PHILIPPINE INFORMATION TECHNOLOGY JOURNAL. VOLUME 1, NUMBER 1, FEBRUARY 2008

# Gene expression levels in *Saccharomyces cerevisiae* strongly reflect energetic cost rates of gene products

Ester Villaprinyo
Dept. de Ciències Mèdiques Bàsiques
Universitat de Lleida
Lleida, Spain
+34973702406
evilaprinyo@cmb.udl.cat

Rui Alves
Dept. de Ciències Mèdiques Bàsiques
Universitat de Lleida
Lleida, Spain
+34973702406
ralves@cmb.udl.cat

Albert Sorribas
Dept. de Ciències Mèdiques Bàsiques
Universitat de Lleida
Lleida, Spain
+34973702406
albert.sorribas@cmb.udl.cat

Armindo Salvador
Ctr. for Neuroscience and Cell Biology
and Chemistry Department
University of Coimbra
3004-535 Coimbra, Portugal
+351-91-9619593
salvador@cnc.uc.pt

## ABSTRACT

We examine how protein abundances under log phase (LP) growth and mRNA change-folds in the shift from LP to late stationary phase (SP) relate to the energetic cost rates (*E*) of maintaining each polypeptide chain's steady state abundance. In LP polypeptide chain abundances decrease very significantly with estimated per-molecule energetic cost (*e*). Turnover numbers, primary sequence length, and mean aminoacid biosynthetic cost are significantly lower for abundant vs. rare proteins, but only the former two factors contribute substantially for variance in *e*. In the LP→SP shift, mRNA change-folds tend to decrease with increasing polypeptide chain cost rates, abundance and length, but increase with turnover numbers. However, the latter trend is reversed at the proteome level if the lack of growth-related protein turnover in SP is accounted for. Altogether, the results suggest that although adaptive plasticity of *e* dampens the selective pressure on gene expression levels for minimizing *E*, expression levels are very sensitive to this selective pressure.

## Categories and Subject Descriptors

Not applicable

## General Terms

None applicable

## Keywords

Gene expression, energetic costs, yeast, stationary phase.

## 1. INTRODUCTION


*ICMSB'08*, February 25–28, 2008, Diliman, Philippines.


Energy expenditures in yeast protein synthesis have substantial fitness costs [1] and span a range of ~$10^5$-fold over pools of different proteins. Gene expression profiles for different conditions should thus reflect the heterogeneity not only of the functional advantages but also of the energetic costs of sustaining expression levels of the gene products. Furthermore, as fitness costs of energy expenditure are lower in richer environments, gene expression changes upon transition to poorer environments should also reflect the heterogeneity of a gene product's energetic costs. Two conditions where an extreme difference in resource availability exists are log phase growth in rich media (LP) and late stationary phase in depleted media (SP). In LP, cells have abundant resources available, rely mostly on glycolysis for energy production, gear gene expression towards maximum growth, and let growth-related dilution contribute substantially to the turnover of many proteins. In SP resources are too scarce for growth, cells rely mostly on respiration for energy production and gear gene expression towards maximum survival.

Here we examine (a) to what extent do polypeptide chain abundances in log phase and mRNA change-folds in the shift from LP to SP relate to the energetic costs of maintaining each polypeptide chain's steady state level; and (b) what basic strategies have evolved that minimize these costs.

## 2. METHODS

The ATP-equivalents expenditures in maintaining the level of each of 3009 polypeptide chains (~45% of the yeast's protein-coding genes) for which enough information is available were computed as follows. For the steady state level of a polypeptide chain to be maintained, growing cells have to synthesize it at a rate that is sufficient to compensate for the decrease in concentration caused by the combined processes of growth-related dilution and protein degradation.

The aminoacids sequestered in polypeptide chains synthesized to compensate for growth are incorporated into biomass, and therefore the growth-related cost rate ($\boldsymbol{E_g}$) of a

polypeptide chain must include the cost of aminoacid biosynthesis. Thus,

$$E_g = \sum_{i=1}^{20} (c_i + 4) f_i V P , \quad (1.1)$$

where $\boldsymbol{P}$ stands for the polypeptide chain's concentration [from ref. 2], $\boldsymbol{V}$ stands for the growth-related dilution rate ($7.7\times10^{-3}$/min for LP, 0 for SP), $\boldsymbol{f_i}$ stands for the frequency of aminoacid $i$ in the chain's primary sequence (from the Comprehensive Yeast Genome Database), $\boldsymbol{c_i}$ stands for the biosynthetic opportunity cost of aminoacid $i$ [from ref. 1], and 4 the polymerization cost per aminoacid [3]. Except where otherwise noted, we used the $c_i$ values for fermentative metabolism for LP.

Upon protein degradation, the constituent aminoacids of the peptide chains are recovered. Therefore, the degradation-related cost rate ($\boldsymbol{E_d}$) of maintaining the chain's concentration does not include the aminoacids' biosynthetic costs:

$$E_d = 4nDP \quad (1.2)$$

Here, $\boldsymbol{D}$ stands for the chain's degradation rate constant [from ref. 4] and $\boldsymbol{n}$ stands for the polypeptide chain's primary sequence length. The total cost rate of maintaining the steady state concentration of a polypeptide chain under LP is thus

$$E_{LP} = E_g + E_d \quad (1.3)$$

For SP, in turn, only $E_d$ contributes to the polypeptide maintenance costs, because cells do not grow under this condition. Per-molecule energetic cost rates are given by $e_g = E_g/P$. $e_d = E_d/P$ and $e = E_{LP}/P$.

Gene expression changes for the LP→SP shift were obtained from ref. [5].

We estimate the effect of changing growth-related dilution on the relationship between transcript and polypeptide chain $i$ change-folds based on the following simple model

$$\frac{dP_i}{dt} = TM_i - (D_i + V)P_i , \quad (1.4)$$

where $M$ stands for mRNA concentration and $T$ for translational activity. At steady state ($\frac{dP}{dt} = 0$) the following relationships apply for LP and SP (V=0) respectively, assuming that D remains unchanged

$$P_i(LP) = \frac{T_{LP} M_i(LP)}{D_i + V}, \quad P_i(SP) = \frac{T_{SP} M_i(SP)}{D_i} . \quad (1.5)$$

Dividing $P_i(SP)$ by $P_i(LP)$, one finds

$$p_i = (1 + \frac{V}{D_i}) \frac{T_{SP}}{T_{LP}} m_i , \quad (1.6)$$

where $\boldsymbol{p_i}$ and $\boldsymbol{m_i}$ stand for the polypeptide chain and mRNA fold changes, respectively. Note that the relationship between $p_i$ and $m_i$ becomes quite sensitive to $D_i$ for very stable proteins ($t_{½}$ > 90 min). Because we are interested in the proteome-scale trends for the overall set of $p_i$ values, rather than in the values of the $p_i$, for specific genes, it is unnecessary to estimate the ratio $T_{SP}/T_{LP}$.

**Table 1. Trends of cost components versus protein abundance (log phase)**

| Property [a] | Tercile medians ratio | M.-W. p [a] | Spearman rank correlation (· vs *P*) |
|---|---|---|---|
| $E_{LP}$ | 12. | $<10^{-266}$ | |
| $e$ | 0.56 | $<10^{-39}$ | -0.26 ($p<10^{-47}$) |
| $e_d$ | 0.41 | $<10^{-50}$ | -0.29 ($p<10^{-63}$) |
| $e_g$ | 0.84 | $<10^{-9}$ | -0.16 ($p<10^{-23}$) |
| $n$ | 0.87 | $<10^{-8}$ | -0.15 ($p<10^{-21}$) |
| $a$ | 0.98 | $<10^{-15}$ | -0.16 ($p<10^{-23}$) |
| $D+V$ | 0.63 | $<10^{-59}$ | -0.32 ($p<10^{-76}$) |

[a] Mann-Whitney U test top vs. bottom protein abundance terciles.

## 3. RESULTS

In LP, growth and degradation-related cost rates for polypeptide chains tend to be balanced: the 0.25, 0.5 and 0.75 quantiles of the ratio $E_g/E_d$ are 0.70, 1.3 and 2.4, respectively. Polypeptide chain abundances decrease substantially and very significantly with estimated per-molecule energetic cost, this trend being most pronounced for the costliest polypeptide chains: regression of log($P$) vs. log($e$) yields a scaling exponent of -0.44, and comparison of chain abundances for the top tercile of *e versus* the bottom tercile yields $p<10^{-40}$ in the Mann-Whitney U test.

The higher the abundance of a given protein is the stronger the selective pressure for minimizing per-molecule energetic costs should be. It is thus interesting to examine how the various adaptable factors that contribute to $e$ respond to this selective pressure. Turnover numbers ($D+V$), $n$, and mean aminoacid biosynthetic costs ($\boldsymbol{a}$) are significantly lower for abundant vs. rare polypeptide chains, but only the former two factors contribute substantially for variation in $e$ (Table 1). Median turnover numbers show a consistent decrease over all the range of protein abundances (data not shown). But in contrast, the median $n$ decreases mainly for $P$ in the range of 300-1000 molecules/cell and does not show any additional consistent decrease for higher abundances. Likewise, the median $a$ decreases mainly for $P$ in the range of 4700-14000 molecules/cell and does not show any additional consistent decrease for higher abundances.

In the LP→SP shift, transcripts coding for the polypeptide chains that would be costliest to maintain should their abundances and half-lives remain the same as in LP (*i. e.* polypeptide chains in the top tercile of $E_d$) show significantly lower change folds than those coding for less costly polypeptide chains (*i. e.* in the bottom tercile of $E_d$) (Table 2). These lower change folds correspond mainly to stronger down-regulation, as few transcripts are up-regulated and the median change-folds in each group are well below 1. Down-regulation of expression is also stronger for large $P$ or $n$ in a statistically significant manner, despite $n$ correlating negatively with $P$ in LP. In contrast, transcripts coding for polypeptide chains with high $e_d$, $a$ or $D$, are less down-regulated than those coding for polypeptide chain with lower values of these indices. (Values of $a$ in these calculations were based on the aminoacid biosynthetic costs for respiratory metabolism.) However, these latter trends are due in part to the negative correlation of $e_d$, $a$ and $D$ with $P$ in LP, combined with the

**Table 2. Comparison of mRNA abundance changes (*m*) between top and bottom terciles of each cost component (LP→SP shift)**

| Property | Median change-fold bottom tercile | Median c.-f. top tercile | M.-W. p |
|---|---|---|---|
| $E_d$ | 0.85 | 0.56 | $<10^{-29}$ |
| $e_d$ | 0.77 | 0.80 | $<10^{-2}$ |
| $n$ | 0.95 | 0.82 | $<10^{-5}$ |
| $a$ | 0.84 | 1.02 | $<10^{-13}$ |
| $D$ | 0.74 | 0.86 | $<10^{-7}$ |
| $P$ | 0.90 | 0.48 | $<10^{-76}$ |

**Table 3. Comparison of estimated polypeptide chain abundance changes (*p*) between top and bottom terciles of each cost component (LP→SP shift)**

| Property | Tercile medians ratio | Mann-Whitney p |
|---|---|---|
| $E_d$ | 0.52 | $<10^{-66}$ |
| $e_d$ | 0.67 | $<10^{-22}$ |
| $n$ | 0.76 | $<10^{-11}$ |
| $a$ | 1.02 | 0.15 |
| $D$ | 0.73 | $<10^{-16}$ |
| $P$ | 0.66 | $<10^{-26}$ |

preferential down-regulation of transcripts for polypeptide chains that are abundant in LP.

Because about half of the polypeptide chains in the considered sample have values of $D$ in the range of $V$ or lower, the changes in polypeptide chain abundances in the LP→SP shift may follow trends that differ substantially from those characterized above for the transcripts. The trend of decreasing change-folds with increasing $E_d$, $P$ and $n$ hold also — and in the cases of $E_d$ and $n$ become even more pronounced — for the polypeptide chain abundance change-folds estimated according to (1.6) (Table 3). However, in remarkable contrast to the transcriptome-level trends, $m$ decreases with increasing $e_d$ and $D$.

## 4. DISCUSSION

Altogether, the results point to cost rates of maintaining the steady state concentrations of polypeptide chains as a major factor shaping gene expression at a genome-wide scale. This assertion is supported by the following two main observations. First, in LP a relatively strong and highly significant negative correlation between polypeptide chain abundances and per-molecule cost rates. Second, in the shift from a substrate-rich to a substrate-depleted medium a stronger down-regulation of the transcripts coding for the polypeptide chains that would be most expensive to maintain at their initial levels.

The first result might also ensue from adaptation of per-molecule cost rates, as these are determined by three evolutionarily adaptable factors: mean aminoacid biosynthetic cost (a function of aminoacid usage), primary sequence length, and turnover numbers. Indeed, our results indicate that each of these three factors tends to decrease as polypeptide chain abundances increase. However, they have a comparatively small adaptive capacity, as the following calculations highlight. The ratios between highest and lowest values of $a$, $n$, $D$+$V$ and $e$ for the set of 3009 polypeptide chains under consideration are 2.3, 91 and 46, respectively, compared to 25545 for $P$ under LP. Furthermore, the fact that the medians of $a$ and $n$ decrease mainly at relatively modest polypeptide chain abundances and do not further decrease for higher $P$ suggest that typical abundant proteins ($P$>50000 molecules/cell) have reached the adaptability limits for size and aminoacid composition. Smaller proteins, or proteins made of energetically cheaper aminoacids may fail to fold stably, to effectively interact with other proteins, or to be functional in some other way. In turn, the cell doubling time sets the value of $V$ as a lower limit on turnover numbers. Presumably owing to the various constraints discussed above, $e$ varies by a factor of 511, which can account by less than 1% of the >70000-fold variation of $E$ in the same set of polypeptide chains. On the other hand, polypeptide chain abundances can adapt over a wide range and through several alternative means: promoter mutations, transcription factor mutations, mutations in the Kozak sequence, changes in codon usage, etc.. Therefore, the observed correlation between $e$ and $P$ should reflect mainly evolutionary adaptation of $P$ via changes in protein synthesis rates.

In the LP→SP shift, the trend of decreasing mRNA and (estimated) polypeptide chain abundance change-folds with increasing $E_d$ ensues mainly from stronger down-regulation of transcripts coding for initially abundant polypeptide chains than of those coding for rare chains. The simple fact that proteins that are very abundant in the first condition are less likely to be needed at similarly high abundance in the second condition than proteins that are initially rare may explain this differential down-regulation without invoking a stronger selective pressure for decreasing $E_d$ in SP. However, the inference (Table 3) that polypeptide chains with higher per-molecule cost rate or primary sequence length are more down-regulated than those with lower values of these indices suggests that this selective pressure is indeed stronger in SP. This assertion is not inconsistent with the observation that transcripts coding for polypeptide chains with higher values of $a$ are equally or less down-regulated in the LP→SP shift as those coding for polypeptide chains with lower values of $a$. Contrary to growing cells, which accumulate protein into biomass over time, overall, quiescent cells (in SP) only recycle protein. Because the latter process does not entail a net incorporation of aminoacids into protein, aminoacid biosynthetic costs do not contribute to cost rates of polypeptide chains in SP. Therefore, the selective pressure for decreasing $a$ is weaker in SP than in LP.

## 5. ACKNOWLEDGMENTS

Work financed in part through Portuguese-Spanish integrated action E-6/07 and grant BFU2005-0234 from Spanish Ministerio de Educacion y Ciencia.

## 6. REFERENCES

[1] Wagner, A. Energy constraints on the evolution of gene expression. Mol Biol Evol, 22, 6 2005), 1365-1374.

[2] Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K. and Weissman, J. S. Global analysis of protein expression in yeast. Nature, 425 2003), 737-741.

[3] Moldave, K. Eukaryotic protein synthesis. Annu Rev Biochem, 541985), 1109-1149.

[4] Belle, A., Tanay, A., Bitincka, L., Shamir, R. and O'Shea, E. K. Quantification of protein half-lives in the budding yeast proteome. PNAS, 103(08/17/ 2006), 13004-13009.

[5] Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D. and Brown, P. O. Genomic expression programs in the response of yeast cells to environmental changes. Molecular Biology of the Cell, 11, 12 2000), 4241-4257.