Chapter 11

Mechanism-Based Clustering of Genome-Wide RNA Levels: Roles of Transcription and Transcript-Degradation Rates

Sungchul Ji

Department of Pharmacology and Toxicology, Rutgers University
170 Frelinghuysen Rd., Piscataway, NJ 08855, USA
E-mail: sji@rci.rutgers.edu

W. Art Chaovalitwongse*

Department of Industrial and Systems Engineering, Rutgers University
96 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Email: wchaoval@rci.rutgers.edu

Nina Fefferman

DIMACS, Rutgers University
96 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Department of Public Health and Family Medicine, Tufts School of Medicine
Boston, MA, 02111
Email: nina.fefferman@tufts.edu

Wonsuk Yoo

Department of Internal Medicine, Wayne State School of Medicine
4201 St. Antoine St., UHC 4H-30, Detroit, MI 48201
Email: wyoo@med.wayne.edu

Jose E. Perez-Ortin

Department of Biochemistry and Molecular Biology, University of Valencia
Dr. Moliner 50, 46100 Burjassot, Spain
Email: jose.e.perez@uv.es

DNA array techniques invented over a decade ago enable biologists to measure tens of thousands of mRNA levels in cells simultaneously as functions of envi-

*Corresponding Author.

ronmental perturbations. In a few cases the same technique has been employed to measure not only genome-wide transcript levels (*TL*) but also the associated transcription rates (*TR*) simultaneously. Since *TL* is determined by the balance between two opposing processes, *i.e.*, transcription and transcript degradation, simple theoretical considerations indicate that it would be impossible to determine *TR* based on *TL* data alone. This conclusion is supported by the finding that *TL* and *TR* do not always vary in parallel. In fact, the genome-wide measurements of *TL* and *TR* in budding yeast undergoing glucose-galactose shift indicate that *TL* can decrease even though *TR* increases and *TL* can increase despite the fact that *TR* decreases. These counter-intuitive findings cannot be accounted for unless transcript-degradation rates (*TD*) are also taken into account. One of the main objectives of this contribution is to derive a mathematical equation relating *TL* to *TR* and *TD*. Based on this equation, it was predicted that there would be 9 different mechanisms by which *TL* can be altered in cells. The *TL* and *TR* data measured in budding yeast demonstrate that all of the 9 predicted mechanisms are found to be activated in budding yeast during glucose-galactose shift, except Mechanisms 5 (i.e., decreasing *TL* with no change in *TR*) and 9 (i.e., no change in *TL* nor in *TR*). It was also shown that the opposite changes in the mRNA levels of glycolytic and respiratory genes observed between 5 and 360 minutes following the glucose-galactose shift could be quantitatively accounted for in terms of what is referred to as the transcript-degradation/transcription (*D/T*) ratios calculated here for the first time. Our results suggest that the predicted 9 mechanisms of controlling *TL* may be employed to cluster the genome-wide measurements of mRNA levels as a means to characterize the functional states of both normal and diseased cells..

11.1. Introduction

The DNA array technique allows cell biologists to measure the intracellular levels of tens of thousands of different kinds of mRNA molecules in living cells simultaneously. [1, 4, 22, 28, 30, 33] When mRNA levels are measured from a cell preparation as a function of time after some perturbation and the resulting data are subjected to a cluster analysis, it is frequently found that the mRNA levels can be grouped into a set of distinct clusters, each cluster exhibiting a common kinetics or a temporal pattern of the changes in mRNA levels.

It has been an almost universal practice in the field of microarray technology since its inception to interpret mRNA level changes in terms of transcription (*i.e.*, the synthesis of mRNA using DNA as template) rates only, without taking into consideration the transcript-degradation step. [23] It is well known that this transcript-degradation process can occur with rates comparable to those of transcription itself. [1, 28, 29] Ignoring the role of the transcript degradation step is tantamount to equating *transcript levels* with *transcription rates* and this has led

to the following erroneous interpretations of DNA microarray data:

(i) When mRNA levels increase, it is interpreted as an indication of increased rates of transcription of the corresponding genes;
(ii) When mRNA levels undergo no change, it is taken as the evidence for unchanged transcription rates; and
(iii) When mRNA levels decrease, it is interpreted as an indication for decreased transcription rates.

For convenience, we will refer to this way of interpreting mRNA levels as the *1-to-1 interpretation*. Based on this approach, it has been widely assumed that, when a *mRNA cluster* was found by various clustering techniques, this fact can be used to infer that the underlying genes are transcribed with similar rates and hence that there exists a corresponding *gene cluster*. It is one of the main objectives of this chapter to demonstrate that this way of interpreting mRNA clusters is theoretically invalid and factually unsupported, leading to Type I and Type II errors. A Type I error (or a false positive) can arise, for example, when an increase in mRNA level is interpreted as indicating an increase in the associated transcription rate (which can be true sometimes but not always), since the level of a mRNA molecule can increase even if the associated transcription rate does not change as long as the number of mRNA molecules synthesized during the time period of observation is greater than the number of mRNA molecules degraded during the same time period. Similarly, a Type II error (or a false negative) can arise if a gene is inferred to undergo no change in its transcription rate based on the fact that its mRNA level did not change. This is because, even if a mRNA level did not change, the transcription rate could have increased (or decreased) if the changes in the rate of mRNA degradation happened to exactly counterbalance the effect of an increased transcription rate.

Direct experimental evidence for the concept that mRNA levels, also known as transcript levels (*TL*), is determined by a dynamic balance between transcription and transcript degradation has been obtained only recently when *TL* and transcription rates (*TR*) were measured simultaneously from human lung carcinoma cells [8], tobacco plant cells [19] and the budding yeast *Sacharomyces cerevisiae (S. cerevisiae)* subjected to glucose-galactose shift. [10] Here we present the results of a genome-wide analysis of the yeast *TL* and *TR* data reported in Garcia-Martinez *et al.* [10] based on a kinetic equation relating *TL* to *TR* and transcription degradation rates (*TD*), demonstrating the following points:

- *TL* and *TR* increase together (see Mechanism 2 in Table 1 and Fig. 11.3) in 51% of the time and decrease together (see Mechanism 6) 40% of the

time.

- The opposite changes in glycolytic and respiratory *TL* (See Fig. 11.4) induced by glucose-derepression are due to opposite changes in *YD*, thus establishing an instance of degradational control in contrast to the well-known transcriptional control.
- There are 9 mechanisms underlying underlying the changes in *TL* (See Fig. 11.3).

11.2. Materials and Data Acquisition

11.2.1. *Glucose-galactose shift experiments*

The *S. cerevisiae* yeast strain BQS252 was grown overnight at 28 degrees Celsius in YPD medium (2% glucose, 2% peptone, 1% yeast extract) to exponential growth phase ($OD_{600} = 0.5$). [10] Cells were recovered by centrifugation, resuspended in YPGal medium (2% galactose, 2% peptone, 1% yeast extract), and allowed to grow in YP Gal medium for 14-15 hours after the glucose-galactose shift. Cell samples were taken at 0 (denoted as t_0), 5 (t_1), 120 (t_2), 360 (t_3), 450 (t_4) and 850 (t_5) minutes after the glucose-galactose shift. The t_5 sampling time corresponds to the exponential growth phase in YPGal medium. Two different aliquots were taken from the cell culture at each sampling time. One aliquot was processed to measure *TR* according to the genomic run-on protocol (see the next section), and the other was processed to measure *TL* using the same DNA arrays recovered after *TR* measurements.

11.2.2. *Measuring transcription rates (*TR*) using the genomic run-on (*GRO*) method*

The experimental details of the genomic run-on procedures are given in Garcia-Martinez *et al.* [10]. This technique is a scaled-up version of the usual nuclear run-on method. [13] Lysed cells contain transcription complexes stalled on the DNA template due to lack of ribonucleotides. Transcription is re-initiated *in vitro* by adding new nucleotides, one of which is radiolabled (*e.g.*, [$\alpha-^{33}$P]UTP). After allowing transcription to finish, one can determine via autoradiography the density of RNA polymerases for each gene. Assuming a constant speed for RNA polymerase II molecules this density allows us to estimate the transcription rates in the cells of interest. By comparing the amount of gene-specific radiolabeled RNA synthesized in one nuclei preparation with another, it is possible to estimate the extent of the RNA polymerase densities and, therefore, transcription rates in

the cells of interest. The *TR* data utilized in the *TL-TR* plots shown in Fig. 11.1 are available at *http://scsie.uv.es/chipsdna/chipsdna-e.html*.

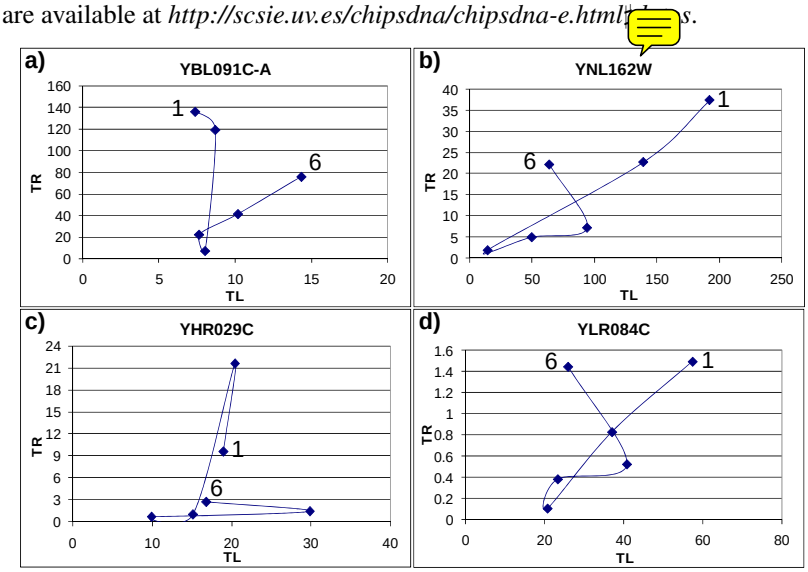

Fig. 11.1. Plots of the fold changes in transcription rates (*TR*) against those of transcript levels (*TL*) measured in budding yeast at 6 time points (1 = 0 min, 2 = 5 min, 3 = 120 min, 4 = 360 min, 5 = 450 min, and 6 = 850 min) after the glucose-galactose shift.

11.2.3. *Measuring mRNA or transcript levels (*TL*)*

The total RNA isolated from the same cell culture used in the previous section was reverse-transcribed into cDNA in the presence of [$\alpha-^{33}$P]dCTP. [10] The labeled cDNAs were purified and hybridized under the same conditions as described for *GRO* in order to minimize the variability due to artifacts of DNA membrane arrays. Again, the raw *TL* data in arbitrary units measured in triplicates at 6 time points are available at *http://evalga.uv.es/scsie-docs/chipsdna/chipsdna-e.html*.

11.2.4. *The* TL-TR *Plots*

The *TL* and *TR* data measured as described in the previous sections can be visualized in a 2-dimensional plane as shown in Fig. 11.1. The genes depicted in this figure were randomly chosen out of the 5,725 genes showing no missing values in their triplicate measurements of *TL* and *TR*. The notation given on the top of each figure is the name of the open reading frame (ORF) whose *TL* and TR values were measured. The trajectory of each plot can be divided into 5 segments or vectors bounded by two time points (*e.g.*, vector 1-2 between 0 to 5 min after the glucose-

galactose shift and vector 2-3 between 5 to 120 min, etc.). Each vector can be characterized in terms of the angle measured counterclockwise starting from the positive x-axis (see Fig. 11.2). For example, vector 5-6 in Fig. 11.1a is approximately $45°$ and vector 1-2 in Fig. 11.1b is approximately $225°$. Thus, the angle α determining the direction of the vector from the i^{th} point to the $(i+1)^{th}$ point in a *TL* versus *TR* plot with coordinates (x_i, y_i) and (x_{i+1}, y_{i+1}), respectively, can be calculated from the relation $\alpha = \tan^{-1}[\frac{(y_{i+1}-y_i)}{(x_{i+1}-x_i)}] + \Theta$, where $\Theta = 0°$ if both the numerator and the denominator are positive, $\Theta = 180°$ if either the numerator is positive and the denominator is negative or both the numerator and the denominator are negative, $\Theta = 360°$ if the numerator is negative but the denominator is positive.

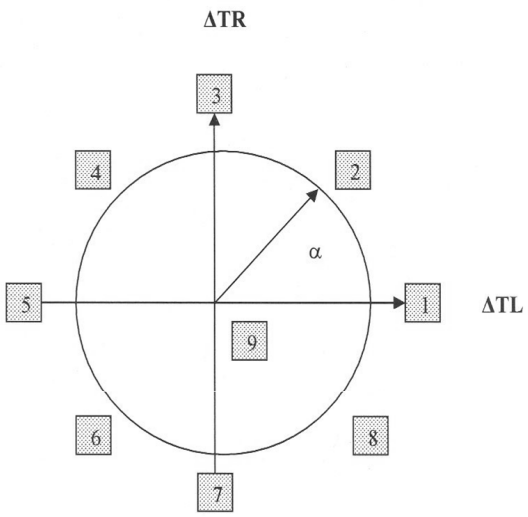

Fig. 11.2. The "unit" circle whose x-axis indicates the changes in *TL* and y-axis those in *TR* values of a trajectory in the *TL-TR* plot. The direction of the radius of the circle coincides with the direction of the component vector of a *TL-TR* trajectory. For convenience, each direction is defined by the following values of angle α: α_1 = 357 − 3; α_2 = 3 − 87; α_3 = 87 − 93; α_4 = 93 − 177; α_5 = 177 − 183; α_6 = 183 − 267; α_7 = 267 − 273; α_8 = 273 − 357. The center of the circle denoted as 9 indicates that there is no change in *TR* nor in *TL* between two time points.

11.3. Statistical analysis

The total number of genes whose *TL* and *TR* values were measured was 5,913, of which 5,725 genes were successfully measured in triplicates without any missing values. The rest of the genes were measured in less than triplicates. A statistical analysis for the comparison between the "expected" and the "observed" distribu-

tions of genes over 8 possible mechanisms (excluding Mechanism 9) of controlling mRNA level in cells(defined in the previous section) was performed using the binomial test to make inferences about a proportion for response rate based on a series of independent observations. A hypothesis test was performed to compare the average trajectories (also called sequences, time series, or profiles) of two groups of genes coding for glycolytic and respiratory mRNA molecules. Each group generates two average profiles, labeled *TL* and *TR* (See Fig. 11.3a). Thus, the null hypothesis states that "there is no difference between the two profiles associated with glycolysis and respiration", and the alternative hypothesis states that "There exists a significant difference between the two profiles". In order to test these hypotheses, a multivariate approach for repeated measures analysis was used to calculate *p*-values. The *p*-value for the two trajectories shown in Fig. 11.3a is less than 0.0001, indicating that there are significant differences between the average glycolytic and respiratory mRNA level trajectories. The *p*-value for the two curves shown in Fig. 11.3b is 0.2292, indicating that there is no significant difference between the average trajectories of the glycolytic and respiratory mRNA synthesis rates.

11.3.1. *Calibration of* TL *data*

To convert the *TL* data expressed in arbitrary unit to the corresponding values in absolute unit (mRNA molecules per cell), we utilized the reference mRNA abundance data complied by Beyer *et al.* [5] based on 36 datasets reported in the literature. Out of about 5,700 mRNA abundance data that these authors collected, we selected a total of 59 glycolytic and respiratory mRNA abundance values and plotted them against the corresponding *TL* data, leading to a linear regression line satisfying the relation, $n = 0.1114TL_0 - 7.220$, where n is the number of mRNA molecules per cell, and TL_0 is the mRNA level measured in arbitrary unit at $t = 0$. The correlation coefficient of the straight line was 0.94245. This equation was used to convert all the *TL* data (measured at five different time stamps) in arbitrary unit into the corresponding values in absolute unit.

Examples of the time courses of the average changes in *TL*, *TR*, and the transcript-degradation/transcription (*D/T*) ratios of glycolytic and respiratory genes are shown in Fig. 11.3 and 11.4. The glycolytic genes considered here include *PDC2, PFK1, PFK2, ADH5, PDC6, LAT1, PDA1, ADH2, [illegible]K1, PGM1, ADH3, TPI1, GCR1*, and *PDB1* and the respiratory genes include *QCR6, COX13, COX9, NDI1, CYT1, COX8, COX12, SDH4, COX5A, COX4, CYC1, QCR2, QCR8*, and *COR1*.

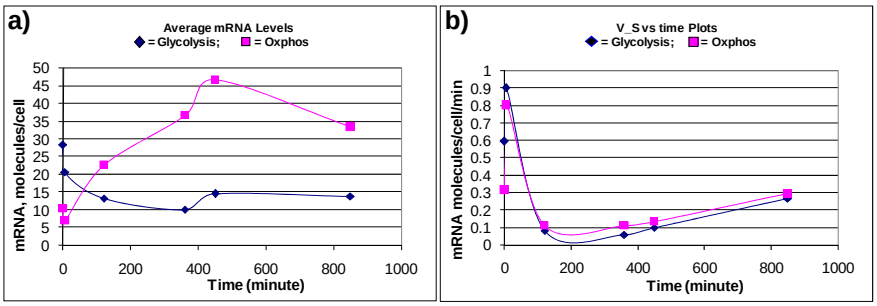

Fig. 11.3. The time courses of the average changes in the transcript level (*TL*) and transcription rate (*TR*) of 12-15 glycolytic (⧫) and 14 respiratory genes (■).

11.3.2. *Calibration of* TR *data*

Transcription rates of yeast genes at t_0 were estimated by utilizing the genome-wide mRNA decay half-lives reported by Wang *et al.* [32] The half-life data were down-loaded from their web site (*http://www-genome.stanford.edu/turnover*). The mathematical relation between transcription rate, $dn_{s,i}/dt$, where $n_{s,i}$ is the number of the i^{th} mRNA molecules synthesized per cell during the time interval dt, and the $mRNA_i$ decay half-life, denoted by $t_{\frac{1}{2},i}$, can be derived based on the following assumptions.

(i) At t_0, budding yeast cells are at a steady state with respect to transcription and transcript degradation. In other words, at t_0, $dn_{S,i}/dt = dn_{D,i}/dt$, where $n_{D,i}$ is the number of the i^{th} mRNA molecules per cell degraded during the time interval dt.

(ii) The decay of the i^{th} mRNA molecules obeys a first-order rate law given by

$$-dn_i/dt = dn_{D,i}/dt = k_{D,i}[mRNA_i], \tag{11.1}$$

where $k_{D,i}$ is the first-order degradation rate constant and $[mRNA_i]$ is the concentration of the i^{th} mRNA in the cell. Integrating Eq. (11.1) with respect to time leads to

$$[mRNA_i] = [mRNA_i]_0 e^{-k_{D,i}t}, \tag{11.2}$$

where $[mRNA_i]_0$ is the $mRNA_i$ abundance at t_0. Substituting the values, $[mRNA_i] = [mRNA_i]_0/2$ at $t = t_{\frac{1}{2},i}$, and solving the resulting equation for $k_{D,i}$ yields

$$k_{D,i} = \frac{\ln 2}{t_{\frac{1}{2},i}} = \frac{0.693}{t_{\frac{1}{2},i}}. \tag{11.3}$$

Using this relation, it is possible to convert mRNA decay half-lives measured by Wang *et al.* [32] into the corresponding transcript decay rate constants.

(iii) Since budding yeast cells are assumed to be at steady states at t_0 with respect to transcription and transcript decay (previously mentioned in the first assumption), it would follow that

$$dn_{S,i}/dt = dn_{D,i}/dt = k_{D,i}[mRNA_i]_0 = \frac{0.693[mRNA_i]_0}{t_{\frac{1}{2},i}}. \quad (11.4)$$

Thus Eq. (11.4) allows us to estimate the transcription rate, $dn_{S,i}/dt$, at t_0 from the $[mRNA_i]_0$ and $t_{\frac{1}{2},i}$ values.

(iv) The conversion factor, defined as $a_i = (dn_{S,i}/dt)/TR_i$ calculated at t_0 is assumed to apply to all the other TR_i values measured at t_1, t_2, t_3, t_4, and t_5, where TR_i denoting the *TR* values associated with the i^{th} gene.

(v) Since the $mRNA_i$ decay measurements were made at 37° Celsius (C) [32], whereas the TR_i measurements were carried out at 28° C [10], the absolute rate values for $dn_{S,i}/dt$ calculated in (iii) were corrected for the temperature difference by dividing $dn_{S,i}/dt$ by the factor $2 \times (9/10) = 1.9$, which results from the assumption that the Q10 value (*i.e.*, the factor by which the rate increases due to a 10° C increase in temperature) was 2.

11.3.3. *Kinetic analysis of the changes in mRNA levels*

In the absence of any exchange of mRNA molecules between budding yeast cells and their environment, it is possible to equate the rate of change of the i^{th} mRNA molecules per cell, dn_i/dt, with the balance between the rate of $mRNA_i$ synthesis, $dn_{S,i}/dt$, and the rate of its degradation, $dn_{D,i}/dt$ given by

$$dn_i/dt = dn_{S,i}/dt - dn_{D,i}/dt \quad (11.5)$$

In general, the rate of synthesis of the i^{th} mRNA, $dn_{S,i}/dt$, would be a function of many variables including the levels (or concentrations) of RNA polymerase II, various transcription factors (encoded by the j^{th} gene, where $j \neq i$, except when the j^{th} gene happens to code for a transcription factor that acts on the j^{th} gene itself), and small molecules such as $ATP, ADP, AMP, Mg^{++}, H^+$, etc. The same would hold true for the rate of the degradation of the i^{th} mRNA, $dn_{D,i}/dt$. A system of ordinary differential equations describing the dynamics of mRNA

levels in a cell taking into account all the variables mentioned above has been derived in several previous studies. [6, 26, 27] Our model in Eq. (11.5) is similar to (but not identical with) Eq. (1) in the studies by Savageau [26, 27].

The change in the number of $mRNA_i$ molecules per cell, Δn_i, during the time interval, Δt, between two time points, t_k and t_{k+1}, can be expressed as

$$\Delta n_i = \mathsf{int}(d)n_i = \mathsf{int}(d)n_{S,i} - \mathsf{int}(d)n_{D,i} = \Delta n_{S,i} - \Delta n_{D,i}, \tag{11.6}$$

where the integration is from t_k to t_{k+1}. The resulting equation, $\Delta n_i = \Delta n_{S,i} - \Delta n_{D,i}$, simply states that the change in the number of the i^{th} mRNA molecules in a cell during the time interval, $\Delta t = t_{k+1} - t_k$, is determined by the balance between the number of $mRNA_i$ molecules synthesized and the number of $mRNA_i$ molecules degraded during that time interval. It should be noted here that the variables in Eq. (11.6) are in absolute units. We can estimate the variables in Eq. (11.6) as follows:

$$\Delta n_i = n_i - n_{i-1} \tag{11.7}$$

$$\Delta n_{S,i} = AUC_S \tag{11.8}$$

$$\Delta n_{D,i} = AUC_S - (n_i - n_{i-1}), \tag{11.9}$$

where the subscript i refers to the i^{th} mRNA molecule, and AUC_S is the area under the curve of the *TR*-time plots.

11.3.4. *Transcript-degradation to transcription (*D/T*) ratios*

The transcript-degradation/transcription (*D/T*) ratio for the i^{th} mRNA molecule can be defined by

$$D/T = \frac{\Delta n_{D,i}}{\Delta n_{S,i}}. \tag{11.10}$$

Note that the subscript i is omitted from the term, D/T, for simplicity. Combining Eqs. (11.6) and (11.10) results in

$$\Delta n_i = \Delta n_{S,i}(1 - D/T). \tag{11.11}$$

We observe that the sign of Δn_i is determined solely by the magnitude of the *D/T* ratio relative to 1. The relationship can be described by

$$\Delta n_i > 0 \text{ if } D/T < 1 \tag{11.12}$$

$$\Delta n_i = 0 \text{ if } D/T = 1 \tag{11.13}$$

$$\Delta n_i < 0 \text{ if } D/T > 1. \tag{11.14}$$

The time-dependent variations of the *D/T* ratios of the glycolytic and respiratory genes were calculated as explained in the previous section. Fig. 11.4 illustrates

the temporal evolution of *D/T* ratios of the glycolytic and respiratory genes given in Fig. 11.3. It should be pointed out that there are 2 less respiratory genes than the one given in Fig. 11.3b due to missing values in the degradation rates. The *D/T* ratio for the i^{th} mRNA molecule (DTR_i) is defined as $\frac{\Delta n_{D,i}}{\Delta n_{S,i}}$ where n_i is the number of the i^{th} mRNA molecules synthesized (S) or degraded (D) per cell over a given time period (see the previous section for more details). The calculation of $\Delta n_{S,i}$ requires integrating the TR_i versus time curves between two sampling time points, say t_1 and t_2. This is why there are only 5 $DTR's$, each of which was plotted at the mid-point of the two time points involved.

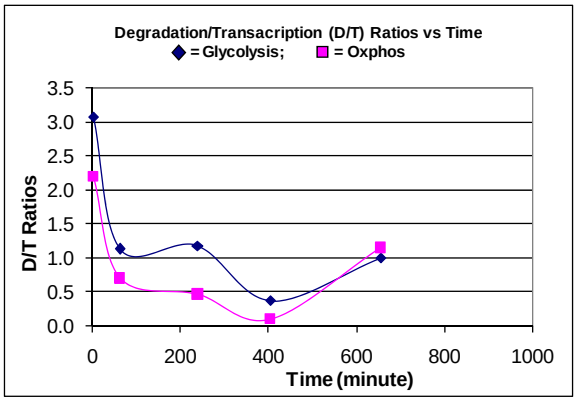

Fig. 11.4. The temporal variations of the transcript-degradation/transcription (D/T) ratios of the 15 glycolytic (◆) and 12 respiratory genes (■) given in Figure 11.3.

11.4. Experimental Results

In our previous publication [10], we reported that, when yeast cells were grown in a glucose-containing medium to exponential growth phase, harvested by centrifugation, and resuspended in a galactose-containing medium replacing glucose, extensive metabolic changes were found to occur as reflected in *TL* and *TR* values for almost all of the 6,400 genes in the yeast genome. The *TL* trace averaged over 5,753-5,829 genes decreased by about 65% during the first 5 minutes following the glucose-galactose shift and continued to decline until 360 minutes by further 20%. There was a slight increase (17%) in *TL* between 360 and 450 minutes, which was followed by a smaller decrease (12%) between 450 and 850 minutes. In contrast, the *TR* values averaged over 5,753-5,829 genes exhibited very different kinetic behaviors: A rapid increase by about 60% during the first 5 minutes followed by a 150% decrease to the 10% level of the control by 120 minutes. This low level of *TR* was maintained until 360 minutes when it started to increase

linearly up to the 45% level of the control by 850 minutes. The *TL* and *TR* values were normalized with respect to their values at t = 0, and corrected for the increasing number of cells during the observational period.

These observations clearly indicate that the average *TL* and *TR* values do not change in parallel (as would have been expected if *TL* reflected the rates of gene expression, namely *TR*) but exhibit seemingly *independent* kinetic behaviors, especially during the first (0-5 minutes) and the last (450 - 850 minutes) periods of observation. The relation between *TL* and *TR* can be visualized by plotting *TL* against *TR* on a 2-dimensional plane as shown in Fig. 11.1a through 11.1d. A trajectory in the *TR* versus *TL* plane consists of a set of 5 vectors, each representing the change in *TL* and *TR* over a time interval which varies from 5 minutes to 400 minutes. The trajectory associated with the gene YBL091C-A consists of a series of five connected vectors starting from time point 1 (0 minute) and ending at time point 6 (850 minutes) having approximate angles of $315^\circ, 270^\circ, 135^\circ, 45^\circ$, and 45°. A given vector is associated with a mechanism of RNA metabolism expressed in terms of *TL* and *TR*. Specifically, when $\alpha = 315^\circ$, it indicates that, during the first time interval (*i.e.*, from 0 to 5 minutes), the transcript level *TL* of gene TBL091C-A is increased but its transcription rate *TR* is decreased; when $\alpha = 270^\circ$, it indicates that the same gene experienced no change in *TL* but a decrease in *TR*, *etc*. It is important to note that this gene experienced an increase in both *TL* and *TR* only during the last two time intervals, since their associated α values are approximately 45°. The 9 mechanisms of RNA metabolism depicted in Fig. 11.2 are defined as follows:

(1) **Mechanism 1** is defined by the α values ranging from 357° (or -3°) to 3° and indicates that *TL* increases without any change in *TR*;
(2) **Mechanism 2** is defined by the α values ranging from 3° to 87° and indicates that both *TL* and *TR* increase;
(3) **Mechanism 3** is defined by the α values ranging from 87° to 93° and indicates that *TL* does not change although *TR* increases;
(4) **Mechanism 4** is defined by the α values ranging from 93° to 177° and indicates that *TL* decreases even though *TR* increases;
(5) **Mechanism 5** is defined by the α values ranging from 177° to 183° and indicates that *TL* decreases without any change in *TR*;
(6) **Mechanism 6** is defined by the α values ranging from 183° to 267° and indicates that both *TL* and *TR* decreases;
(7) **Mechanism 7** is defined by the α values ranging from 267° to 273° and indicates that *TL* does not change although *TR* decreases; and
(8) **Mechanism 8** is defined by the α values ranging from 273° to 357° and

indicates that *TL* increases even though *TR* decreases.

Note that **Mechanisms 9** is not shown above because it is defined not by α but by the values of ΔTL and ΔTR both being zero.

Since there are 5 vectors per *TL-TR* trajectory, there are a total of $5 \times 5,725 = 28,625$ values of α to be calculated. These values are calculated using the formula given in the previous section and displayed in Table 11.1. The rows numbered 1 through 5 refer to the 5 time intervals and the columns numbered 1 through 8 refer to the mechanisms of controlling mRNA levels. Due to a large number of genes involved, we used the normal approximation approach, leading to the results that the observed proportions for Mechanisms 1, 2, 3, 4, 6, 7, and 8 are significantly different from the expected except for Mechanism 5. In other words, all of the 8 mechanisms are observed to occur in budding yeast cells under the experimental conditions employed, except Mechanisms 5 and 9.

Table 11.1. The frequency distributions of the 8 modules of RNA metabolism (defined in the legend to Fig. 11.2) as the functions of the 5 time periods following the glucose-galactose shift.If the angles are randomly distributed over 360°, the expected distributions can be calculated as shown in the 9^{th} row. The *p*-values for the difference between the observed and the expected distributions are given in the last row. The differences are all significant, except for Mechanism 5.

	Mechanism								
Vector	1	2	3	4	5	6	7	8	Total
1	0	142	234	3470	96	1732	12	39	5725
2	14	18	3	37	5	3729	617	1302	5725
3	340	1914	52	638	314	1471	28	968	5725
4	477	4237	21	151	61	143	19	616	5725
5	12	1151	238	4213	38	56	4	13	5725
Total Observed	843	7462	548	8509	514	7131	680	2938	28625
Total Expected	477	6678	477	6678	477	6678	477	6678	28625
p-value	< 0.0001	< 0.0001	0.001	< 0.0001	0.092	< 0.0001	< 0.0001	< 0.0001	

The numbers in Table 11.1 represent the frequencies of the different mechanisms that occur in budding yeast during one of the five time intervals. Thus, during the first time interval (*i.e.*, from 0 to 5 minutes), no gene experienced (or exhibited) Mechanism 1; 142 genes experienced Mechanism 2; 234 genes exhibited Mechanism 3; 3,470 genes experienced Mechanism 4; 96 experienced Mechanism 5; 1,732 genes experienced Mechanism 6; 12 genes experienced Mechanism 7; and 39 experienced Mechanism 8. If the various mechanisms occur randomly during this time interval, their frequency of distribution would be expected to be proportional to the magnitude of the angle α associated with the vector spanning the time intervals involved, *i.e.*, 6° for Mechanisms 1, 3, 5, and 7, and 84° for Mechanisms 2, 4, 6 and 8. The theoretically predicted distributions of the mechanisms based on the angular sizes are given in the 9^{th} row in Table 11.1. Comparing Rows 8 and 9, it is clear that all of the 8 mechanisms occur with frequencies

different from those expected on the basis of random distributions, except Mechanism 5, as evidenced by the fact that the associated p-values are all less than 0.001 except that associated with Mechanism 5.

In 2002, Gorospe and her group measured for the first time the *TL* and *TR* data for about 2,000 genes from non-small cell human lung carcinoma H1299 [8] and found that *TL* could increase or decrease without any changes in *TR* (see Groups *IV* and *V* in Table 1 [8]), from which they concluded that transcript degradation played a critical role in determining mRNA levels. The first genome-wide measurements of *TL* and *TR* in budding yeast *S. cerevisiae* were reported by Garcia-Martinez et al. [10], whose results also indicated that there were no *1-to-1* correlation between *TL* and *TR*. However, neither of these publications included any mathematical equation relating *TL, TR,* and *TD*. One of the main objectives of this paper is to fill this gap in our knowledge and use the derived equation to analyze the *TL* and *TR* data of functionally well-defined groups of mRNAs in order to investigate the possible functional roles of mRNA levels in cell biology. For this purpose, we chose the glycolytic and respiratory mRNA molecules for a detailed analysis because the biochemistry of glycolysis and respiration (leading to oxidative phosphorylation) and their antagonistic interactions are well known in *S. cerevisiae* during glucose-galsctose shift. [2, 7, 15, 16, 18]

The unicellular organism *S. cerevisiae* (also known as budding yeast, baker's yeast, or wine yeast) has the capacity to metabolize glucose and galactose but prefers the former as the carbon and energy sources when both nutrients are present in its environment. In the presence of glucose, the organism turns on those genes coding for the enzymes needed to convert glucose to ethanol (which phenomenon is known as *glucose induction*) and turns off those genes needed for galactose metabolism (which phenomenon is known as *glucose repression*). [2, 7, 15, 16, 18] The detailed molecular mechanisms underlying these phenomena (called *diauxic shift*) are incompletely understood at present and are under intensive studies. [11, 25, 31] When glucose is depleted, *S. cerevisiae* increases its rate of metabolism of ethanol to produce ATP via the Krebs cycle and mitochondrial respiration. [11, 25] This metabolic control is exerted by reversing the glucose repression of the genes encoding the enzymes required for respiration (*i.e.*, oxidative phosphorylation) – the process referred to as *glucose de-repression*. [11]

11.5. Conclusion and Discussion

Fig. 11.3a shows the time courses of the average levels of 14 each of the glycolytic and respiratory mRNA molecules during the 850 minutes of observation after shifting glucose to galactose. The time course of the average transcription rates of

the same sets of glycolytic and respiratory genes are displayed in Fig. 11.3b. The most striking feature of these figures is that, despite the similarity between the time courses of the transcription rates (*TR*) of glycolytic and respiratory genes, those of the corresponding transcript levels (*TL*) are quite different. In fact, the *TL* trajectories of glycolytic and respiratory genes change in opposite directions during the period between 5 and 360 minutes after glucose-galactose shift, whereas the corresponding *TR* trajectories almost coincide. As shown in Fig. 11.4, these opposite changes in *TL* appear to be the consequences of the opposite changes in the degradation rates of glycolytic and respiratory mRNA molecules.

The qualitative features of the temporal behaviors of *TL* and *TR* changes displayed in Fig. 11.3a and 11.4b are summarized in Table 11.2. As indicated in the first two rows, the total observational period of 850 minutes are broken down to 5 phases, labeled *I* through *V*. During Phase I, the transcript levels of both glycolytic and respiratory genes decrease precipitously although the corresponding transcription rates increase, most likely because the stress induced by glucose-galactose shift increase transcript degradation rates more than can be compensated for by increased transcription. This interpretation is supported by the transcription/degradation ratio of 0.5 calculated for Phase I (see Fig. 11.4). During Phases II and III, the glycolytic transcript levels decrease by 2 fold, whereas the respiratory transcript levels increase by 4 fold. Since the corresponding transcription rates of both the glycolytic and respiratory genes decline rapidly followed by a plateau, the increased respiratory mRNA levels cannot be accounted for in terms of transcriptional control but must implicate degradational control. That is, just as the removal of glucose "de-induces" glycolytic mRNA molecules (leading to the declining *TL* and *TR* trajectories for glycolysis in Fig. 11.3a and 11.3b), so it might repress the degradation of respiratory mRNA molecules, leading to a rise in respiratory mRNA levels as seen in Fig. 11.3a between the second and fourth time points. This phenomenon may be referred to as "glucose de-induction" in analogy to glucose induction. [7, 18, 20] If this interpretation is correct, one intriguing hypothesis suggests itself that glucose normally keeps the respiratory mRNA levels low by both enhancing the degradation and repressing the synthesis of respiratory mRNA molecules. During Phase IV, both *TL* and *TR* for glycolytic and respiratory genes increase, and this may be attributed to galactose induction. [11, 18, 34] In support of this interpretation, it was found that glucose-galactose shift induced an increase in both *TL* and *TR* of the *Leloir* genes (GAL 1, 2, 3, 7 and 10) between 120 and 450 minutes by more than 10 folds (data not shown). The *Leloir* genes code for the enzymes and transport proteins that are involved in converting extracellular galactose to intracellular glucose-1-phosphate [9], which is then metabolized via the glycolytic and respiratory pathways. Finally, during Phase

V, the glycolytic mRNA levels remain constant while the respiratory mRNA levels decline slightly, the latter probably due to galactose repression (in analogy to the well-known glucose repression [18]) of respiration following the formation of *glucose-1-phosphate* via the Leloir pathway. [9] The transcription rate of glycolytic genes continue to increase during Phase V probably due to galactose induction [21, 25], although the corresponding transcript levels remain unchanged, which may also indicate the degradational control of glycolytic mRNA molecules during this time period. That is, budding yeast seems able to keep glycolytic *TL* constant in the face of increasing *TR*, by increasing *TD* - the transcript degradation rate. The *TR* trajectory of respiratory genes also continue to increase during Phase V despite the fact that their *TL* trajectory decline, which can be best explained in terms of the hypothesis that that respiratory mRNA levels are controlled by transcript degradation. It is quite evident that the *TL* and *TR* data presented in Fig. 11.3a and 11.3b cannot be accounted for in terms of *TR* alone but requires taking into account both *TR* and *TD* on an equal footing for their logically consistent explications, which is tantamount to the conclusion that *TL* is determined by the *D/T* ratio (see Fig. 11.3a and 11.4).

Table 11.2. A summary of the kinetics of the *TL* and *TR* changes depicted in Fig. 11.3a and 11.4b. The upward and downward arrows indicate an increase and decrease, respectively.

Time (min)		0-5	5-120	120-360	360-450	450-850
Phase (or Time Period)		I	II	III	IV	V
Transcript Level (TL)	Glycolysis	↓	↓	↓	↑	No Change
	Oxidative Phosphorylation	↓	↑	↑	↑	↓
Transcription Rate (TR)	Glycolysis	↑	↓	↓	↑	↑
	Oxidative Phosphorylation	↑	↓	↓	↑	↑

In conclusion, the genome-wide *TL* and *TR* data of *S. cerevisiae* measured by Garcia-Martinez et al. [10] have provided us with a concrete experimental basis to establish the concept that mRNA levels measured with cDNA arrays cannot be interpreted in terms other than what is here called the transcription/degradation (D/T) ratios. These ratios have been found useful in characterizing the temporal evolution of the molecular mechanisms underlying mRNA level changes induced by glucose-galactose shift in this microorganism, which may be extended to other cellular systems for similar determinations. Since these mRNA level changes reflect the dynamic metabolic states of cells (*i.e.*, cell states) supported by dissipation of free energy, they qualify as examples of what was referred to as intracellular dissipative structures (*IDSs*), an example of the application of Prigogine's dissipative structure concept to molecular cell biology. [2, 14, 17] It was postulated that *IDSs* serve as the immediate driving forces for all cell functions and

reflect the functional states of the cell. If this interpretation turns out to be correct upon further investigations, the DNA array technique, due to its ability to measure D/T ratios as demonstrated here, may prove to be an invaluable experimental tool to characterize and investigate *IDSs* and their biological functions, leading to numerous applications in basic cell biology, biotechnology, and medicine, including developments of diagnostic procedures to recognize cancer cells in their early developmental stages and testing drug candidates for their ability to reverse such pathological cell states.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 0546574. The first author thanks R. Miura and S. Dhar of NJIT for their valuable contributions in the early phase of the research program described here.

References

[1] Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra,S., Mack , D. and Levine, A.J. (1999) "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays". *Proc. Nat. Acad. Sci. USA*, 96, 6745-6750.

[2] Ashe, M.P., De Long, S.K. and Sachs, A.B. (2000) "Glucose depletion rapidly inhibits translation initiation in yeast". *Mol. Biol. Cell*, 11, 833-848.

[3] Babloyantz, A. *Molecules, Dynamics and Life: An Introduction to Self-Organization of Matter*. Wiley-Interscience, New York, 1986.

[4] Berg, J.M., Tymoczko, J.L., Stryer, L. *Biochemistry, Fifth Edition*. W. H. Freeman and Company, New York, 2002, 159-160.

[5] Beyer, A., Hollunder, J., Nasheuer, H.-P., and Wilhelm, T. (2004) "Post-transcriptional expression regulation in the yeast saccharomyces cerevisiae on a genomic scale". *Mol. Cell. Proteomics*, 3, 1083-1092.

[6] Chen, T., He, H.L. and Church, G.M. (1999) "Modeling gene expression with differential equations". In: R.B. Altman, A.K. Dunker, L. Hunter, T.E. Klein, and K. Lauderdale, eds. *Pacific Symposium on Biocomputing*, World Scientific, Singapore, 29-40.

[7] DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997) "Exploring the metabolic and genetic control of gene expression on a genomic scale". *Science*, 278, 680-686.

[8] Fan, J., Yang, X., Wang, W., Wood, W.H., Becker, K.G. and Gorospe, M. (2002) "Global analysis of stress-regulated mRNA turnover by using cDNA arrays". *Proc. Nat. Acad. Sci.*, 99(16), 10611-10616.

[9] Fu, L., Bounelis, P., Dey, N., Browne, B.L., Marchase, R.B. and Bedwell, D.M. (1995) "The posttranslational modification of phosphoglucomutase is regulated by galactose induction and glucose repression in Saccharomyces cerevisiae". *J. Bacteriol.*, 7(11), 3087-3094.

[10] Garcia-Martinez, J., Aranda, A. and Perez-Ortin, J.E. (2004) "Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms". *Mol. Cell*, 15, 303-313.

[11] Gasch, A.P. (2003) "The environmental stress response: a common yeast response to diverse environmental stresses". In: S. Hohmann and W.H. Mager, eds. *Yeast Stress Responses*, Springer, Berlin, 11-70.

[12] Hargrove, J.L., and Schmidt, F.H. (1989) "The role of mRNA and protein stability in gene expression". *FASEB J.*, 3, 2360-2370.

[13] Hirayoshi, K. and Lis, J.T. (1999) "Nuclear run-on assays: assessing transcription by measuring density of engaged RNA polymerases". *Methods in Enzymology*, 304, 351-362.

[14] Ji, S. (1985) "The Bhopalator: a molecular model of the living cell based on the concepts of conformons and dissipative structures". *J. theoret. Biol.*, 116, 399-4265.

[15] Jona, G., Choder, M. and Gileadi, O. (2000) "Glucose starvation induces a drastic reduction in the rates of both transcription and degradation of mRNA in yeast". *Biochim. Biophys. Acta*, 1491, 37-48.

[16] Johnnston, M. (1999) "Feasting, fasting and fermenting: glucose sensing in yeast and other cells". *Trends Genetics*, 15(1), 29-33.

[17] Kondepudi, D., and Prigogine, I. *Modern Thermodynamics: From Heat Engine to Dissipative Structures*. John Wiley and Sons, Inc., Chichester, 1998.

[18] Kuhn, K.M., DeRisi, J.L., Brown, P.O. and Sarnow, P. (2001) "Global and specific translational regulation in the genomic response of saccharomyces cerevisiae to nonfermental carnbon source". *Mol. Cell. Biol.*, 21(3), 916-927.

[19] Legen, J., Kemp, S., Krause, K., Profanter, B., Hermann, R.G., and Maier, R.M. (2002) "Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries". *Plant J.*, 31, 171-188.

[20] Leuther, K.K. and Johnston, S.A. (1992) "Nondissociation of GAL4 and GAL80 in vivo after galactose induction". *Science*, 256(5061), 33-1335.

[21] Mosley, A.L., Lakshmann, J., Aryal, B.K. and zcan, S. (2003) "Glucose-mediated phosphorylation converts the transcription factor Rgt1 from a repressor to an activator". *J. Biol. Chem.*, 278 (12), 10322-10327.

[22] Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P. and Fodor, P.A. (1994) "Light-generated oligonucleotide arrays for rapid DNA sequence analysis". *Proc. Nat. Acad. Sci. USA*, 91, 5022-5026.

[23] Prigogine, I. (1978) "Time, structure, and fluctuations". *Science*, 201, 777-785.

[24] Rhodes, D.R. and Chinnaiyan, A.M. (2004) "Bioinformatics strategies for translatingh genome-wide expression analyses into clinically useful cancer markers". *Ann. N.Y. Acad. Sci.*, 1020, 32-40.

[25] Ronne, H. (1995) "Glucose repression in fungi". *Trends Genetics*, 11(1), 12-17.

[26] Savageau, M.A. (1969) "Biochemcial Systems Analysis I: Some mathematical properties of the rate law for the component enzymatic reactions". *J. theoret. Biol.*, 25, 365-369.

[27] Savageau, M.A. (1969) "Biochemcial Systems Analysis II: The steady-state solutions for an n-pool system using a power-law approximation". *J. theoret. Biol.*, 25, 370-379.

[28] Schena, S.M., Shalon, D., Davis D.R. and Brown, P.O. (1995) "Quantitative monitor-

ing of gene expression patterns with a complementary DNA microarray". *Science*, 91, 270, 467-470.

[29] Shapiro, D.J., Blume, J.E. and Nielsen, D.A. (1986) "Regulation of messenger RNA stability in eukaryotic cells". *BioEssays*, 6(5), 221-226.

[30] Troester, M.A., Hoadley, K.A., Sorlie, T. , Herbert, B.-S., Borresen-Dale, A.-L., Lonning, P.E., Shay, J.W., Kaufmann, W.K., and Perou, C.M. (2004) "Cell-type-specific responses to chemotherapeutics in breast cancer". *Cancer Research*, 64, 4218-4226.

[31] Tu, B.P., Kudlicki, A., Towicka, M. and McKnight, S.L. (2005) "Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes". *Science*, 310, 1151-1158.

[32] Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. (2002) "Precision and functional specificity in mRNA cay". *PNAS*, 99(9), 5860-5865.

[33] Watson, S.J. and Akil, U. (1999) "Gene chips and arrays revealed: A primer on their power and their uses". *Biol. Psychiatry*, 45, 533-543.

[34] Winderickx, J., Holsbeeks, I., Lagatie, O., Giots, F., Thevelein, J., and de Winde, H. (2003) "From feast to famine; adaptation to nutrient availability in yeast". In: S. Hohmann and W.H. Mager, eds. *Yeast Stress Responses*, Springer, Berlin, 305-386.

[35] Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schoeder, M., Magnasco, M., and Darnell, J.E., Jr. (2003) "Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes". *Genome Research*, 13, 1863-1872.