

Chapter 3

From transcription to translation: global translational properties of fission yeast mRNAs and integration with other genome-wide data sets on gene expression

From transcription to translation: global translational properties of fission yeast mRNAs and integration with other genome-wide data sets on gene expression

This chapter will provide a global view of translational efficiency of mRNAs in vegetatively growing fission yeast cells measured by translational profiling. Furthermore, other genome-wide data sets on various aspects of gene expression regulation such as mRNA steady-state levels, poly(A) tail length of mRNAs, mRNA half-lives and transcriptional efficiency and the connections between these diverse layers of gene expression regulation will be presented. Data from this chapter have been published in *Molecular Cell* (Lackner et al. 2007).

Introduction

It is important to recognize that gene expression can be regulated at multiple levels, and cells need to coordinate different regulatory processes to function properly. Transcriptional rates, mRNA features such as poly(A) tail length, association with RNA-binding proteins and mRNA half-lives as well as translational rates all make a contribution to regulating gene expression in the cell (Hieronymus and Silver 2004; Mata et al. 2005). There is increasing evidence that these processes at the complex interplay between DNA, RNA and the regulatory apparatus are integrated with each other (Maniatis and Reed 2002; Orphanides and Reinberg 2002; Proudfoot et al. 2002; Moore 2005). Most data supporting this idea have been generated through numerous in-depth studies focussing on single genes which are regulated at several levels.

Complementary to the single gene approach, large-scale approaches have given us new insights into gene expression regulation from a genome-wide perspective. For many of these studies, microarrays were mostly used to measure mRNA steady-state levels for expression profiling (Lockhart and Winzeler 2000). Recently, sophisticated variations of microarray-based approaches made genome-wide measurements of additional aspects of gene expression possible (Hieronymus and Silver 2004; Mata et al. 2005). Many of these approaches were pioneered in the budding yeast *Saccharomyces cerevisiae*. Examples include genome-wide studies on mRNA half-

lives (Wang et al. 2002; Grigull et al. 2004), RNA-binding proteins (Gerber et al. 2004), and translation (Arava et al. 2003; Preiss et al. 2003; MacKay et al. 2004). These global data sets provide supplementary and unique views on specific aspects of gene expression and allow the discovery of unexpected connections; for example, translational profiling revealed that long mRNAs are less densely associated with ribosomes than are short mRNAs (Arava et al. 2003).

While traditional studies can address multiple aspects of regulation for one or a few genes, genome-wide studies typically are restricted to one aspect of regulation. It is not clear to what degree different regulatory levels of gene expression are interconnected at a global scale, and whether any global patterns are conserved during evolution. For a comprehensive understanding of gene expression, it will be important to obtain and integrate global data sets covering as many regulatory aspects as possible, given that the cell itself regulates and coordinates multiple levels of gene expression.

This work was aimed at gaining insights into key aspects of gene expression in the fission yeast *Schizosaccharomyces pombe* on a genome-wide scale with a focus on translation. To this end, a detailed analysis of genome-wide translational properties was complemented by a range of other large-scale data for context and comparisons. Besides using established methods to determine translational profiles, mRNA steady-state levels and mRNA half-lives, novel microarray-based approaches were applied to estimate poly(A) tail lengths and transcription rates. The integrated analyses further incorporated publicly available data on *S. pombe* ORF lengths (Wood et al. 2002) and protein levels (Matsuyama et al. 2006). The systematic and quantitative data sets from this multi-dimensional approach, all acquired using a standardized growth condition and coherent methodology, helped to uncover global connections and trends that would not be apparent from studies involving only a few genes, and they revealed remarkably widespread relationships between multiple layers of gene expression. Furthermore, these data provided us with a valuable basis for the measurement of changes in translation efficiency in cells under different conditions such as after the exposure to stress (see Chapters 4 and 5).

Establishing polysome fractionation

To obtain data on translation in fission yeast, we wanted to determine the association of ribosomes with mRNAs on a global scale. To this end, polysome profiling combined with microarray analysis was used: Cycloheximide is added to the cells, which blocks translation and "freezes" the ribosomes on the mRNA. Cell lysates are then subjected to ultracentrifugation on a sucrose gradient to resolve mRNA-ribosome particles according to density, which corresponds to the number of bound ribosomes. The gradient is then fractionated by upward displacement with 55% sucrose and RNA absorbance is measured at 254 nm. RNA can then be extracted from the collected fractions, which correspond to mRNAs with increasing numbers of bound ribosomes, and probed on microarrays.

At the early stages of this study, 5-45% sucrose gradients were used to fractionate mRNA-ribosome particles and ultracentrifuge runs were done at 39,000 rpm for 160 min. A representative polysome profile is depicted in Figure 3.1A. To confirm that the peaks in Figure 3.1A correspond to the indicated ribosomal subunits (40S, 60S), monosome (80S) and polysome fractions (2 and more bound ribosomes), the distribution of ribosomal RNA (rRNA) along the profile was determined.

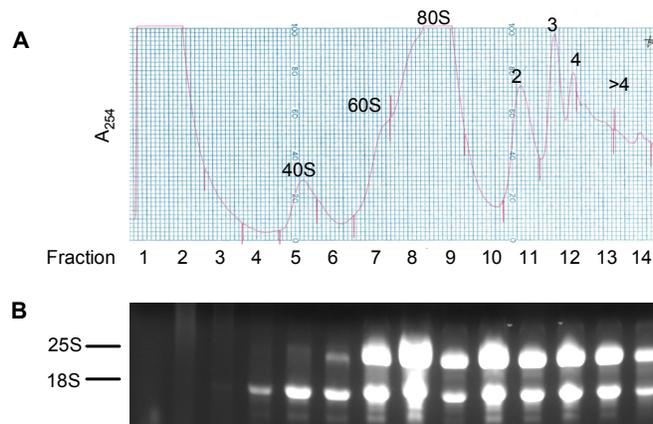


Figure 3.1 Polysome profile of ribosomes isolated from *S. pombe* and resolved by velocity sedimentation through a 5-45% sucrose gradient

(A) The positions of free small (40S) and large (60S) ribosomal subunits, monosomes (80S), and polysomes (2–4 ribosomes and above) are indicated in the profile.

(B) RNA was extracted from each fraction and an aliquot of each fraction was resolved using a 1% agarose gel and stained with ethidium bromide. The 2 most prominent bands correspond to 25S and 18S ribosomal RNA.

RNA was extracted from 14 fractions equally spaced along the profile (see Chapter 2), and equal amounts of RNA from each fraction were loaded onto a 1% agarose gel and stained with ethidium bromide, which is sufficient to visualize rRNAs given that they are the most abundant RNA species in growing cells. The occurrence of 18S RNA, which is part of the 40S ribosomal subunit, and 25S RNA, which is part of the 60S ribosomal subunit corresponded well with the peaks in the polysome profile (Figure 3.1B).

It was also important to confirm that mRNA was efficiently extracted from these fractions and that the amount of a given mRNA in the individual fractions would correspond to its translational efficiency. Actin is a very abundant protein in the cell (Futcher et al. 1999; Lu et al. 2007), and it could be assumed that *actin* mRNA is efficiently translated and most of its mRNA should be associated with the heavy polysome fractions. To test this, aliquots of mRNA extracted from each fraction were resolved on a 1% formaldehyd-agarose gel and analyzed by Northern blotting with a probe specific for *act1* mRNA (see Chapter 2). As expected, most *act1* mRNA was associated with fractions 12-14 of the polysome profile (Figure 3.2).

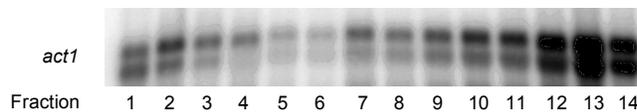


Figure 3.2 Association of *actin* mRNA across the polysome profile

Aliquots of each fraction were resolved on a 1% formaldehyd-agarose gel and analysed by Northern Blotting with a probe specific for actin. Most *act1* mRNA could be found associated with polysomal fractions. The 2 bands of different size probably belong to 2 transcripts with alternative 3' UTRs (Mertins and Gallwitz 1987).

In the 5-45% sucrose gradients, we were able to obtain single peak resolution of up to 4 ribosomes bound to mRNA in the polysome fractions (Figure 3.1A, Figure 3.3 top panel). To enhance this resolution, we used various gradients with differing sucrose concentrations for the polysome profiling and also varied the time and speed for the ultracentrifugation step. Best results were obtained using a 10-50% sucrose gradient and running the gradients at 35,000 rpm for 160 min. Using these conditions, we could resolve up to 8 ribosomes bound to mRNA as a singleton peak (Figure 3.3). Thus, unless otherwise indicated, polysome profiling in this study was done using these conditions.

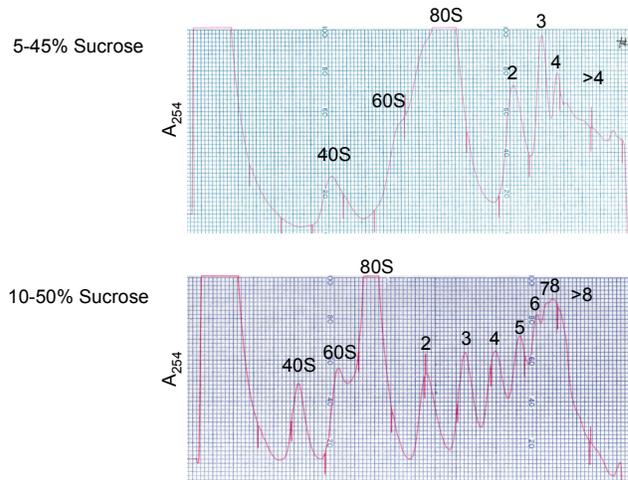


Figure 3.3 Comparison of polysome profiles obtained using sucrose gradients with different concentrations

Polysome profile of ribosomes isolated from *S. pombe* and resolved by velocity sedimentation through a 5-45% sucrose gradient run at 39,000 rpm for 160 minutes (top panel) or a 10-50% sucrose gradient run at 35,000 rpm for 160 minutes (bottom panel). The positions of free small (40S) and large (60S) ribosomal subunits, monosomes (80S), and polysomes (2–8 ribosomes and above) are indicated in the profile.

Genome-wide translational profiling

To determine the translational characteristics of mRNAs in fission yeast at a genome-wide scale, we prepared polysome profiles and hybridized microarrays with twelve mRNA fractions representing different numbers of associated ribosomes (Figure 3.4A). Normalization of the microarray data was done based on spiked-in bacterial mRNAs to correct for different RNA amounts in each fraction (see Chapter 2). Using this approach, we obtained high-resolution translational data for vegetative *S. pombe* cells growing exponentially in minimal medium at 32°C.

Figure 3.4B provides examples of translation profiles from three independently repeated experiments. There was high reproducibility between these experiments. We verified that transcripts peaked in the expected fractions. For instance, the non-coding *rrk1* RNA (RNase P K-RNA; Krupp et al. 1986) peaked in fraction 2, reflecting an absence of associated ribosomes as expected for an RNA that is not translated. The *fbal* mRNA, encoding fructose-bisphosphate aldolase, peaked in fraction 11, reflecting an association with many ribosomes for most of the mRNA. Consistent with this, Fba1p is highly expressed and within the top 1% with respect to protein

levels (Hwang et al. 2006). The 78-nucleotide *rpl4101* is the shortest mRNA in *S. pombe* and is therefore not expected to be associated with many ribosomes; accordingly, it peaked around fraction 6, which corresponds to the binding of a single ribosome (Figure 3.4A,B). *Actin* mRNA showed a similar distribution throughout the fractions measured using microarrays (Figure 3.4B) or Northern blotting (Figure 3.2). Furthermore, these profiles obtained by microarrays corresponded well with independent profiles obtained by quantitative PCR in another study (Bachand et al. 2006).

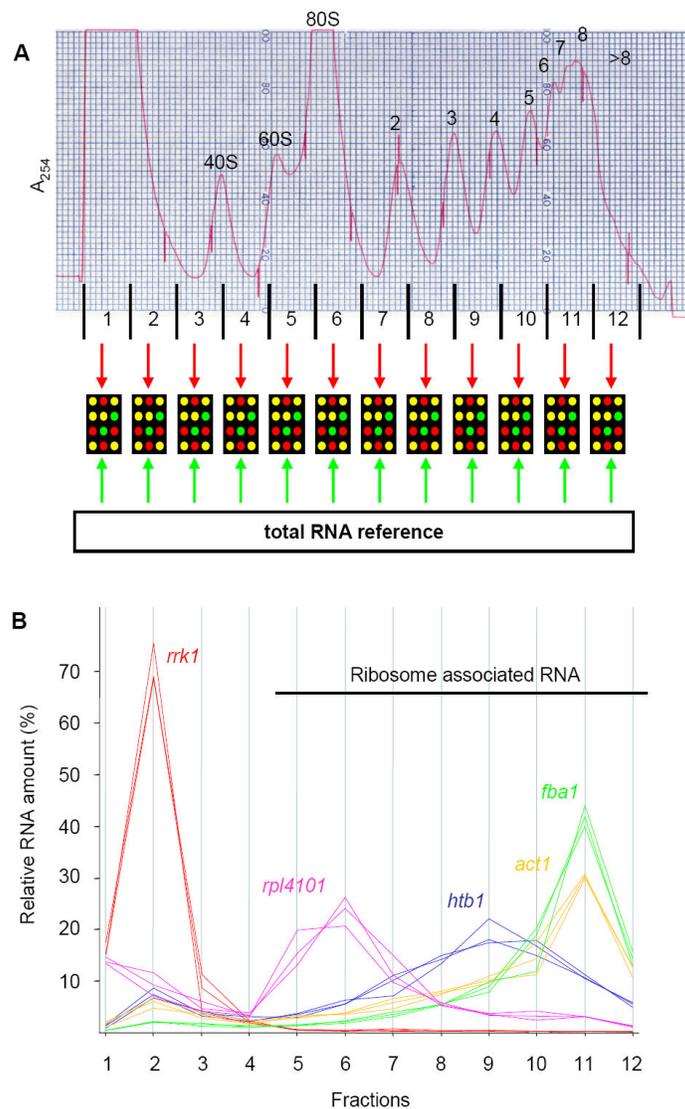


Figure 3.4 High-resolution polysome profiling

(A) Polysome profile of ribosomes isolated from vegetatively growing *S. pombe* cells and resolved by velocity sedimentation through a 10-50% sucrose gradient. The positions of free small (40S) and large (60S) ribosomal subunits, monosomes (80S), and polysomes (2–8 ribosomes and above) are indicated in the profile. RNA extracted from 12 fractions equally

spaced throughout the profile (bottom) was labelled and hybridized against a total RNA reference on microarrays containing all *S. pombe* genes.

(B) Translation profiles for selected transcripts obtained by microarray analysis, showing the relative RNA amounts for a given transcript contained in each of the 12 fractions. Fractions associated with ribosomes are indicated. Different transcripts are colour-coded, and polysome profiles from three independent biological repeats are shown for *rrk1* (RNase P K-RNA), *rpl4101* (encoding ribosomal protein), *htb1* (encoding histone H2B), *fba1* (encoding fructose-biphosphate aldolase) and *act1* (encoding actin).

Figure 3.5A shows average translational profiles for selected groups of transcripts. The profile of all mRNAs that provided translational data showed a peak in fraction 3 (reflecting free mRNA) along with a broad peak covering fractions 7-11 (reflecting polysomes of different sizes). Introns that were included on the microarrays peaked in fraction 3, which is not associated with ribosomes, as expected given that translation occurs on spliced mRNA. Conversely, mRNAs associated with Gene Ontology (GO) terms for translational regulation were associated with many ribosomes as expected for these highly expressed genes (Hwang et al. 2006). A list of 377 mRNAs encoding secreted proteins, which are translated on the endoplasmic reticulum membrane, showed an almost identical average translation profile to the one for all mRNAs (Figure 3.5B), indicating that the ribosome distribution for this specialized group is similar.

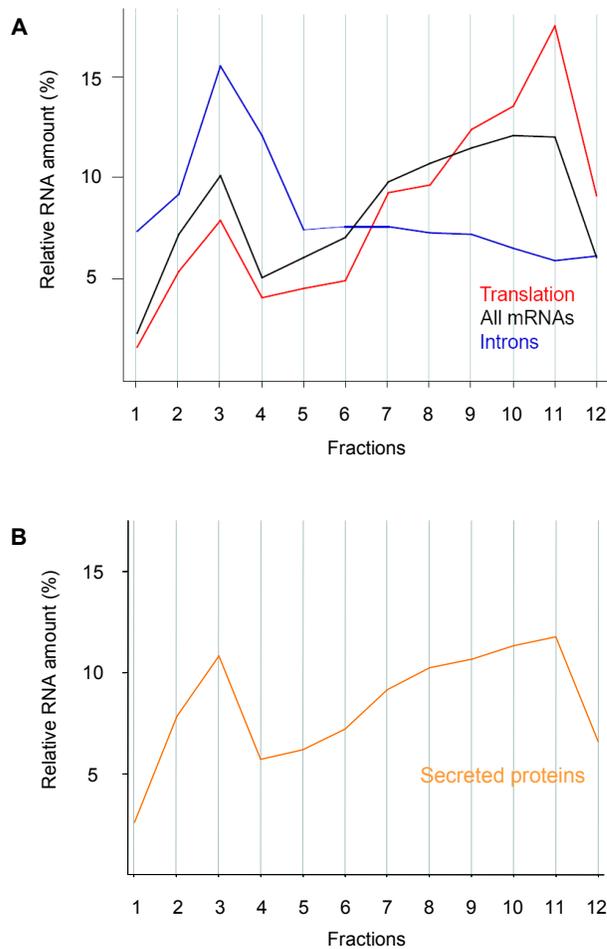


Figure 3.5 Average translation profiles for selected groups of RNAs

(A) Average translation profiles for selected groups of RNAs, plotted as in Fig. 3.4B for one experiment. All mRNAs, the 3505 high-confidence mRNAs with complete profiles in this experiment; Introns, 11 long introns included on the microarray; and Translation, 62 mRNAs associated with the GO terms “translational initiation,” “translational elongation,” or “translational termination”.

(B) Average translation profiles from the same experiment as in **(A)** for 377 mRNAs encoding secreted proteins.

Global translational properties of mRNAs

Although polysome profiles for almost all mRNAs were obtained, for further analysis we focussed on a conservative, high-confidence set of 3598 (72.5%) out of the 4962 nuclear encoded protein-coding genes. For a mRNA to be included in the high-confidence it had to fulfil the following criteria: (1) microarray data had to be available for at least 2 out of the 3 experiments for all 12 fractions, and (2) there had to be a minimum correlation of the profiles from different repeats (see Chapter 2 for details). Most of the excluded mRNAs were not or only weakly expressed under the condition used, which could be seen by looking at the distribution of these excluded genes according to relative expression levels measured using Affymetrix chips (Figure 3.6). These genes were also most enriched for GO terms related to meiosis ($P < 4e^{-24}$). From the translation profiles of the mRNAs included in the analysis, we determined different properties reflecting translational efficiency (see Chapter 2); these data on translational properties described below are provided in supplementary Table S1, which can be downloaded from http://www.sanger.ac.uk/PostGenomics/S_pombe/projects/translation/.

Ribosome occupancy indicates the percentage of a given type of mRNA that is associated with one or more ribosomes as opposed to free mRNA. The average ribosome occupancy was 77.3% with a relatively small standard deviation (SD) of 7.0%. This suggests that during exponential growth the majority of high-confidence mRNAs from most genes are engaged in translation, although a substantial fraction of >20% of mRNAs is not associated with any ribosomes.

The mean number of ribosomes bound to a given mRNA was calculated based on a weighted average by using the relative amount of the mRNA associated with each fraction and the number of ribosomes corresponding to that fraction. Only fractions associated with ribosomes were included for this (Figure 3.4A, B; fractions 5-12) so that the mean ribosome number is independent of ribosome occupancy. On average, 4.1 ribosomes were associated with mRNAs with a surprisingly small SD of 0.6. If the mRNAs not associated with ribosomes were also taken into account, this value was lowered to 3.6 ribosomes. As expected, the mean number of associated ribosomes generally increased as a function of open reading frame (ORF) length (Figure 3.7A), but this correlation breaks down for mRNAs with a length of over ~1.2 kilobases (kb)

and is more pronounced for shorter mRNAs such as mRNAs encoding ribosomal proteins (Figure 3.7B).

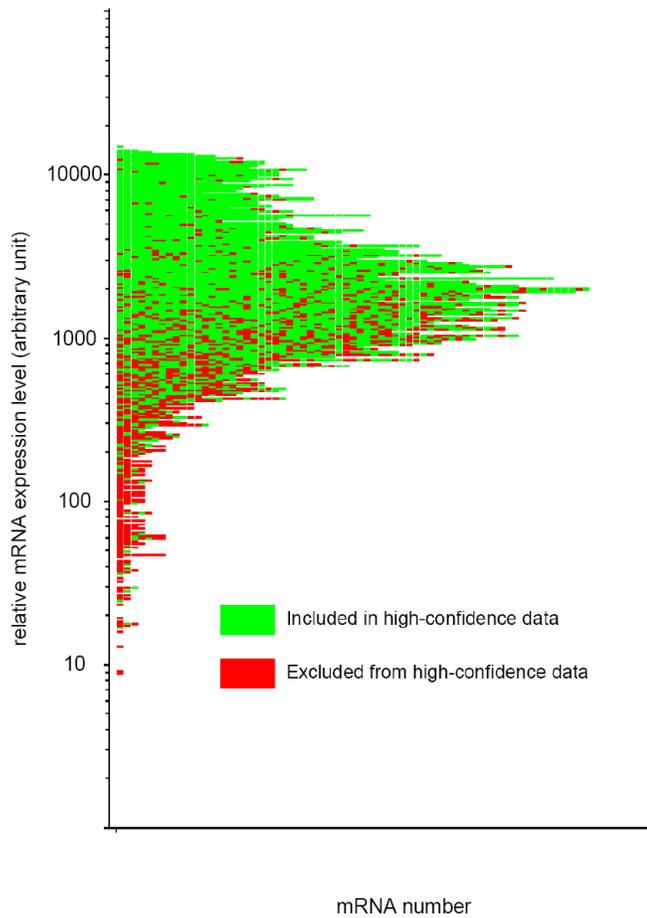


Figure 3.6 Distribution of mRNA levels for protein-coding genes included or excluded from high-confidence translational profiling data

Histogram showing the mRNA levels for 4818 protein-coding genes that provided signal data on Affymetrix chips. The average of two independent experiments is shown. Green: genes included in the high-confidence data set from the translational profiling experiments (3567 genes with measurable chip signals). Red: genes not included in the high-confidence data set (1251 genes with measurable chip signals).

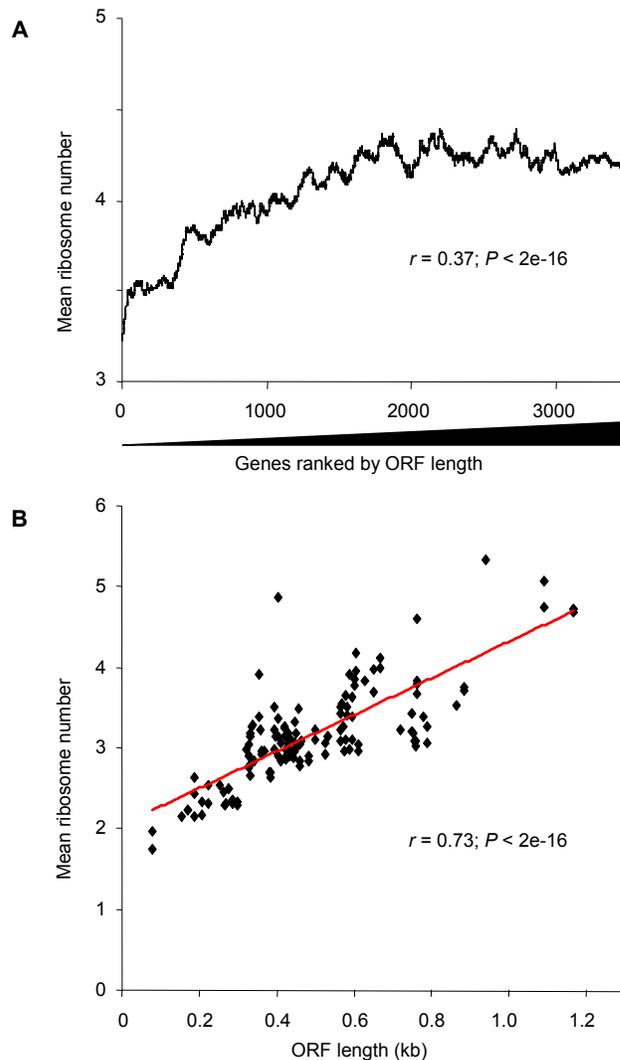


Figure 3.7 Correlation between ORF length and mean number of associated ribosomes
(A) Graph showing moving averages (100-gene window) of mean ribosome number as a function of genes ranked by ORF length ($n = 3598$), along with the corresponding Spearman rank correlation.
(B) Scatter plot of ORF length against mean number of associated ribosomes for mRNAs encoding ribosomal proteins ($n = 134$). The red line represents the linear trend-line for this correlation. The corresponding Spearman rank correlation between ORF length and mean ribosome number is also shown.

Arguably, the ribosome density is a better measure than the mean ribosome number to estimate translational efficiency as it normalizes for different mRNA lengths that influence the numbers of bound ribosomes (Figure 3.7A) (Arava et al. 2003; Beyer et al. 2004). The average ribosome density for all mRNAs was 4.5 ribosomes per kilobase of ORF, with a relatively large SD of 3.1 ribosomes per kb. On average, the

mRNAs thus contained one ribosome roughly every 222 nucleotides. Given that a eukaryotic ribosome occupies ~35 nucleotides of mRNA (Wolin and Walter 1988), the average density determined here is only about 1/6 of the maximal packing density. This is consistent with initiation being the rate-limiting factor during translation.

The sequence context of the AUG start codon influences the rate of translational initiation (Kozak 1991). To corroborate that high ribosome occupancy and density in our data reflect efficient translational initiation rather than slow elongation or ribosome stalling, we compared ribosome occupancy and density with the "AUG context adaptation index" (AugCAI), a measure for the effectiveness of the AUG context to promote translational initiation (Miyasaka 1999; Miyasaka 2002). Data for the AugCAI in fission yeast were calculated by Samuel Marguerat. Basically, mRNAs are assigned a score depending on the overlap with a consensus sequence around the AUG start codon derived from the 100 most abundant mRNAs, with a higher score for higher translation initiation efficiency. This analysis provided a consensus sequence for optimal translational initiation in *S. pombe* and revealed significant correlations between the AugCAI on one hand and ribosome occupancy and density on the other (Figure 3.8). This provides independent evidence that the translational profiling data are measures of translational efficiency.

We next looked for poorly and strongly translated mRNAs. Of the 3598 high-confidence mRNAs, only 57 showed ribosome occupancies of less than 60%, and 99 showed densities of less than one ribosome/kb on average. Just one mRNA (*urb2*, predicted role in ribosome biogenesis) was present in both of these groups. The 20% of mRNAs showing the highest ribosome occupancy were most enriched for transcripts repressed during stress ($P \sim 8e^{-30}$; Chen et al. 2003) and for those associated with the GO terms 'metabolism' and 'biosynthesis' ($P \sim 1e^{-30}$ - $2e^{-31}$). The 20% of mRNAs with the lowest ribosome occupancy were diverse and showed no strong enrichment for any particular GO terms or functional groups. The 20% of mRNAs showing the highest ribosome density were most enriched for GO terms such as 'ribosome', 'organelle', and several terms related to mitochondria ($P \sim 1e^{-12}$ - $7e^{-42}$) and for transcripts containing introns ($P \sim 5e^{-17}$), which is notable given that introns can enhance translation in mammals (Nott et al. 2003). The 20% of mRNAs showing the lowest ribosome density were most enriched for genes with GO terms such as 'ATP-binding', 'hydrolase activity', 'signal transduction', and related terms ($P \sim 2e^{-10}$ - $4e^{-26}$). The mRNAs with low ribosome density were also strongly enriched for the

Short mRNAs are more efficiently translated

Whereas the mean ribosome numbers varied by less than 4 fold (1.8 to 6.8 ribosomes per mRNA), the ORF lengths varied more than 180 fold (78 to 14154 bp; Wood et al. 2002). Accordingly, the ribosome numbers showed only modest increase relative to ORF length, and they did not increase above ~ 4.3 ribosomes on average for mRNAs longer than ~ 1200 bp (Figure 3.7). Consistent with this, the group of mRNAs with the lowest ribosome densities was highly enriched for the longest mRNAs, and the SD for ribosome density was much larger than for ribosome numbers (see above). These observations indicate that the range of associated ribosomes does not keep up with the range of ORF lengths, and ORF length is therefore expected to be a major factor determining ribosome density. There was indeed a strong inverse correlation between ORF length and ribosome density (Figure 3.9A). Short mRNAs were much more tightly packed with ribosomes than were long mRNAs. This inverse correlation was evident over the whole range of ORF sizes and ribosome densities. A similar inverse correlation was obtained when using mRNA lengths instead of ORF lengths based on 198 mRNAs for which untranslated regions (UTRs) are available from *S. pombe* GeneDB ($r = -0.9$; $P < 2e^{-16}$).

We were concerned that this inverse correlation might reflect a systematic artefact of the translational profiling approach. A bias could arise from underestimating the numbers of ribosomes in the more poorly resolved higher fractions where single-peak resolution for polysomes cannot be achieved (Figure 3.4A). We observed a similar negative correlation, however, when using only the relatively short mRNAs encoding ribosomal proteins (Figure 3.9B; Figure 3.7B); these mRNAs showed defined peaks in the well-resolved fractions of the polysome profiles where ribosome numbers can be determined with confidence (Figure 3.4A, fractions 6-10). To further exclude a possible error due to underestimating ribosomes, we associated double the originally estimated number of ribosomes associated with fraction 12 (~ 20 ribosomes instead of 10), which slightly increased the ribosome number for transcripts that strongly peaked in this very heavy polysome fraction. This re-analysis resulted in a similar negative correlation between ribosome density and ORF length (Figure 3.10), suggesting that underestimation of ribosome numbers for long and strongly translated mRNAs is not the cause of the negative correlation between ORF length and ribosome density. Moreover, there was a significant inverse relationship between the AugCAI and ORF

length ($r = -0.15$; $P < 2e^{-16}$), providing independent evidence for a link between ORF length and translational efficiency.

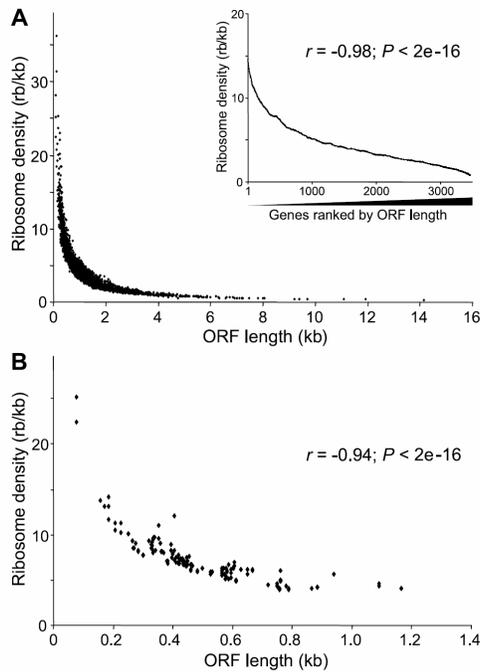


Figure 3.9 Inverse correlation between ribosome density and ORF length

(A) Scatterplot of ribosome density plotted against ORF length for the high-confidence set of protein-coding genes ($n = 3598$). The small inset graph shows moving averages (100-gene window) of ribosome density as a function of genes ranked by ORF length. The corresponding Spearman rank correlation between ribosome density and ORF length is also shown.

(B) Scatterplot of ribosome density plotted against ORF length as in (A) but showing only the mRNAs encoding ribosomal proteins ($n = 134$), along with the corresponding Spearman rank correlation.

We also observed a significant inverse correlation between ORF length and ribosome occupancy, although much less pronounced than for ribosome density ($r = -0.27$; $P < 2e^{-16}$). Together, these data raise the possibility that ORF length is a major factor for translational efficiency. We therefore expected that long proteins should be present in lower levels in the cell than short proteins due to differences in translational efficiency. To test this hypothesis, we took advantage of recent data on expression levels of nearly all *S. pombe* proteins (Matsuyama et al. 2006). For this study, all ORFs were cloned into the same vector, integrated into the same genomic site, and transcribed under the control of the same promoter. These data should therefore be minimally affected by differences in transcription or by differences in regulation of mRNA stability and translation via UTR sequences, as the lengths and sequences of

the ORFs are the only remaining factors that could influence translational efficiency, which (along with protein turnover) will determine protein levels. Protein levels in this study were then determined using a protein-based array technique, where tagged proteins were detected with an anti-His-tag antibody, and as internal reference α -tubulin was simultaneously quantified with an anti- α -tubulin antibody. The ribosome densities showed a significant positive correlation with protein levels from this study, while ORF length negatively correlated with protein levels on a global scale as predicted from our translational profiling data (Figure 3.11). For proteins present at lower levels, the correlations with ribosome density and ORF length were less evident, possibly due to increased noise. The protein levels also showed a significant correlation with ribosome occupancy ($r = 0.31$; $P < 2e^{-16}$) and a weak correlation with ribosome numbers ($r = 0.17$; $P < 2e^{-16}$). This is consistent with the intuition that ribosome density and ribosome occupancy are better measures for translational efficiency than the number of ribosomes associated with mRNAs. We conclude that ORF length is an important factor for translational efficiency and protein levels.

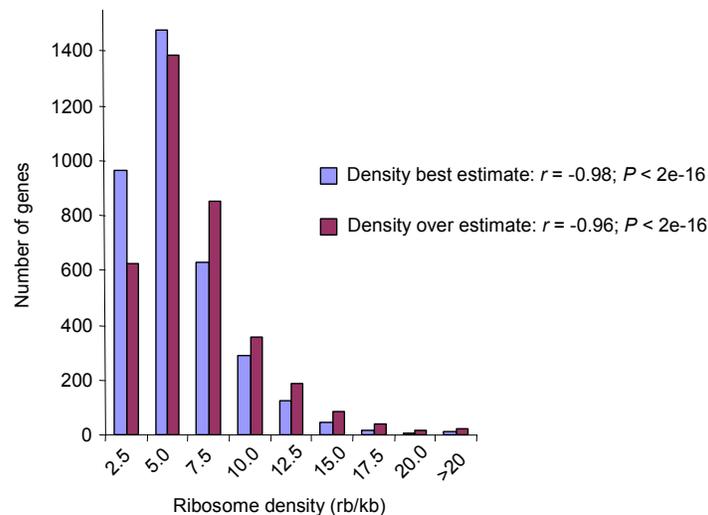


Figure 3.10 Overestimation of ribosome number for fraction 12 does not affect negative correlation between ribosome density and ORF length

Histogram using bins of different ribosome densities (upper bin limits given on X axis). Blue: distribution of ribosome densities calculated as described in Chapter 2. Purple: distribution of ribosome densities calculated the same way, except that the number of ribosomes associated with mRNAs in fraction 12 were 2 fold overestimated. The corresponding Spearman rank correlations between ORF length and ribosome densities calculated in both ways are also shown.

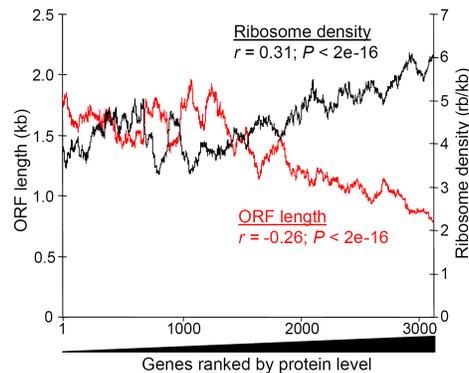


Figure 3.11 Correlations between ORF length/ribosome density and protein level
 Graph showing moving averages (100-gene window) of ribosome density (black) and ORF length (red) as a function of genes ranked by protein level ($n = 3265$). The corresponding Spearman rank correlations between protein level and ribosome density ($n = 3265$) and between protein level and ORF length ($n = 4434$) are also shown.

Genome-wide measurement of poly(A) tail length

The lengths of the poly(A) tails of mRNAs are thought to determine the efficiency of translational initiation based on single-gene studies (Preiss and Hentze 1998; Sachs 2000; Wickens et al. 2000; Stevenson and Norbury 2006). We therefore wondered whether translational efficiency might be reflected in poly(A) tail lengths on a genome-wide scale. To obtain global data on polyadenylation we used a technique called polyadenylation state array (PASTA) analysis: mRNAs are fractionated using a poly-U sepharose column and eluted at five different temperatures. The resulting five fractions correspond to mRNAs with different ranges of poly(A) tail length and were then hybridized to microarrays using total eluate as a reference (Figure 3.12). The fractionation of the mRNAs according to their length was done in collaboration with Traude Beilharz and Thomas Preiss (Victor Chang Cardiac Research Institute, Sydney, Australia), who have established this technique in the budding yeast *Saccharomyces cerevisiae* (Beilharz and Preiss 2007). The five fractions contained distinct, but partially overlapping distributions of poly(A) tail lengths (Figure 3.13).

Sizes of the poly(A) tail length ranged from ~10 to 80 nucleotides (nt) with the following peak sizes for each fraction: 12°C, ~10 nt; 25°C, ~22 nt; 30°C, ~30 nt; 35°C, ~40 nt; and 45°C, ~57 nt. These polyadenylation data revealed a continuous distribution of poly(A) tail lengths, both for specific mRNAs and between different mRNAs. There was minor cross-contamination of long-tailed mRNAs in the first 2 elution fractions. These transcripts may have bound non-specifically to the matrix or through poly(A) runs within the body of the transcript. Furthermore, short A-tract fragments were inefficiently precipitated in the bulk end-labelling experiment and as a consequence there is only weak signal on the gel from the first fraction (Figure 3.13). Nevertheless, the poly(A) profiles for different mRNAs were enriched for distinct sizes.

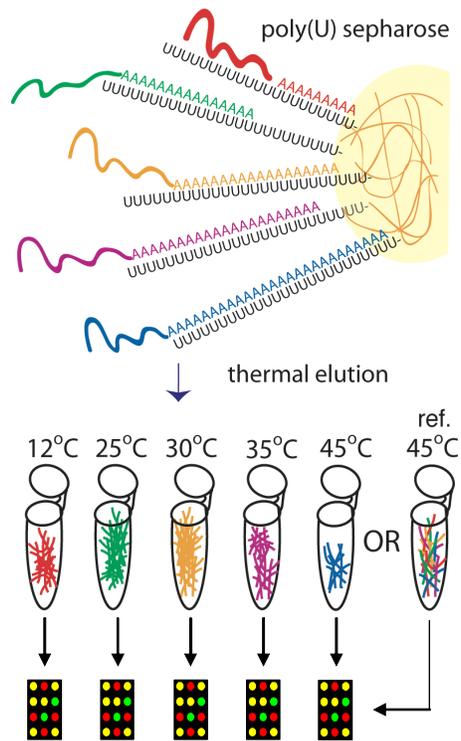


Figure 3.12 Experimental layout for polyadenylation state array (PASTA) mRNAs are fractionated using a poly(U) sepharose column and eluted at different temperatures. The resulting fractions contain mRNA with increasing poly(A) tail length, which are then hybridized onto a DNA microarray against total eluate as reference. Figure adapted from Beilharz et al. (2007).

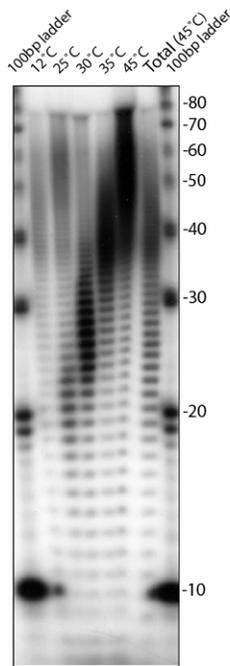


Figure 3.13 mRNAs fractionated using poly(U)-sepharose chromatography

Gel of poly(A) tail length tracts for mRNAs eluted from a poly(U) sepharose column at increasing temperatures as indicated on top, showing mRNAs with increasing poly(A) tail length from *S. pombe*. Aliquots of each mRNA fraction were end labeled with [³²P]-pCp, digested with RNases A and T1n, and poly(A) fragments analyzed by denaturing PAGE. Nucleotide numbers corresponding to the 100 bp ladder are indicated at right.

A modified RT-PCR assay, termed ligation-mediated poly(A) test (LM-PAT) (Salles and Strickland 1995; Beilharz and Preiss 2007) was used to verify the poly(A) profiles derived from the PASTA analysis (Figure 3.14). In this assay, PCR products between a gene-specific primer at the end of the ORF or within the 3' UTR and an anchored primer at the end of the poly(A) tail are generated. These PCR products reflect the poly(A) tail length of a specific mRNA. It should be mentioned that the use of oligo(dT)₁₂₋₁₈ mix and (dT)₁₂-anchor primes introduces a certain laddering of the products. Especially the shortest band represents all short tails (~7-22 nt) that can be bound by a (dT)₁₂-anchor but that cannot accommodate additional oligo(dT)₁₂₋₁₈ primers.

The distribution of poly(A) tails for mRNAs representing different tail lengths showed good agreement between the PASTA analysis and the LM-PAT-assay (Figure 3.15): *for3* mRNA, encoding formin, was enriched for short poly(A) tails in the PASTA analysis (Figure 3.15A), and consequently PCR products from the LM-PAT assay

could only be detected in the first fractions (Figure 3.15B). mRNAs encoding the ribosomal proteins *rps27* and *rpl14* showed peaks in the fractions enriched for poly(A) tails of medium length in both assays (Figure 3.15A,B). *Hhf1* and *hhf2*, two almost identical mRNAs encoding histones, peaked in the fractions enriched for long poly(A) tails (Figure 3.15A,B). Note that histone mRNAs in yeast are polyadenylated in contrast to most higher eukaryotes (Fahrner et al. 1980; Butler et al. 1990). Moreover, mitochondrially encoded mRNAs, which lack long poly(A) tails in fission yeast (Schäfer et al. 2005), showed the expected peak in the first fraction in the PASTA analysis (Figure 3.16).

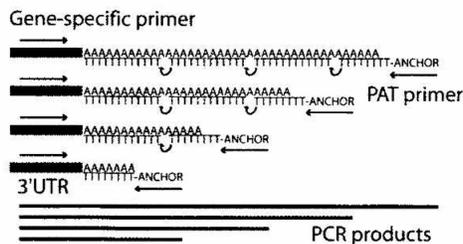


Figure 3.14 Experimental layout of LM-PAT assay

mRNAs are incubated with oligo(dT)₁₂₋₁₈ primers in the presence of T4-DNA ligase at 42°C followed by ligation of an oligo dT₁₂-anchor primer, thus covering the full length of poly(A) tails of mRNAs. cDNA is then synthesised from the ligated primers. Aliquots of this cDNA are used in PCR reactions to amplify a region between a site at the 3' end of the ORF of the mRNA under study and the anchor region. This figure is taken from Beilharz et al. (2007).

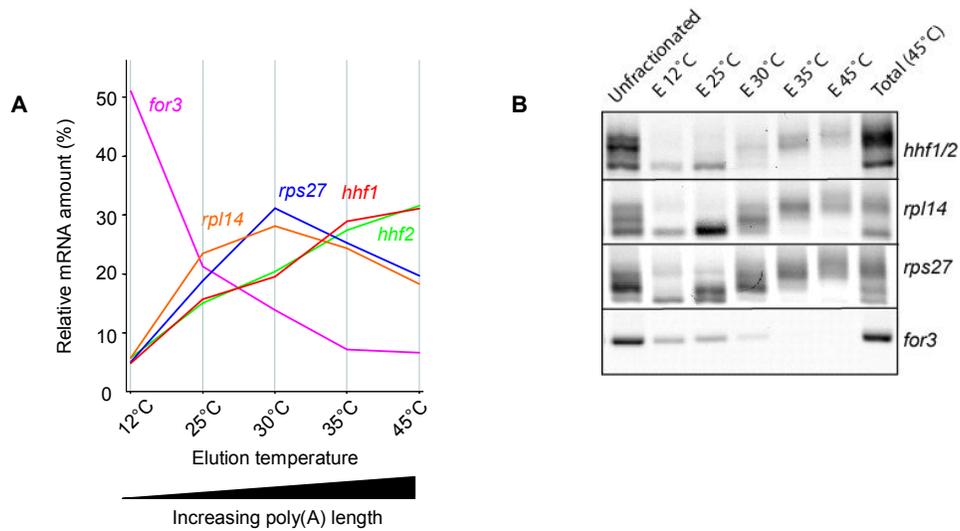


Figure 3.15 Examples of poly(A) tail length determination by LM-PAT assays and PASTA analysis

(A) Profiles of poly(A) tail length distribution for 5 mRNAs as determined by microarray-based PASTA analysis: *hhf1* and *hhf2* (almost identical mRNAs encoding histones, enriched for long poly(A) tails), *rpl14* and *rps27* (mRNAs encoding ribosomal proteins, enriched for poly(A) tails of medium length), and *for3* (mRNA encoding formin, enriched for short poly(A) tails). The curves show the relative amounts of RNA for a given mRNA in each of the five fractions as an average for two independent biological repeats. Different mRNAs are colour-coded.

(B) Distribution of poly(A) tail length determined using LM-PAT assays for the same mRNAs as in **(A)**. Fractions eluted from poly(U) sepharose column at increasing temperatures (corresponding to increasing tail lengths) as indicated on top.

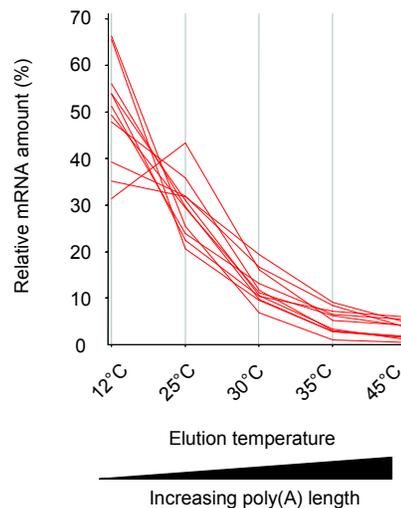


Figure 3.16 Poly(A) tail profiles for mitochondrially encoded mRNAs determined by PASTA analysis

Profiles of poly(A) tail length distribution for 11 mitochondrially encoded mRNAs (*SPMIT.01*, *SPMIT.02*, *SPMIT.03*, *SPMIT.04*, *SPMIT.05*, *SPMIT.06*, *SPMIT.07*, *SPMIT.08*, *SPMIT.09*, *SPMIT.10*, *SPMIT.11*) as determined by microarray-based PASTA analysis. The curves show the relative amounts of RNA for a given mRNA in each of the five fractions as an average of two independent biological repeats.

mRNAs with long poly(A) tails are more efficiently translated

For the further analyses, only mRNAs for which data could be obtained in all five fractions of two repeated experiments in the PASTA analysis were included. In total, 2795 protein-coding mRNAs fulfilled these criteria of which 2575 were also included in the translational profiling data set. The 20% of mRNAs with the longest tails were most enriched for transcripts repressed during environmental stress ($P \sim 1e^{-15}$; Chen et al. 2003) and for GO terms such as ‘biosynthesis’, ‘cytoplasm’, and ‘ribosome’ ($P \sim 2e^{-16} - 2e^{-27}$). The 20% of mRNAs with the shortest tails were most enriched for genes containing predicted nuclear localisation signals ($P \sim 1e^{-18}$) and for GO terms such as ‘nuclear lumen’, ‘nucleolus’, ‘RNA metabolism’, and ‘ribosome biogenesis and assembly’ ($P \sim 3e^{-8} - 3e^{-13}$).

The mRNAs were ranked by relative poly(A) tail length using a weighted average of the relative amounts of mRNA associated with each fraction (Table S1; Chapter 2). Integration of these data with the translational data revealed that poly(A) tail lengths significantly increased with increasing ribosome density (Figure 3.17). Moreover, poly(A) tail lengths increased with decreasing ORF lengths (Figure 3.17), consistent with the strong inverse correlation between ORF length and ribosome density (Figure 3.9). These data were corroborated by genome-wide binding data for the poly(A) binding protein Pab1p using RIP-chip: Pab1p tended to be most enriched in precipitated mRNAs with long poly(A) tails according to the PASTA analysis, and ORF lengths showed a strong inverse correlation with Pab1p enrichment (Juan Mata, unpublished data; see also Beilharz and Preiss 2007). Poly(A) tail lengths also correlated with ribosome occupancy ($r = 0.27$; $P < 2e^{-16}$) and with protein levels ($r = 0.21$; $P < 2e^{-16}$). Together, these data reveal a genome-wide connection between ORF length, poly(A) tail length, and translational efficiency: short mRNAs tend to have long poly(A) tails and are more efficiently translated than longer mRNAs that tend to have shorter poly(A) tails. These connections are ultimately reflected at the protein levels and are most evident for the highly expressed proteins (Figure 3.11).

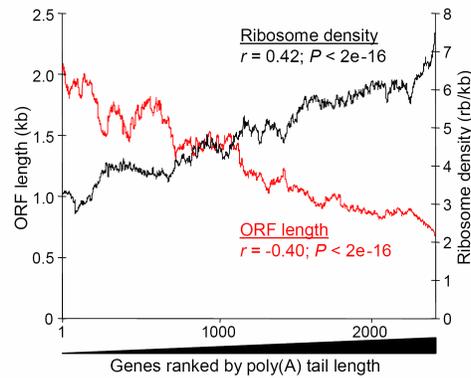


Figure 3.17 Correlations between ORF length and ribosome density and poly(A) tail length Moving averages (100-gene window) of ribosome density (black) and ORF length (red) as a function of 2576 genes ranked by poly(A) tail length. The Spearman rank correlations between poly(A) tail length and ribosome density ($n = 2576$) and between poly(A) tail length and ORF length ($n = 2714$) are also shown.

Abundant mRNAs are more efficiently translated

Steady-state mRNA levels are another important determinant of gene expression. mRNA levels in exponentially growing cells were estimated from the absolute hybridization signal intensities using Affymetrix chips (Table S1). These signal intensities should reflect mRNA abundance quite accurately, given that there are several different probes for each mRNA on the array, which should minimize the influence of differential efficiencies of hybridization to different probes. Furthermore, these data were in good agreement with independent data for mRNA levels obtained by hybridizing mRNA against a genomic DNA reference on our in-house DNA microarrays (data not shown). The 10% most abundant mRNAs were most enriched for transcripts repressed during environmental stress ($P \sim 2e^{-86}$; Chen et al. 2003) and for GO terms such as ‘ribosome’, ‘protein biosynthesis’, ‘cellular metabolism’, and related terms ($P \sim 2e^{-55}$ to $1e^{-128}$). The 10% least abundant mRNAs were most enriched for transcripts induced during meiosis and stress ($P \sim 3e^{-15}$ to $7e^{-19}$; Mata et al. 2002; Chen et al. 2003), for *S. pombe* specific transcripts ($P \sim 1e^{-34}$; Mata and Bähler 2003), and for GO terms such as ‘meiosis’, and ‘M-phase’ ($P \sim 1e^{-18}$ to $1e10^{-25}$). Note, that mRNA levels did not correlate with ORF lengths ($r = -0.02$; $P = 0.11$) and there was also no correlation when transcript length was used instead of ORF

length ($r = -0.06$; $P = 0.41$; calculated using the 198 mRNAs for which 5' and 3' UTR length data are available in *S. pombe* GeneDB (www.genedb.org/genedb/pombe/index.jsp) (Figure 3.18). mRNA levels significantly correlated, however, with poly(A) tail lengths, with the most abundant mRNAs showing a tendency for longer tails (Figure 3.19A).

We then checked for relationships between mRNA levels and translational efficiency. The mRNAs with the lowest expression levels tended to be associated with fewer ribosomes than the mRNAs with the highest levels (Figure 3.19B). This raised the possibility that mRNA abundance is somehow coordinated with translational efficiency. Consistent with this, ribosome densities showed some correlation with mRNA levels ($r = 0.14$; $P < 2e^{-16}$). Stronger correlations throughout the entire population of mRNAs were apparent between ribosome occupancy and mRNA levels (Figure 3.19C). There was also a significant correlation between the AugCAI and mRNA levels ($r = 0.22$; $P < 2e^{-16}$). Taken together, these findings indicate a genome-wide coordination between mRNA levels and translational efficiency: more abundant mRNAs tend to be more efficiently translated as reflected by their higher ribosome occupancy and, to a lesser extent, higher ribosome density.

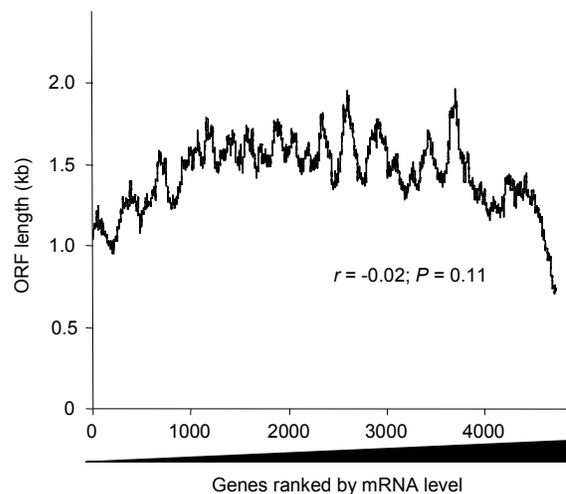


Figure 3.18 No correlation between mRNA levels and ORF length

Graph showing moving averages (100-gene window) of ORF length as a function of genes ranked by mRNA level ($n = 4818$). The corresponding Spearman rank correlation between ORF length and mRNA level is shown within the graph.

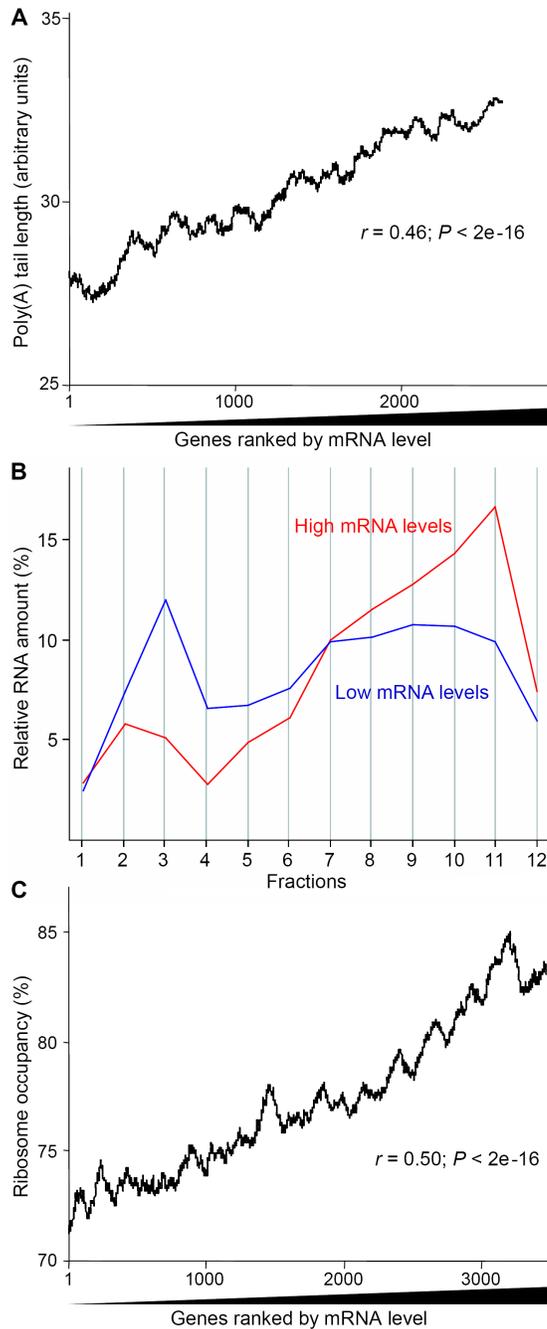


Figure 3.19 Correlations between mRNA level and poly(A) tail length and ribosome occupancy

(A) Graph showing moving averages (100-gene window) of poly(A) tail length as a function of genes ranked by mRNA level ($n = 2688$), along with the corresponding Spearman rank correlation.

(B) The curves show the average translation profiles of the mRNAs with the 500 highest (red) or 500 lowest (blue) levels. Curves are plotted as in Figure 3.4B.

(C) Graph showing moving averages (100-gene window) of ribosome occupancy as a function of genes ranked by mRNA level ($n = 3567$), along with the corresponding Spearman rank correlation.

Stable and highly transcribed mRNAs are more efficiently translated

The steady-state level of a given mRNA is determined by the rate of transcription and the rate of decay, both of which are controlled at genome-wide level (Mata et al. 2005). The correlation between translational efficiency and mRNA abundance could therefore reflect a connection between translation and mRNA stability and/or between translation and transcription.

Abundant mRNAs are expected to be more stable on average than less abundant mRNAs. To test whether mRNA stability is linked to translation, global mRNA half-lives were estimated by blocking transcription and measuring mRNA levels at different times after transcriptional shut-off. These experiments were performed by Samuel Marguerat. Cells were treated with the transcriptional inhibitor 1,10-phenanthroline, and mRNA was isolated before and at 4, 12, and 28 min after transcriptional shut-off and mRNA levels were measured using microarrays. Two lists of genes with short and long half-lives were generated from these data (see Chapter 2) (Figure 3.20; Table S1). These experiments provided reliable estimates on relative half-lives for the 868 least stable mRNAs, with half-lives ranging from ~10 to 96 min and a median of ~33 min. This group of unstable mRNAs is significantly enriched for genes with periodic expression during the cell cycle ($P \sim 6e^{-15}$; Rustici et al. 2004; Marguerat et al. 2006); these mRNAs peak in levels during a short phase of the cell cycle and are therefore expected to have short half-lives. The unstable mRNAs were also enriched for genes associated with the GO terms ‘regulation of biological process’, ‘cell communication’, ‘signal transduction’, and ‘cell septum’ ($P \sim 1e^{-5} - 2e^{-8}$). This probably reflects that mRNAs encoding regulatory proteins or proteins only required during a defined stage such as septation need to be tightly controlled. We also selected a similarly sized group of bona fide stable mRNAs whose expression levels were not altered 30 min after transcriptional shut-off (Figure 3.20). This group was most enriched for genes with the GO terms ‘cytoplasm’ and ‘mitochondrial part’ ($P \sim 5e^{-5} - 9e^{-8}$). As expected, mRNAs with short half-lives were significantly less abundant on average than mRNAs with longer half-lives (Figure 3.21A).

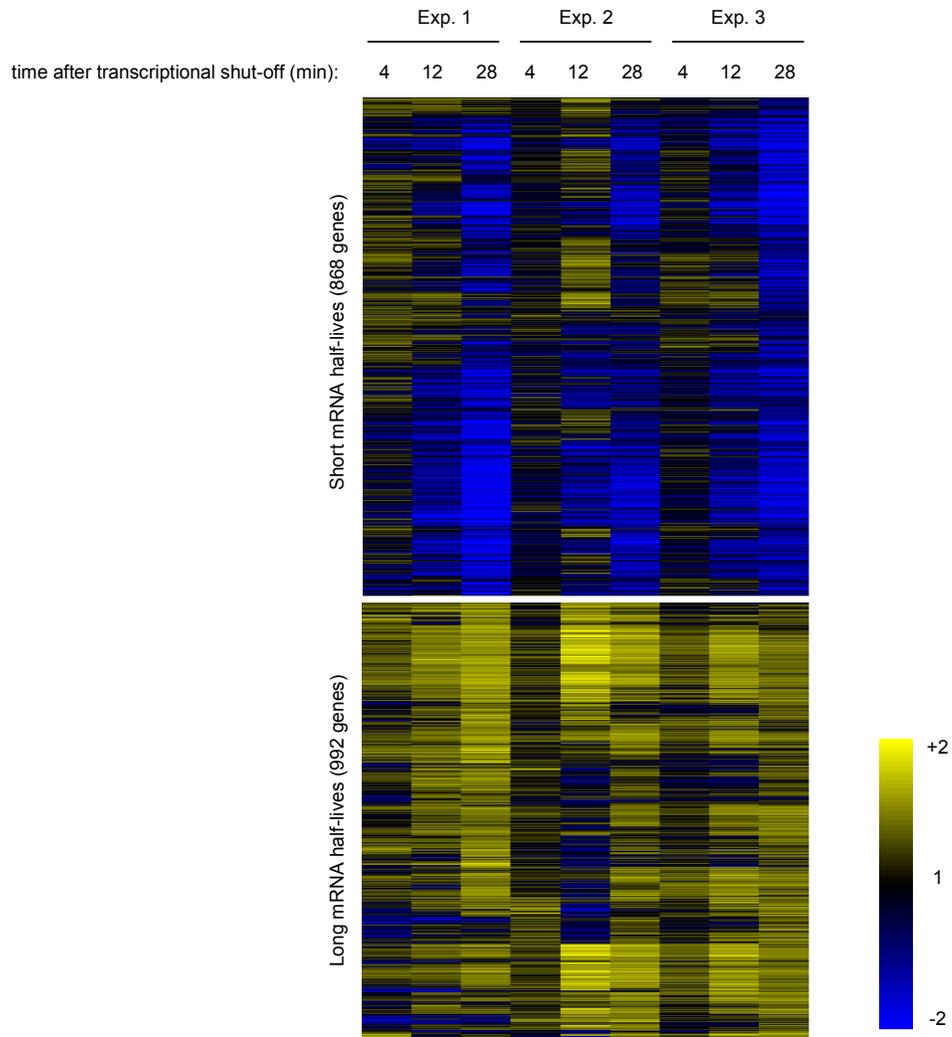


Figure 3.20 Determination of mRNAs with short and long half-lives

To estimate mRNA half-lives, cells were treated with the transcriptional inhibitor 1,10-phenanthroline, and mRNA was isolated before and at 4, 12, and 28 min after transcriptional shut-off. Two lists of genes with short and long half-lives were created from these data. The figure shows heat maps of these two gene lists, clustered using the Spearman correlation. Data from three independent biological experiments are shown. The columns represent experimental time points, and rows represent genes. The data of each array were normalized to the 50th percentile of the measurements taken from that array and colour-coded according to the ratios between experimental samples vs sample before transcriptional shut-off.

We then checked for relationships between mRNA stability and translational efficiency. The mRNAs with long half-lives showed significantly higher ribosome occupancies and densities on average than mRNAs with short half-lives (Figure 3.21B,C). Thus, stable mRNAs seem to be more efficiently translated than less stable mRNAs. Although translational efficiency correlated with both poly(A) tail length and mRNA stability, we did not detect any correlation between mRNA stability and poly(A) tail length (Figure 3.21D).

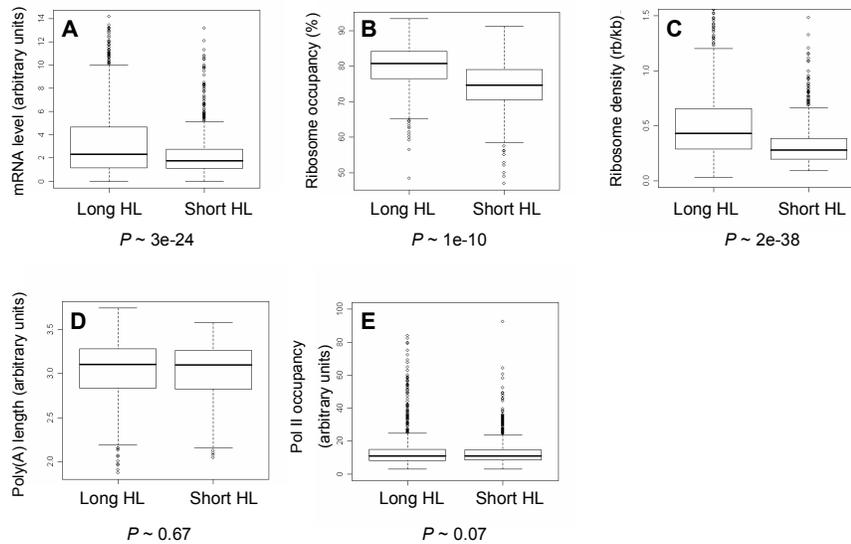


Figure 3.21 Correlations between mRNA half-lives and other gene expression properties

Boxplot showing mRNA levels (**A**), ribosome occupancies (**B**), ribosome densities (**C**), poly(A) tail lengths (**D**), and RNA polymerase II occupancies (**E**) for two groups of mRNAs with either long (long HL) or short (short HL) half-lives.

The box contains the middle 50% of the data. The upper edge (hinge) of the box indicates the 75th percentile of the data set, and the lower hinge indicates the 25th percentile. The range of the middle two quartiles represents the inter-quartile range. The thick black line indicates the median of the data. The ends of the vertical lines or "whiskers" indicate the minimum and maximum data values, unless outliers are present in which case the whiskers extend to a maximum of 1.5 times the inter-quartile range. The points outside the ends of the whiskers are outliers or suspected outliers.

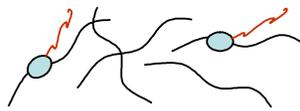
The significance of the difference between the means of the two mRNA groups is given for each panel.

The relationship between mRNA stability and translational efficiency is consistent with data on single genes that indicate a connection between mRNA stability and translation (Sachs 2000; Wickens et al. 2000). This raises the possibility that mRNA stability may be the main determinant for the observed correlation between mRNA levels and translation (Figure 3.19). To test this, we needed to also estimate transcriptional efficiency, the other determinant of mRNA levels. The relative amount of RNA polymerase II associated with a given ORF provides an estimate for transcriptional efficiency (Sandoval et al. 2004). We therefore established a systematic approach to measure RNA polymerase II (Pol II) occupancy for all fission yeast ORFs using a chromatin immunoprecipitation followed by analysis on microarrays (Figure 3.22, Table S1; Material and Methods). These experiments were performed by Samuel Marguerat. Functional analysis of the 10% of genes that were either most or least associated with RNA polymerase II showed highly similar

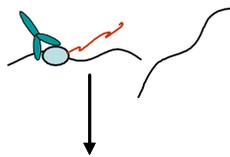
enrichments for GO terms and functional groups as the 10% most or least abundant mRNAs, respectively. The mitochondrially encoded genes were a notable exception; they showed high mRNA levels but were strongly under-enriched in the polymerase II precipitations, consistent with these genes being transcribed by a different RNA polymerase (Schäfer et al. 2005). Transcriptional efficiency did not show any significant correlation with mRNA stability (Figure 3.21E). As expected, however, it was correlated with mRNA levels (Figure 3.23A).

polymerase II occupancy

Extraction of cross-linked chromatin and sonication



Immunoprecipitation of DNA fragments bound to RNA polymerase II



DNA labelling and hybridisation

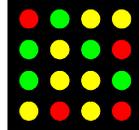


Figure 3.22 Experimental layout for estimating Pol II occupancy on a genome-wide scale

DNA and protein-complexes are crosslinked with formaldehyde, chromatin is extracted from the cells and sonicated. An immunoprecipitation using an antibody recognizing Pol II is performed. Immunoprecipitated DNA is labelled and probed on a microarray against input DNA. Genes that are more efficiently transcribed will have more Pol II bound and will be more strongly enriched in the immunoprecipitates.

We next checked for relationships between transcriptional and translational efficiencies. Pol II occupancy showed a correlation with ribosome occupancy (Figure 3.23B) and a marginal, albeit significant, correlation with ribosome density ($r = 0.11$; $P \sim 3e^{-11}$). Thus, both transcription and mRNA turnover are reflected at the level of translation: efficiently transcribed and stable mRNAs tend to be more efficiently translated.

Surprisingly, transcriptional efficiency was also correlated with poly(A) tail lengths (Figure 3.23C). This is in contrast to the apparent absence of any connection between mRNA stability and poly(A) tails (Figure 3.21D), but it is consistent with the correlation between mRNA levels and poly(A) tails (Figure 3.19A).

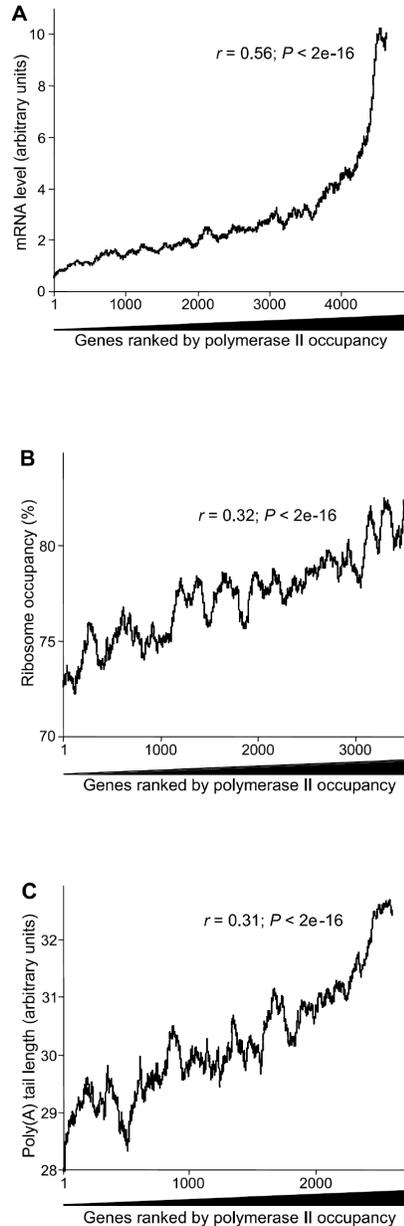


Figure 3.23 Correlations between Pol II occupancy and other gene expression properties

(A) Moving averages (100-gene window) of relative mRNA level as a function of 4724 genes ranked by Pol II occupancy, along with corresponding Spearman rank correlation.

(B) Moving averages (100-gene window) of ribosome occupancy as a function of 3598 genes ranked by Pol II occupancy, along with corresponding Spearman rank correlation.

(C) Moving averages (100-gene window) of poly(A) tail length as a function of 2713 genes ranked by Pol II occupancy, along with corresponding Spearman rank correlation.

Changes in mRNA polyadenylation in response to transcriptional switch-on

Given the correlation between poly(A) tail length and transcription (Figure 3.23C) and the finding that the carboxy-terminal domain (CTD) of Pol II is important for 3' end processing of the mRNA (Proudfoot and O'Sullivan 2002; Proudfoot et al. 2002), it was tempting to hypothesize that poly(A) tail lengths are actually determined by transcription rates. To test this hypothesis, we used LM-PAT assays to analyse polyadenylation for specific mRNAs that were transcribed at different rates using regulatable promoters. At first, *pom1* (Bähler and Nurse 2001) and *rpb4* (Sharma et al. 2006) were used as reporter genes. Pom1p is a cell cycle regulated kinase and essential for cell symmetry, Rpb4p is a subunit of RNA polymerase II. Both genes were regulated using *nmt1* promoters of different strength integrated into the genomic locus of the two genes. Expression from the *nmt1* promoter is induced by removal of thiamine from the medium. *3nmt1* is the strongest promoter, *41nmt1* is the intermediated promoter, and *81nmt1* is the weakest promoter (Basi et al. 1993). mRNAs before induction as well as 16 and 21 hours after induction were analyzed for poly(A) tail length by LM-PAT assays (Figure 3.24). Whereas mRNA levels increased for both mRNAs as expected, poly(A) tail length was not affected by the transcriptional up-regulation.

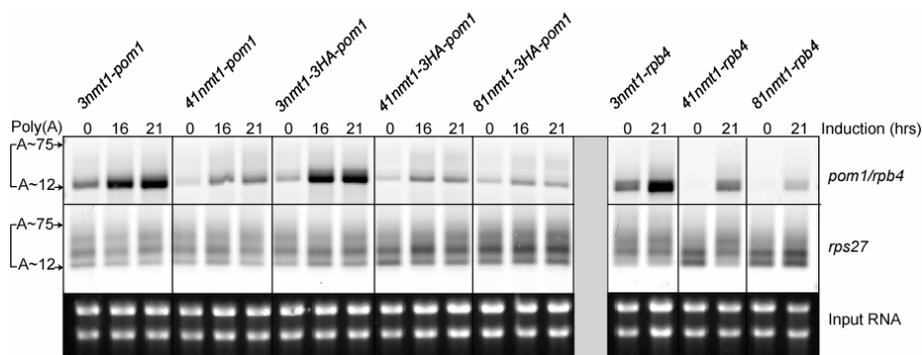


Figure 3.24 No changes in poly(A) tail length for mRNAs induced in expression using *nmt1* promoters with long induction time

The *pom1* (left panels) and *rpb4* (right panels) genes were transcriptionally induced by thiamine removal using regulatable *nmt1* promoters of different strength: *3nmt1*, strongest promoter; *41nmt1*, intermediate promoter; and *81nmt1*, weakest promoter. mRNAs before induction (0) as well as 16 and 21 hours after induction were analyzed for poly(A) tail length by LM-PAT assays. Both mRNAs showed short poly(A) tails independently of transcription rates. The longer-tailed *rps27* mRNA is included as a control (middle panels), and the input RNA is shown below.

These data indicate that the transcription rate does not influence poly(A) tail length for the tested mRNAs. A problem with this experiment and its interpretation might arise through the long induction time needed for expression from the *nmt1* promoter. Therefore we wanted to test whether the same result would be obtained looking at mRNAs induced for a short period only. To this end we made use of the promoter of *urg1* (uracil regulatable gene), which shows a fast induction time after addition of uracil to the medium (S. Watt, J. Mata, G. Burns and J. Bähler, manuscript in preparation). *pom1* under the regulation of this promoter, *urg1* under its own promoter and SPAC1002.17, which shows a similar short induction time under its own promoter after uracil addition, were tested by LM-PAT assay (Figure 3.25). A 30 min time-course of induction was followed by a 30 min time-course of repression 4 hours later. A weak band indicating some long-tailed form of *pom1* was present 5 min after induction, but from 10 min onwards the short-tailed form predominated. *Urg1* and SPAC1002.17 were present mainly in a long-tailed form 5 min after induction, and short-tailed forms appeared later after 15 and 30 min. However, these mRNAs seemed to have slower deadenylation kinetics, as long-tailed forms were still evident 30 min after induction, but were gone at the time of repression. Thus, when transcription was induced within a short time, a transient population of longer tailed mRNAs was apparent, which were then deadenylated with different kinetics depending on the particular mRNA.

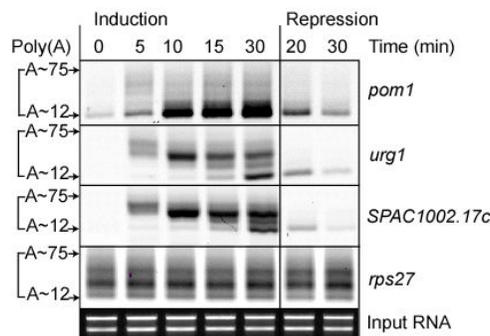


Figure 3.25 Transient changes in poly(A) tail length for mRNAs induced in expression using a promoter with short induction time

Transcription of *pom1* was induced under the control of the regulatable *urg1* promoter showing fast induction time. A 30-min timecourse of induction was followed by a 30-min timecourse of repression 4 hours later, and mRNAs were analyzed for poly(A) tail length by LM-PAT assays. Some long-tailed form of *pom1* mRNA is present at 5 min after induction, but from 10 min onwards the short-tailed form predominates. Corresponding LM-PAT assays from the same cells are also shown for *urg1* (under its own promoter) and SPAC1002.17c (which shows a similar short induction time under its own promoter). These mRNAs have slower deadenylation kinetics, and long-tailed forms are still evident 30 min after induction, but are gone at the time of repression. The unregulated *rps27* mRNA and input RNA are shown as controls.

We conclude that the transcription rate does not directly influence poly(A) tail lengths, although increased transcription can lead to transiently increased tail lengths before reaching steady-state conditions. Similar observations were made in budding yeast for mRNAs with predominantly long poly(A) tails measured by LM-PAT assay after replenishing stationary phase cells with fresh media (Beilharz and Preiss 2007).

Conclusion

Comparisons between our genome-wide data sets on key aspects of gene expression control, ranging from transcription to translation, highlight a remarkable degree of global interconnectivity between different layers of gene expression. For a summary of all correlations between the diverse data sets see Table 3.1.

Table 3.1 Summary of all correlations between the different genome-wide data sets on key aspects of gene expression

	Ribosome density	Ribosome occupancy	ORF length	Poly(A) tail	mRNA level	Pol II occupancy	Protein levels
Ribosome occupancy	0.40 ($< 2e-16$)						
ORF length	-0.98 ($< 2e-16$)	-0.27 ($< 2e-16$)					
Poly(A) tail	0.42 ($< 2e-16$)	0.27 ($< 2e-16$)	-0.40 ($< 2e-16$)				
mRNA level	0.14 ($< 2e-16$)	0.50 ($< 2e-16$)	-0.02 (0.11)	0.46 ($< 2e-16$)			
Pol II occupancy	0.11 ($8e-11$)	0.32 ($< 2e-16$)	-0.05 (0.11)	0.31 ($< 2e-16$)	0.56 ($< 2e-16$)		
Protein levels	0.31 ($< 2e-16$)	0.31 ($< 2e-16$)	-0.26 ($< 2e-16$)	0.21 ($< 2e-16$)	0.23 ($< 2e-16$)	0.10 ($2e-10$)	
AugCAI	0.20 ($< 2e-16$)	0.25 ($< 2e-16$)	-0.15 ($< 2e-16$)	0.15 ($2e-14$)	0.22 ($< 2e-16$)	0.15 ($< 2e-16$)	0.17 ($< 2e-16$)

This table shows the Spearman rank correlations between all data-sets, for which ranked lists were available. The corresponding P values are shown in brackets. Positive correlations with a P value smaller than $2e-16$ are coloured green; negative correlations with a P value smaller than $2e-16$ are coloured red.

The large network of correlations between all aspects of regulation suggests widespread coordination between multiple gene expression levels for coherent and efficient protein production. Some of these relationships may reflect direct mechanistic links (e.g., translational efficiency could influence mRNA stability), while others may reflect independent evolutionary selection at different levels of regulation (e.g., alignment of transcriptional and translational efficiencies). These rich data sets, all acquired under one standardized condition in a simple model organism,

are a basis to interpret global and specific regulation of gene expression in response to environmental or genetic perturbations. The findings presented here also provide a framework to gain comprehensive mechanistic insight into multi-layered gene expression programs and should advance a system-wide understanding of gene expression also in more complex organisms.