with the length of pre-AUG poly(A), but decreases dramatically when the poly(A) length reaches 12. The straight line indicates the regression line for genes with poly(A) length shorter than 12. The vertical bars show one standard error above and below the mean for each poly(A) length category.

Is low protein abundance in cyclin due to rapid degradation?

The cyclin gene *PCL5/YHR071W* has a pre-AUG A_{29}, which is the longest among all yeast genes. The Pcl5 protein has a half-life of only 2–3 min, and its low abundance is usually attributed to its rapid degradation (Aviram *et al.* 2008). Our results suggest that the low abundance may be partially attributable to inefficient translation initiation presumably mediated by the binding of Pab1p to the long pre-AUG A_{29}. The ribosomal occupancy is extremely low for *PCL5* mRNA relative to other yeast genes (Figure 8).

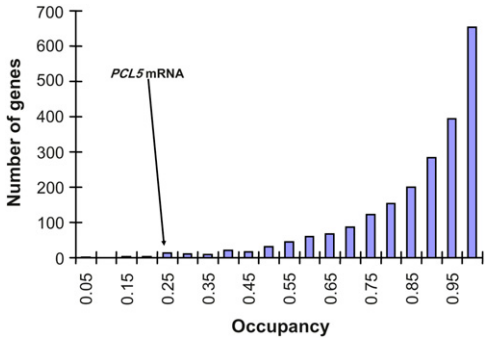

Figure 8 Frequency distribution of occupancy (proportion of mRNA associated with polysomes) of 2175 yeast genes (MacKay *et al.* 2004) with 5′-UTR characterized by Miura *et al.* (2006). The mean occupancy of *PCL5* mRNA from two measurements is 0.2360.

Discussion

Interaction between pre-AUG poly(A) and Pab1p

Our finding that yeast genes with a pre-AUG $A_{\geq 12}$ have a much reduced predicted protein synthesis rate (Figures 6) and ribosomal density (Figure 7) is consistent with the hypothesis that, while a pre-AUG A_N may enhance translation by binding to translation initiation factors (Gallie and Tanguay 1994; Shirokikh and Spirin 2008), a long pre-AUG A_N, by binding tightly to PABP, may repress translation by interfering with ribosomal scanning (Sachs *et al.* 1987; Wu and Bag 1998; Bag 2001; Melo *et al.* 2003a,b; Patel *et al.* 2005; Ma *et al.* 2006; Patel and Bag 2006). This hypothesis would predict that removing PABP would remove its inhibitory effect on mRNA with long pre-AUG A_N. This is exactly what has been observed in a previous *in vitro* experiment (Shirokikh and Spirin 2008) without PABP, where the translation-enhancing effect is greater for longer poly(A) than for shorter poly(A), *i.e.*, the ranking order of translation initiation efficiency is $A_{25} > A_{12} > A_5$.

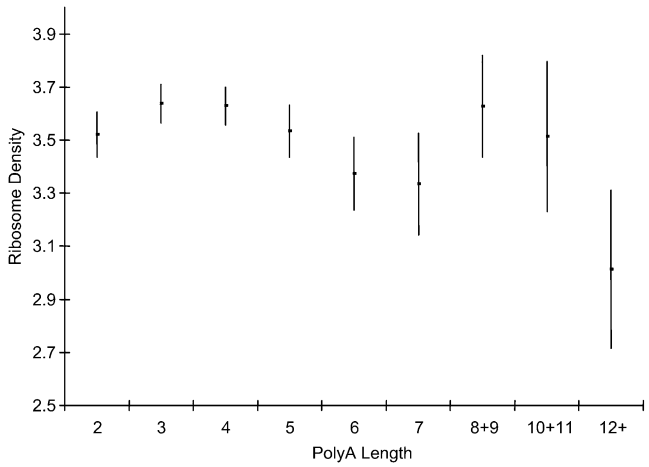

Figure 7 Relationship between ribosomal density (log-transformed normalized read density) and poly(A) length. The vertical bars show one standard error (SE) above and below the mean for each poly(A) length category. Some neighboring poly(A) length groups were lumped to reduce SE. Ribosomal density for genes in the "12+" group is significantly lower ($t = 2.1932$, DF $= 2911$, $P = 0.0284$, two-tailed test) than that of all other categories treated as one group.

Our results suggest an alternative to the dominant hypothesis concerning the function of PABP on translation initiation. Several studies have shown that exogenous poly (A) can inhibit translation initiation (Lodish and Nathan 1972; Jacobson and Favreau 1983; Grossi De Sa *et al.* 1988), and the inhibitive effect can be eliminated by the addition of PABP (Grossi De Sa *et al.* 1988; Gilbert *et al.* 2007). Similarly, noncoding poly(A) sequences, such as BC1 and BC200 RNA expressed in neurons, are known to bind PABP (Muddashetty *et al.* 2002) and to inhibit translation initiation when highly expressed (Wang *et al.* 2002, 2005; Kondrashov *et al.* 2005). The dominant hypothesis is that PABP, in addition to its function in mRNA stabilization and circularization, also serves as a translation initiation factor (Kahvejian *et al.* 2005; Khanam *et al.* 2006) that functions by binding to pre-AUG A_N. Thus, either exogenous poly(A) RNA or intrinsically produced poly(A) RNA such as BC1 and BC200 RNA that sequesters PABP would reduce translation initiation (Khanam *et al.* 2006; Gilbert *et al.* 2007). Consistent with this hypothesis, adding PABP eliminated the inhibitive effect of the exogenous poly(A) RNA (Grossi De Sa *et al.* 1988; Gilbert *et al.* 2007). The hypothesis also explains why poly(A) is over-represented in 5′-UTR in yeast genes, especially those highly expressed ones because such pre-AUG A_N would gain enhanced translation initiation by interacting with PABP. However, this hypothesis has three difficulties. First, it cannot explain why genes with a long pre-AUG A_N have a reduced protein synthesis rate as well as a reduced ribosomal density shown in this article. Second, it cannot explain why, in the complete absence of PABP, pre-AUG A_N can still enhance translation initiation for both capped and uncapped mRNA (Shirokikh and Spirin 2008). Third, it cannot explain why adding translation initiation factors eIF-4B and eIF-4F (including eIF-4A) in combination also eliminated the inhibitive effect of exogenous poly(A) RNA on translation initiation (Gallie and Tanguay 1994). Our new hypothesis is that pre-AUG A_N binds to translation initiation factors such as eIF-4B and eIF-4F to facilitate translation initiation. Exogenous or intrinsic poly (A) RNAs can inhibit translation initiation not only because

they compete for PABP but also because they would sequester the translation initiation factors eIF-4B and eIF-4F. Adding PABP can eliminate the inhibitive effect of exogenous poly (A) RNA because PABP would bind to the poly(A) and free the translation initiation factors sequestered by these poly(A) RNAs. This new hypothesis, which was implicitly proposed in a previous study (Gallie and Tanguay 1994) demonstrating the binding of poly(A) RNA to eIF-4B and eIF-4F, eliminates all three difficulties plaguing the other hypothesis.

Presence of a pre-AUG A_N appears to be a key feature in a set of internal ribosomal entry sites (IRESs) empirically verified in a recent study on yeast translation (Gilbert *et al.* 2007). All those poly(A) tracts are shorter than 12 consecutive A's. These include not only the genes involved in the invasive growth in the yeast, but also transcripts that are routinely transcribed and translated, such as *eIF-4G* and *Pab1* transcripts. The IRES activity mediated by the pre-AUG A_N does seem to require Pab1p (Gilbert *et al.* 2007). A recent study using mRNAs without a poly(A) tail (Kahvejian *et al.* 2005) suggests that PABP may serve as a translation initiation factor independent of its binding to the poly (A) tail. It is possible that the multiple PABP functions may depend on how strong it binds to pre-AUG A_N, with strong binding inhibiting translation and weak binding enhancing translation. It is also possible that the association between the IRES activity and the pre-AUG A_N is coincidental. A recent study on IRESs from both the yeast and *Drosophila melanogaster* shows that IRES activity increases consistently with decreasing stability of secondary structure (Xia and Holcik 2009). A pre-AUG poly(A) would contribute to a weak RNA secondary structure when nucleotide U usage is dramatically reduced (Figures 1–3).

While there is empirical evidence that mammalian PABP expression may be autoregulated by PABP binding to the pre-AUG A_N of its own mRNA (Wu and Bag 1998; Bag 2001; Ma *et al.* 2003a,b, 2006; Patel *et al.* 2005; Patel and Bag 2006; Bag and Bhattacharjee 2010), there is no evidence that Pab1p abundance in yeast is autoregulated. Pab1p abundance is high, being the top 39th among the 3841 yeast genes with characterized protein abundance (Ghaemmaghami *et al.* 2003). Its mRNA ranked the top 114th in ribosomal density among the 5164 yeast genes with ribosomal density characterized by Ingolia *et al.* (2009). Such a high protein abundance and a high ribosomal density is strong evidence that the high protein abundance in Pab1p does not interfere with the translation of its mRNA. If the autoregulation requires a pre-AUG A_{12}, then *Pab1* mRNA would escape autoregulation because it has only a pre-AUG A_{11}. The mammalian PAPB seems to be less strict on contiguity of poly(A), especially its RNA-recognition motif (RRM) 3+4 (Khanam *et al.* 2006).

Relevance to the translation of early and late genes in vaccinia virus

The finding that the length of pre-AUG A_N is strongly associated with ribosomal loading and protein synthesis sheds light on the evolutionary significance of the difference in the length of pre-AUG A_N between early and late genes in vaccinia virus. The early vaccinia viral genes have a pre-AUG A_N with 4–14 A residues (Ahn *et al.* 1990; Ink and Pickup 1990), but the poly(A) tracts in late genes are often around 35 A residues (Bertholet *et al.* 1987; Schwer *et al.* 1987; Schwer and Stunnenberg 1988). The early viral genes are translated in the presence of abundant PABP, which would repress the translation of mRNAs with a long pre-AUG A_N. This implies that the transcripts of the viral early genes should have only short poly(A) to avoid repression. In contrast, late viral genes are translated when the cellular protein production has been much reduced, *i.e.*, when PABP is expected to be less abundant. So mRNAs from late viral genes can have long pre-AUG A_N without suffering from translation repression mediated by PABP. It has been experimentally demonstrated that, in the absence of PABP, the translation enhancing effect of pre-AUG A_N increases with its length (Shirokikh and Spirin 2008).

There is some controversy concerning whether the PABP level is reduced during the infection cycle of vaccinia virus. The degradation of host mRNA appears nearly complete 6 hr after the viral infection as no host poly(A) mRNA is detectable at/after this time (Katsafanas and Moss 2007). Furthermore, a large-scale characterization of mRNA of HeLa cells infected with vaccinia virus (Yang *et al.* 2010) showed that PABP mRNA was reduced to 50% by 4 hr. Although no mRNA characterization is done after this time, intuition would suggest continued reduction, and such a suggestion is consistent with the finding that no host mRNA is detectable after 6 hr after the viral infection (Katsafanas and Moss 2007).

The study by Katsafanas and Moss (2007) also showed that the viral mRNAs are located in the cavities of viral factories (VFs), where they are transcribed and translated. A number of translation initiation factors such as eIF4E and eIF4G are also localized in these cavities (Katsafanas and Moss 2007; Walsh *et al.* 2008). In contrast, PABP is localized on the periphery of a VF (Walsh *et al.* 2008), which suggests that PABP does not participate in translation of the viral genes. It is known that vaccinia virus produces poly(A) nontranslated small RNA sequences that selectively inhibit cap-dependent translation of host messages (Bablanian and Banerjee 1986; Bablanian *et al.* 1986, 1987, 1993; Lu and Bablanian 1996), presumably by binding to PABP and preventing it from interacting with other translation initiation factors. Both Rubella virus and Bunyamwera virus inhibit translation of host genes by producing a capsid protein that binds to PABP and prevents it from binding to other translation initiation factors (Ilkow *et al.* 2008; Blakqori *et al.* 2009).

Walsh *et al.* (2008) found a persistent level of PABP during the infection cycle of vaccinia virus, but did not provide any evidence that PABP is in fact produced during the viral infection cycle. However, a subsequent paper (Perez *et al.* 2011) from the same laboratory found that PABP is

continuously synthesized in cytomegalovirus-infected cells. They suggested that this selective translation of PABP is through the mTOR+4E-BP pathway. However, this suggestion does not seem coherent. Preventing 4E-BP from binding to eIF-4E would lead to a general increase of cap-dependent translation, not selective translation of the *PABP* mRNA.

In summary, multiple lines of empirical evidence suggest that a pre-AUG A_N shorter than 12 may enhance translation in the yeast. However, yeast genes with a pre-AUG $A_{\geq 12}$ tend to be translated inefficiently with a low ribosomal density and output a reduced amount of protein, consistent with the interpretation that such long poly(A) tracts may bind to Pab1p, resulting in repression of translation.

Acknowledgments

We thank Y. Fang, M. Ragonnet, A. van Weringh, and S. Zhao for discussion, comments, and suggestions. B. Moss provided references on PABP and vaccinia virus. This study is supported by National Sciences and Engineering Research Council's Discovery, Research Tools and Instrument, and Strategic Research Grants to X.X.

Literature Cited

Ahn, B. Y., E. V. Jones, and B. Moss, 1990 Identification of the vaccinia virus gene encoding an 18-kilodalton subunit of RNA polymerase and demonstration of a 5′ poly(A) leader on its early transcript. J. Virol. 64: 3019–3024.

Arava, Y., Y. Wang, J. D. Storey, C. L. Liu, P. O. Brown *et al.*, 2003 Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 100: 3889–3894.

Aviram, S., E. Simon, T. Gildor, F. Glaser, and D. Kornitzer, 2008 Autophosphorylation-induced degradation of the Pho85 cyclin Pcl5 is essential for response to amino acid limitation. Mol. Cell. Biol. 28: 6858–6869.

Bablanian, R., and A. K. Banerjee, 1986 Poly(riboadenylic acid) preferentially inhibits in vitro translation of cellular mRNAs compared with vaccinia virus mRNAs: possible role in vaccinia virus cytopathology. Proc. Natl. Acad. Sci. USA 83: 1290–1294.

Bablanian, R., G. Coppola, P. S. Masters, and A. K. Banerjee, 1986 Characterization of vaccinia virus transcripts involved in selective inhibition of host protein synthesis. Virology 148: 375–380.

Bablanian, R., S. K. Goswami, M. Esteban, and A. K. Banerjee, 1987 Selective inhibition of protein synthesis by synthetic and vaccinia virus-core synthesized poly(riboadenylic acids). Virology 161: 366–373.

Bablanian, R., S. Scribani, and M. Esteban, 1993 Amplification of polyadenylated nontranslated small RNA sequences (POLADS) during superinfection correlates with the inhibition of viral and cellular protein synthesis. Cell. Mol. Biol. Res. 39: 243–255.

Bag, J., 2001 Feedback inhibition of poly(A)-binding protein mRNA translation. A possible mechanism of translation arrest by stalled 40 S ribosomal subunits. J. Biol. Chem. 276: 47352–47360.

Bag, J., and R. B. Bhattacharjee, 2010 Multiple levels of post-transcriptional control of expression of the poy (A)-binding protein. RNA Biol. 7: 5–12.

Bertholet, C., E. Van Meir, B. ten Heggeler-Bordier, and R. Wittek, 1987 Vaccinia virus produces late mRNAs by discontinuous synthesis. Cell 50: 153–162.

Blakqori, G., I. van Knippenberg, and R. M. Elliott, 2009 Bunyamwera orthobunyavirus S-segment untranslated regions mediate poly(A) tail-independent translation. J. Virol. 83: 3637–3646.

Coghlan, A., and K. H. Wolfe, 2000 Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. Yeast 16: 1131–1145.

Duret, L., and D. Mouchiroud, 1999 Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. Proc. Natl. Acad. Sci. USA 96: 4482–4487.

Gallie, D. R., and R. Tanguay, 1994 Poly(A) binds to initiation factors and increases cap-dependent translation in vitro. J. Biol. Chem. 269: 17166–17173.

Ghaemmaghami, S., W. K. Huh, K. Bower, R. W. Howson, A. Belle *et al.*, 2003 Global analysis of protein expression in yeast. Nature 425: 737–741.

Gilbert, W. V., K. Zhou, T. K. Butler, and J. A. Doudna, 2007 Cap-independent translation Is required for starvation-induced differentiation in yeast. Science 317: 1224–1227.

Grossi de Sa, M. F., N. Standart, C. Martins de Sa, O. Akhayat, M. Huesca *et al.*, 1988 The poly(A)-binding protein facilitates in vitro translation of poly(A)-rich mRNA. Eur. J. Biochem. 176: 521–526.

Holstege, F. C., E. G. Jennings, J. J. Wyrick, T. I. Lee, C. J. Hengartner *et al.*, 1998 Dissecting the regulatory circuitry of a eukaryotic genome. (Transcriptomic data at http://web.wi.mit.edu/young/pub/data/orf_transcriptome.txt). Cell 95: 717–728.

Ilkow, C. S., V. Mancinelli, M. D. Beatch, and T. C. Hobman, 2008 Rubella virus capsid protein interacts with poly(a)-binding protein and inhibits translation. J. Virol. 82: 4284–4294.

Ingolia, N. T., S. Ghaemmaghami, J. R. S. Newman, and J. S. Weissman, 2009 Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324: 218–223.

Ink, B. S., and D. J. Pickup, 1990 Vaccinia virus directs the synthesis of early mRNAs containing 5′ poly(A) sequences. Proc. Natl. Acad. Sci. USA 87: 1536–1540.

Jacobson, A., and M. Favreau, 1983 Possible involvement of poly (A) in protein synthesis. Nucleic Acids Res. 11: 6353–6368.

Kahvejian, A., Y. V. Svitkin, R. Sukarieh, M. N. M'Boutchou, and N. Sonenberg, 2005 Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. Genes Dev. 19: 104–113.

Katsafanas, G. C., and B. Moss, 2007 Colocalization of transcription and translation within cytoplasmic poxvirus factories coordinates viral expression and subjugates host functions. Cell Host Microbe 2: 221.

Khanam, T., R. S. Muddashetty, A. Kahvejian, N. Sonenberg, and J. Brosius, 2006 Poly(A)-binding protein binds to A-rich sequences via RNA-binding domains 1+2 and 3+4. RNA Biol. 3: 170–177.

Kondrashov, A. V., M. Kiefmann, K. Ebnet, T. Khanam, R. S. Muddashetty *et al.*, 2005 Inhibitory effect of naked neural BC1 RNA or BC200 RNA on eukaryotic in vitro translation systems is reversed by poly(A)-binding protein (PABP). J. Mol. Biol. 353: 88–103.

Lawless, C., R. Pearson, J. Selley, J. Smirnova, C. Grant *et al.*, 2009 Upstream sequence elements direct post-transcriptional regulation of gene expression under stress conditions in yeast. BMC Genomics 10: 7.

Lodish, H. F., and D. G. Nathan, 1972 Regulation of hemoglobin synthesis. Preferential inhibition of and globin synthesis. J. Biol. Chem. 247: 7822–7829.

Lu, C., and R. Bablanian, 1996 Characterization of small nontranslated polyadenylylated RNAs in vaccinia virus-infected cells. Proc. Natl. Acad. Sci. USA 93: 2037–2042.

Ma, S., T. Musa, and J. Bag, 2006 Reduced stability of mitogen-activated protein kinase kinase-2 mRNA and phosphorylation of poly(A)-binding protein (PABP) in cells overexpressing PABP. J. Biol. Chem. 281: 3145–3156.

MacKay, V. L., X. Li, M. R. Flory, E. Turcott, G. L. Law *et al.*, 2004 Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. Mol. Cell. Proteomics 3: 478–489.

Melo, E. O., O. P. de Melo Neto, and C. Martins de Sa, 2003a Adenosine-rich elements present in the 5′-untranslated region of PABP mRNA can selectively reduce the abundance and translation of CAT mRNAs in vivo. FEBS Lett. 546: 329–334.

Melo, E. O., R. Dhalia, C. Martins de Sa, N. Standart, and O. P. de Melo Neto, 2003b Identification of a C-terminal poly(A)-binding protein (PABP)-PABP interaction domain: role in cooperative binding to poly (A) and efficient cap distal translational repression. J. Biol. Chem. 278: 46357–46368.

Miura, F., N. Kawaguchi, J. Sese, A. Toyoda, M. Hattori *et al.*, 2006 A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. Proc. Natl. Acad. Sci. USA 103: 17846–17851.

Miura, F., N. Kawaguchi, M. Yoshida, C. Uematsu, K. Kito *et al.*, 2008 Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs. BMC Genomics 9: 574.

Muddashetty, R., T. Khanam, A. Kondrashov, M. Bundman, A. Iacoangeli *et al.*, 2002 Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles. J. Mol. Biol. 321: 433–445.

Patel, G. P., and J. Bag, 2006 IMP1 interacts with poly(A)-binding protein (PABP) and the autoregulatory translational control element of PABP-mRNA through the KH III–IV domain. FEBS J. 273: 5678–5690.

Patel, G. P., S. Ma, and J. Bag, 2005 The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex. Nucleic Acids Res. 33: 7074–7089.

Perez, C., C. McKinney, U. Chulunbaatar, and I. Mohr, 2011 Translational control of the abundance of cytoplasmic poly(A) binding protein in human cytomegalovirus-infected cells. J. Virol. 85: 156–164.

Sachs, A. B., R. W. Davis, and R. D. Kornberg, 1987 A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. Mol. Cell. Biol. 7: 3268–3276.

Schwer, B., and H. G. Stunnenberg, 1988 Vaccinia virus late transcripts generated in vitro have a poly(A) head. EMBO J. 7: 1183–1190.

Schwer, B., P. Visca, J. C. Vos, and H. G. Stunnenberg, 1987 Discontinuous transcription or RNA processing of vaccinia virus late messengers results in a 5′ poly(A) leader. Cell 50: 163–169.

Serikawa, K. A., X. L. Xu, V. L. MacKay, G. L. Law, Q. Zong *et al.*, 2003 The transcriptome and its translation during recovery from cell cycle arrest in Saccharomyces cerevisiae. Mol. Cell. Proteomics 2: 191–204.

Shabalina, S. A., A. Y. Ogurtsov, I. B. Rogozin, E. V. Koonin, and D. J. Lipman, 2004 Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals. Nucleic Acids Res. 32: 1774–1782.

Sharp, P. M., and W. H. Li, 1987 The codon adaptation index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15: 1281–1295.

Shirokikh, N. E., and A. S. Spirin, 2008 Poly(A) leader of eukaryotic mRNA bypasses the dependence of translation on initiation factors. Proc. Natl. Acad. Sci. USA 105: 10738–10743.

van Weringh, A., M. Ragonnet-Cronin, E. Pranckeviciene, M. Pavon-Eternod, L. Kleiman *et al.*, 2011 HIV-1 modulates the tRNA pool to improve translation efficiency. Mol. Biol. Evol. 28: 1827–1834.

Walsh, D., C. Arias, C. Perez, D. Halladin, M. Escandon *et al.*, 2008 Eukaryotic translation initiation factor 4F architectural alterations accompany translation initiation factor redistribution in poxvirus-infected cells. Mol. Cell. Biol. 28: 2648–2658.

Wang, H., A. Iacoangeli, S. Popp, I. A. Muslimov, H. Imataka *et al.*, 2002 Dendritic BC1 RNA: functional role in regulation of translation initiation. J. Neurosci. 22: 10232–10241.

Wang, H., A. Iacoangeli, D. Lin, K. Williams, R. B. Denman *et al.*, 2005 Dendritic BC1 RNA in translational control mechanisms. J. Cell Biol. 171: 811–821.

Wu, J., and J. Bag, 1998 Negative control of the poly(A)-binding protein mRNA translation is mediated by the adenine-rich region of its 5′-untranslated region. J. Biol. Chem. 273: 34535–34542.

Xia, X., 1998 How optimized is the translational machinery in Escherichia coli, Salmonella typhimurium and Saccharomyces cerevisiae? Genetics 149: 37–44.

Xia, X., 2001 *Data Analysis in Molecular Biology and Evolution*. Kluwer Academic Publishers, Boston.

Xia, X., 2005 Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. Gene 345: 13–20.

Xia, X., 2007 An improved implementation of codon adaptation index. Evol. Bioinform. 3: 53–58.

Xia, X., and M. Holcik, 2009 Strong eukaryotic IRESs have weak secondary structure. PLoS ONE 4: e4136.

Xia, X., and Z. Xie, 2001 DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. 92: 371–373.

Yang, Z., D. P. Bruno, C. A. Martens, S. F. Porcella, and B. Moss, 2010 Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing. Proc. Natl. Acad. Sci. USA 107: 11513–11518.

Zhang, Z., and F. S. Dietrich, 2006 Mapping of transcription start sites in Saccharomyces cerevisiae using 5′ SAGE. Nucleic Acids Res. 33: 2838–2851.

Communicating editor: K. M. Arndt

GENETICS

Supporting Information
http://www.genetics.org/content/suppl/2011/08/12/genetics.111.132068.DC1

Translation Initiation: A Regulatory Role for Poly(A) Tracts in Front of the AUG Codon in *Saccharomyces cerevisiae*

Xuhua Xia, Vivian MacKay, Xiaoquan Yao, Jianhua Wu, Fumihito Miura, Takashi Ito, and David R. Morris

DOI: 10.1534/genetics.111.132068

File S1

Concordance among MacKayData, AravaData and IngoliaData

MacKayData and AravaData are highly concordant, which is expected because the two data sets were collected using the same technology of translation state array analysis (TSAA). The AravaData contains 5701 genes of which 2936 genes have their 5' UTR determined by Miura *et al.* (2006). Both data sets have the ribosomal density reaching low values for genes with poly(A) length equal to or greater than 12. However, a number of short CDSs included in AravaData have low ribosomal density and are annotated as putative genes. These genes were mostly excluded in MacKayData, so the average ribosomal density values are significantly higher in MacKayData than in AravaData.

What is remarkable is the concordance between MacKayData/AravaData and IngoliaData, with the two datasets collected by dramatically different methodologies. In MacKay et al. (2004), the mRNA molecules are fractionated into mRNAs with different numbers of ribosomes attached. These mRNAs are then subject to microarray so that the mRNA for each gene can be identified. The average number of ribosomes attached to each mRNA (say M) can then be estimated and the ribosomal density for each mRNA is obtained by dividing M by the mRNA length. The ribosomal density in Ingolia et al. (2009) is obtained by a ribosome-profiling strategy based on deep sequencing. In short, each ribosome bound to mRNA protects about 30 bases of its bound mRNA from being digested by nuclease. These 30 bases were then sequenced and aligned to its source gene (except for paralogous genes with identical sequences, which fortunately are few in the yeast). The ribosomal density for each mRNA may then be estimated as the number of such 30mers divided by the mRNA length (Their actual normalized ribosomal density is based on the number of 30mers in the first 151 codons). This method has the potential advantage of identify mRNAs which may have an efficient 5'UTR for loading ribosomes but the ribosomes are not able to move down to the coding part, e.g., when there are uORFs in the 5' UTR of certain mRNAs.

There are 3596 genes (MacKayIngoliaShared.xls in Supplemental Materials) with the predicted protein synthesis rate from MacKay et al. (2004) and the normalized ribosome density from Ingolia et al. (2009), allowing an assessment of the concordance between the two data sets. The high concordance between the two data sets (Fig. 1) suggests that the large-scale experimental characterization of translation activity by ribosomal profiling has come of age.

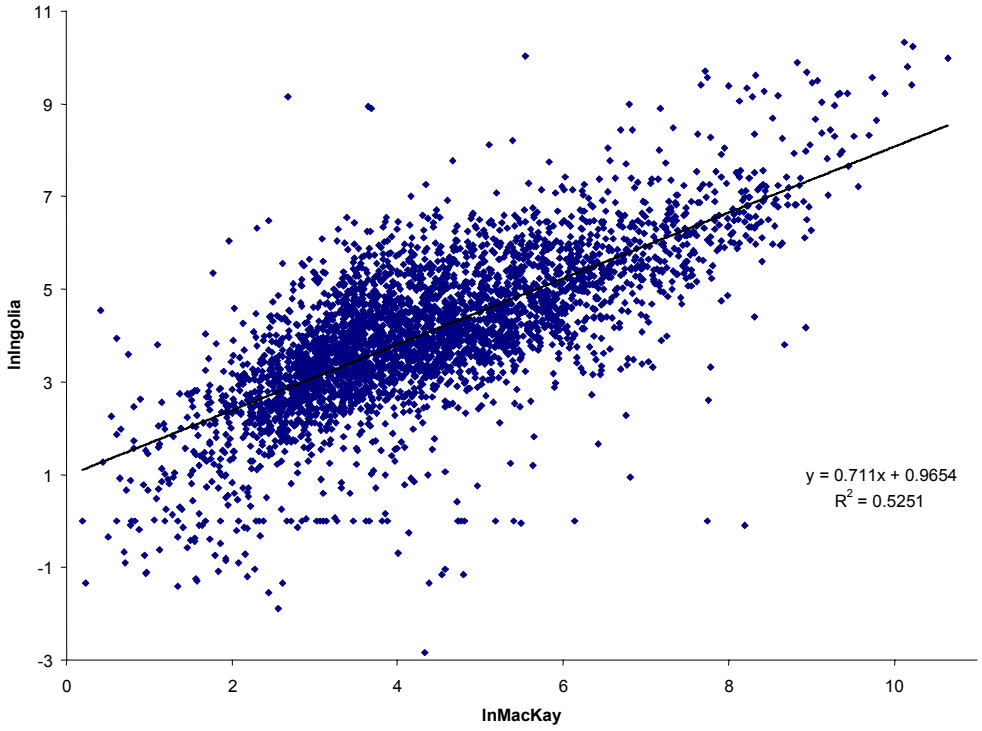

Fig. 1. Concordance between the MacKayData (log-transformed predicted protein synthesis rate) and IngoliaData (log-transformed normalized read density). Based on 3596 yeast genes with ribosome density characterized in both studies (see Supplementary online data for the actual data from which the graph is generated).

Files S2-S5

Files S2-S5 are available for download at http://www.genetics.org/content/suppl/2011/08/12/genetics.111.132068.DC1.

File S2 MacKayIngoliaShared.xls: Data used to generate the figure in FileS1.pdf
File S3 AllData.xls: Data for replicating tables and figures in the manuscript
File S4 GhaemmaghamiProtein.xls: Protein abundance from Ghaemmaghami et al. 2003
File S5 MiuraTSS.FAS: 5' UTR sequences extracted based on Miura et al. 2006, in FASTA file