

Free Energy Rhythms in *Saccharomyces cerevisiae*: A Dynamic Perspective with Implications for Ribosomal Biogenesis

A. Gross,[†] Caroline M. Li,[‡] F. Remacle,^{†,§} and R. D. Levine^{*,†,||}

[†]The Fritz Haber Research Center, Hebrew University, Jerusalem 91904, Israel

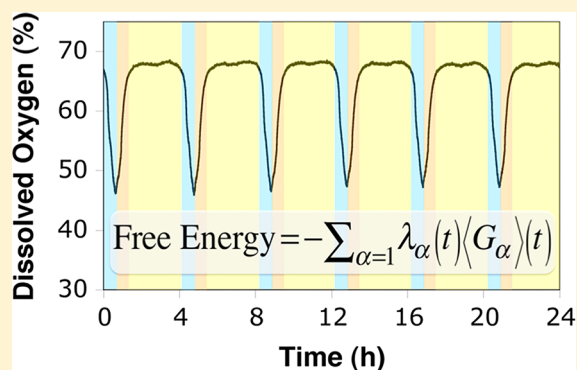
[‡]Department of Molecular and Cellular Biology, City of Hope Beckman Research Institute, Duarte, California 91010, United States

[§]Département de Chimie, B6c, Université de Liège, B4000 Liège, Belgium

^{||}Department of Chemistry and Biochemistry, Crump Institute for Molecular Imaging, and Department of Molecular and Medical Pharmacology, University of California, Los Angeles, California 90095, United States

Supporting Information

ABSTRACT: To describe the time course of cellular systems, we integrate ideas from thermodynamics and information theory to discuss the work needed to change the state of the cell. The biological example analyzed is experimental microarray transcription level oscillations of yeast in the different phases as characterized by oxygen consumption. Surprisal analysis was applied to identify groups of transcripts that oscillate in concert and thereby to compute changes in free energy with time. Three dominant transcript groups were identified by surprisal analysis. The groups correspond to the respiratory, early, and late reductive phases. Genes involved in ribosome biogenesis peaked at the respiratory phase. The work to prepare the state is shown to be the sum of the contributions of these groups. We paid particular attention to work requirements during ribosomal building, and the correlation with ATP levels and dissolved oxygen. The suggestion that cells in the respiratory phase likely build ribosomes, an energy intensive process, in preparation for protein production during the S phase of the cell cycle is validated by an experiment. Surprisal analysis thereby provided a useful tool for determining the synchronization of transcription events and energetics in a cell in real time.



Metabolic processes produce and consume energy.¹ Yeast grown in continuous culture displays transcriptional oscillations that correlate with metabolite concentrations.^{2–9} The periods of the transcriptional oscillations can be influenced by growth conditions, such as glucose availability and the amount of dissolved oxygen (DO), and by drug treatment.^{10–14} Time-resolved microarray data have demonstrated that the levels of certain transcripts correlate with each other during different phases of the cell cycle.^{3,6,15} Taken together, these studies support a possible correlation between gene transcription and metabolic function,¹⁶ as well as a link between transcriptional dynamics and energy requirements and energy release. In the analysis described herein, we sought to make such a link explicit and quantitative.

The state of the yeast cell in a culture is constantly changing, for example, with varying oxygen levels in the culture medium.^{4,15} These cells are thought to self-synchronize through cell signaling of small molecules such as hydrogen sulfide and acetaldehyde.^{8,9} The transcription oscillations or redox cycles are related to the respiratory and reductive phases, and transcription and metabolites are precisely timed to peak at specific points along the redox cycle.¹⁵ These redox cycles can easily be monitored by DO levels in the fermentor. For example, as the level of DO in the fermentor decreases, cells consume oxygen in

the respiratory phase. Then, during the early reductive phase, the cells cease consuming oxygen and the level of DO in the fermentor increases. The late reductive phase corresponds to a high level of DO in the media because cells consume the least oxygen during this phase. The coordination of DO oscillations with timing of DNA replication, mRNA expression, and metabolites has been reported.^{3,6,7,14} Cells are thought to be less prone to genetic mutations when DNA replication occurs during the reductive phase of the cycle.^{3,7,14,15} Specifically, we used the data reported for two different strains in refs 3 and 15 to relate transcription levels to work requirements. The special feature of the experiments is that many time points were measured, and this allows us to identify cycles in energy requirements and energy consumption. See also the data analyses in refs 15, 17, and 18. We used three steps to elucidate the changes in the free energy: (i) relating the thermodynamic notion of free energy to the level of expression of the transcripts, (ii) computing the actual energy changes of the transcription system during the cell cycle by using experimental microarray

Received: December 24, 2012

Revised: January 27, 2013

Published: February 4, 2013

data, and (iii) offering a biological interpretation of the observed dominant energy requirements.

On the basis of the three steps listed above, one can proceed to discuss the response to stimuli, such as increasingly limiting nutrients, oxygen, etc. Quantitatively assessing the role of such perturbations is, however, beyond the scope of this paper. The necessary theoretical machinery for doing so, a quantitative version of the principle of Le Chatelier, is discussed in ref 19.

A general property of physicochemical systems is that they can do work only when they are not in equilibrium. We took the approach that this also holds true for biological systems. The free energy of the current state of a cell is the maximal work that can be done when the cell is brought from its current state to a state that is not constrained by external perturbations. This is the steady-state baseline,^{20,21} where the definition of this value is repeated in eq 1. We take this state to be a reference state, and as part of the analysis, we check if the state is common to all time points of the cell. We can also view the free energy as the minimal work needed to be done for the cell to be brought from the notional steady state to its current state by imposing the external constraints. The physical analogy is deriving work from the energy of water that is otherwise spontaneously falling down a waterfall. The maximal work available is the difference between the potential energy of the water at the top and at the bottom of the fall. It is not only the initial height of the water but also the final height that determines the work available from the fall.

A cell that is in the steady state and is not constrained cannot be used to extract any work, but what about the work needed to assemble the cell when it is in the steady state? This, often much larger, work is the standard free energy. In this study, we set out to compute both terms: (i) the standard free energy (the work required for assembling the cell in the steady state) and (ii) the excess free energy that represents the maximal available work from the current constrained state of the cell that also equals the work required to bring the cell from the steady state to its current state.

To proceed with the free energy analysis, we first must define the state of the cell. We here define only the state of the transcription system and not of the entire cell. We therefore emphasize that the free energy we computed is the thermodynamic free energy, but it is the free energy of the transcription system alone. We did not include other important cellular constituents such as metabolites. Surprisal analysis, which has been well documented^{20–22} for characterization of the expression level of transcripts, was used to characterize the state of the transcription system as it goes through its cycles. This analysis was by itself challenging because of the relatively large number, 48, of time points at which the transcription levels were measured. Moreover, the transcription levels oscillate, and surprisal analysis had to describe this nonmonotonic time dependence. In this paper, we take a significant step beyond earlier studies in that surprisal analysis is used not only to characterize the transcription levels but also to compute the free energy changes during the cellular cycles. This is possible because the surprisal-based analysis has the advantage that the two components of the free energy, the standard free energy and the maximal available work, are readily and directly computable from the output of the analysis. This advantage arises from the thermodynamic background of surprisal analysis²³ and specifically in that we use the thermodynamic and not the statistical definition of entropy. This means that there is a baseline value for each expression level, a level that reflects the thermodynamic weight in the absence of constraints.^{20,21} In particular, the route

from the surprisal to the maximal work that can be derived from a state that is not in equilibrium is explicitly discussed in ref 24. In this study we describe for the first time the use of this connection to determine the work that is being done in a biological process. Specifically, we compute the work needed to drive a cell to different states during its cycles. We then provide experimental evidence that supports the idea that the most work is being done during ribosomal synthesis.

■ EXPERIMENTAL AND THEORETICAL PROCEDURES

Surprisal analysis, the theoretical procedure that is the background to what we do, is discussed here with some technical points in section S1 of the Supporting Information. See refs 20 and 21 for more details in a biological context and refs 23 and 25 for applications in chemical physics. The errors in surprisal analysis that are due to experimental uncertainties are analyzed in section S2 of the Supporting Information. The essential and new procedural point about the theory used in this paper is that once surprisal analysis has been performed, computing the free energy is rather straightforward. This theme is developed as part of the discussion of the results. We also discuss here the materials and experimental methods that were used in the experiment that was performed to validate the theoretical analysis.

Input Microarray Data. We aim to describe the changes over time of genomewide transcription levels in *Saccharomyces cerevisiae* grown in continuous culture and to determine which genes contribute at each time point during the cell cycle. Specifically, the analysis sought to identify groups of genes that change with time in precisely the same manner. In previous applications of surprisal analysis, we have referred to such a group as a transcription pattern.^{20,21} As opposed to the commonly used clustering method, surprisal analysis identifies a dynamical transcription pattern, i.e., a group of transcripts that synchronically progress throughout the time evolution of the cell. Typically, only a few transcription patterns are needed to quantitatively characterize the dynamics of the transcript's expression. In the *S. cerevisiae* system, at most four transcription patterns were shown to be sufficient to capture all the important features in the measured microarray levels of each transcript. We used data for two strains. These had transcription patterns oscillating with a period of 40 min (IFO0233) so that the experiment^{3,4} spans several cycles and a strain (CEN.PK113-7D) with a period of 4 h where the experiment is reported in ref 15.

Surprisal Analysis. The central reason for the use of surprisal analysis is that it includes all the input required to compute the free energy changes and also the standard reference baseline value for the transcriptional state. This is because we use entropy as a thermodynamic quantity and not as a statistical measure of dispersion. A key advantage of surprisal analysis is that it also determines the steady-state or “baseline” value and the corresponding expression of the transcripts. This baseline is indeed a major contributor to the expression level,^{20,21} particularly so for the more strongly expressed transcripts. Deviations from the baseline are caused by constraints that are present on the system.

The procedure of maximizing the thermodynamic entropy subject to constraints can be implemented using the well-studied mathematical method of Lagrange multipliers. Using this method,^{20,21} we seek the maximum of the entropy at time point t and thereby derive the value $X_i(t)$ of the expression level of the i th transcript at time t as

$$\underbrace{X_i(t)}_{\substack{\text{expression level} \\ \text{of transcript } i \\ \text{at time } t}} = \underbrace{X_i^o}_{\substack{\text{base line} \\ \text{expression level} \\ \text{of transcript } i}} \cdot \exp\left(-\sum_{\alpha=1}^{\text{sum over the constraints, } \alpha=1,2,\dots} \underbrace{\lambda_{\alpha}(t)}_{\substack{\text{Lagrange multiplier} \\ \text{of constraint } \alpha \\ \text{at time } t}} \cdot \underbrace{G_{i\alpha}}_{\substack{\text{weight of} \\ \text{transcript } i \\ \text{in constraint } \alpha}}\right) \quad (1)$$

where X_i^o , the baseline expression level, is the expression level of transcript i in a state not subject to perturbations. When all the transcripts reach this baseline value, the entropy reaches its unconstrained maximal value for a given environment. The introduction of constraints lowers the maximal value of the entropy. The presence of constraints is indicated by the exponential correction term, which is a generalization of the more familiar Boltzmann factor for which the constraint is energy exchange with the surroundings.

Toward a Biological Interpretation of Surprisal Analysis. Given an expression level that changes between one time point and another, biochemists will look first at the fold change in that level. That is in essence what surprisal analysis does but with two key qualifications. To see that it is a fold view, we rewrite eq 1 in a strictly equivalent form

$$-\ln[X_i(t)/X_i^o] = -[\ln X_i(t) - \ln X_i^o] = \sum_{\alpha=1} \lambda_{\alpha}(t)G_{i\alpha}$$

What is different is the fact that we scale the actual fold value by the baseline fold value. A transcript can have a rather low expression value and yet be subject to important changes if its baseline value is also low. Indeed, we find in general that such transcripts are more susceptible to change.²⁰ The other point is that we resolve the change as a sum of contributions, each labeled by an index, denoted α in eq 1. The structure of each deviation term for transcript i , $\lambda_{\alpha}(t)G_{i\alpha}$ is also very telling. It is a separable expression meaning that the dependence on time and the dependence on index i of the transcript are uncoupled. The time dependence given by $\lambda_{\alpha}(t)$ is the same for all transcripts. $\lambda_{\alpha}(t)$ is the same for all transcripts in the deviation pattern α . The extent to which a transcript changes by pattern α is independent of time and is specified by the value of $G_{i\alpha}$. We particularly take note of those transcripts whose value of $G_{i\alpha}$ is either very positive or very negative. These are the transcripts that are either very upregulated or very downregulated in the pattern α . A transcript can vary substantially in one pattern and hardly vary in another pattern. The level of the housekeeping transcripts often hardly varies in any pattern.²⁰

Weight of Transcription Patterns. The thermodynamic importance of a constraint is determined by how much it reduces the value of the entropy at time point t . The importance is measured by the value of the Lagrange multiplier $\lambda_{\alpha}(t)$, and as can be seen in eq 1, less important constraints result in a smaller value of the exponent and hence do not greatly influence the level of expression. Outright unimportant constraints have Lagrange multipliers that equal zero.

Target Preparation and Processing. Purified RNA samples from each time point were adjusted to a final concentration of 1 $\mu\text{g}/\mu\text{L}$. The RNA samples were processed as described in the Affymetrix GeneChip Expression Analysis Technical Manual (see ref 4).

Western Blotting Sample Preparation. Fermentor culture conditions for CEN.PK113-7D oscillating with a 4 h period have been previously described.¹⁵ The cell concentration was determined using a Coulter counter (Beckman Coulter, Miami, FL). For each time point, a constant volume of cells (~ 5

$\times 10^7$) was removed from the fermentor, pelleted by centrifugation, resuspended in 0.05 mL of lysis solution, and boiled for 7 min. The lysis solution consisted of 50 mM Tris (pH 7.5), 5% SDS, 5 mM EDTA, 5% glycerol, 50 mM DTT, 0.01 mg/mL aprotinin, 0.1 mg/mL Pefabloc, 0.01 mg/mL leupeptin, 0.01 mg/mL pepstatin, 0.01 mg/mL chymostatin, 0.06 mg/mL antipain, 0.15 mg/mL benzamidine, and $\sim 0.08\%$ bromophenol blue. Lysed samples were returned to room temperature, and extracts were clarified by centrifugation (5 min at 16000 relative centrifugal force). The supernatant was stored at -80°C . For Western blot analysis, the supernatant was boiled for 5 min and mixed, and equal volumes were loaded into lanes of an 18% polyacrylamide gel (Criterion Cell, Bio-Rad Laboratories, Hercules, CA).

Western Blotting. Proteins were transferred to a 0.45 mm nitrocellulose membrane using the Criterion Blotter (Bio-Rad Laboratories). Primary antibodies against phosphorylated Rps6p (Cell Signaling Technology, Danvers, MA) were used at a 1:1000 dilution. IRDye 800CW goat anti-rabbit secondary antibodies (LI-COR Biosciences, Lincoln, NE) were used at a 1:10000 dilution. Blots were developed on an Odyssey Imager (LI-COR Biosciences), which yields an arbitrary density value. The density for each time point was divided by the average density of all samples from the same blot to yield the protein average centered ratio. To normalize the density of the Rps6p signal, the protein average centered ratio was divided by the total protein correction factor at an equivalent phase angle from the DO curve.

Time Adjusted to Phase Angle. To compare protein levels gathered on different days, DO levels and Western blot data were plotted together with phase angle rather than time. The inflection point from decreasing levels of dissolved oxygen represents a 0° phase angle. Each cycle represents 360° .

Protein Quantification. Because the lysis solution interfered with colorimetric assays (Bio-Rad Protein Assay and RC DC Protein Assay, Bio-Rad Laboratories) for measuring protein concentration, Coomassie blue staining of the polyacrylamide gel was used to semiquantify the total protein concentrations. Gels were stained with Bio-Safe Coomassie Stain (Bio-Rad Laboratories), and the protein density was measured with an Odyssey Infrared Imaging system (LI-COR Biosciences). The loaded extract volume was in the linear density range in the 700 nm channel. Data were scaled with a total protein correction factor for each phase angle that was determined by dividing the total protein density of each lane by the average total protein density from the same gel.

RESULTS AND DISCUSSION

Surprisal Analysis of Transcription Oscillations. Surprisal analysis uses eq 1 to quantitatively represent the time variation of transcription levels. The time evolution for each transcript i is determined by its Lagrange multipliers $\lambda_{\alpha}(t)$ that correspond to transcription pattern α where $\alpha = 1, 2, \dots$. The transcripts of each transcription pattern collectively evolve according to the time dependency of the constraint as given by its Lagrange multiplier.

The left-hand side of eq 1, $X_i(t)$, is the measured expression value of transcript i at time point t . Then, using a numerical procedure (discussed in the Supporting Information and in ref 21), surprisal analysis seeks to accurately fit the data by including as few constraints as possible. The values of many quantities in eq 1 are not known before we begin our analysis, but there were also many more measured values of the expression levels, e.g., 5280 transcripts measured at 48 different time points for the data reported in ref 4 than unknowns. Thus, there were a sufficient number of data points, and a careful count showed that we could, in principle, determine weights $\lambda_\alpha(t)$ of up to 47 different, independent constraints, meaning that α assumes all values from 1 to 47. In practice, the majority of the constraints were determined to be strictly marginal, meaning that their contribution was within the noise level of the data. Indeed, the first output of the analysis was the ranking of the constraints by their importance. For the data of ref 4, the importance of the constraints decreased exponentially with their values of index α (Figure S1 of the Supporting Information). Of the 47 possible constraints, at most 11 were outside the noise level that could be estimated from the data.²⁶ Of these 11, four constraints accounted for all critical features in the oscillation dynamics (Figure S2 of the Supporting Information) in the data for IFO 0233.^{3,4} For the CEN.PK113-7D data,¹⁵ even just two multipliers provide a realistic approximation.

The quantitative assessment of how many multipliers are warranted by the data is based on translating the possible margin of error in the data to error bounds on the value of the multipliers. See the error analysis section of the Supporting Information and ref 26. The principle of the criterion is easy to state. A Lagrange multiplier is not needed when the error in its value is comparable to the value. Why? Because when such is the case, zero is a possible value for the multiplier being within the error bar. Such a multiplier only serves to fit and describe the noise in the data and not the real data.

Baseline. The baseline expression values, X_i^0 , are the expression levels at the global maximum of the entropy, when there are only the steady, time-invariant, constraints. Therefore, it is often useful to use the notation $X_i^0 = \exp(-\lambda_0 G_{i\alpha})$. In this equation, the values of $G_{i\alpha}$ are determined as described in the Supporting Information, which subjects them to a criterion of normalization, so λ_0 sets the absolute scale of the $G_{i\alpha}$ values. In the experiments reported in refs 4 and 15, the environment was possibly changing. Therefore, the weight of the constraint that reflects a “steady” environment could change with time. The change that we found was limited, <5%, but unlike other examples of cell cultures, here λ_0 did change some with time, particularly so when the drug phenelzine was added at time point 12.⁴

Determining the Lagrange Multipliers. We determined the Lagrange multipliers $\lambda_\alpha(t)$ by determining the eigenvalues and eigenvectors of the covariance matrix of the surprisal (see the Supporting Information). The eigenvalues were arranged in decreasing order (Figure S1 of the Supporting Information) and thereby ordered the constraints by their importance. The results for the Lagrange multipliers for the data of ref 4 are shown in Figure 1.

The Lagrange multipliers shown in Figure 1 already provided a realistic time point by time point representation of the measured expression levels. To obtain an accurate description of period doubling, stronger constraints were also needed. With 11 constraints, the oscillations in the levels of every transcript were fully and accurately captured, as shown in Figure S2 of the

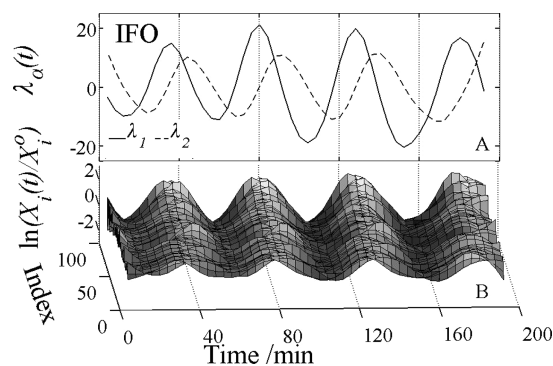


Figure 1. Results of surprisal analysis for strain IFO0233 for which the cycle period is ~ 40 min. The top panel shows the weight of the two most important transcription patterns vs time. Shown is the value of the Lagrange multiplier $\lambda_\alpha(t)$ vs time, where α is the index of the constraint ($\alpha = 1, 2, \dots$). Note how the value of the second multiplier is smaller than the first. The values of the higher multipliers are significantly lower. Note that the values of the Lagrange multipliers of the dynamical constraints can be negative. The error bars on these dominant multipliers are less than a few percent, computed as discussed in the Supporting Information. The bottom panel shows $\ln[X_i(t)/X_i^0]$, the expression levels, shown as a deviation from their time-independent baseline value, vs time for the 100 most heavily weighted transcripts for constraint $\alpha = 1$. This measure of the deviation is what we call the surprisal. The weight of constraint i in pattern α is, cf. eq 1, denoted as $G_{i\alpha}$. The value of this weight is determined by surprisal analysis as discussed in the Supporting Information and ref 18. The surprisals of those transcripts that are heavily weighted in the $\alpha = 2$ pattern (not shown) oscillate with a phase difference with respect to the oscillations seen in the bottom panel.

Supporting Information; 47 constraints provided a mathematically perfect fit. However, within the noise level, this fit could not be distinguished from the fit using 11 constraints (Figure S2 of the Supporting Information). On the time resolution of minutes, there are at most 11 transcription patterns that contributed to the process. Of these, four patterns dominate and the two main constraints ($\alpha = 1$ and 2) represent the respiratory and reductive phases.

A key result that is also found in all other transcription data²⁰ was the separation in scales between the value of the Lagrange multiplier of the baseline as compared to the constraints due to the dynamical process ($\lambda_0 \gg \lambda_1, \lambda_2, \dots$). We use the much greater than sign because λ_0 is larger than the others by more than an order of magnitude (Figure 1 and Figure S1 of the Supporting Information).

As shown in Figure 1, the values of the Lagrange multipliers of the dynamical constraints can be negative. Equation 1 shows that a change in sign corresponds to a qualitative switch in the role of the constraint. For one sign, transcripts are induced, and for the other sign, they are repressed. We have previously noted^{21,22} that such switches in the sign of the constraints reflect transitions in the dynamics.

From Surprisal Analysis to Dynamics. The time-dependent role of a constraint is completely described by the time dependence of its Lagrange multiplier, $\lambda_\alpha(t)$. All the transcripts that contribute significantly to a given pattern α will increase or decrease coherently with the same time dependence as shown in Figure 1.

Surprisal analysis therefore describes the deviations from a balanced steady state in terms of very few clusters of transcripts that we call transcription patterns. All transcripts that contribute to a particular transcription patterns do so in a coherent fashion.

Participation of Transcripts in Constraints. The extent to which the expression of transcript i responds to the imposition of constraint α is measured by $G_{i\alpha}$ (see eq 1). When we examined a particular constraint, we found that of the several thousand transcripts of *S. cerevisiae*, most did not participate in that constraint to any considerable extent. A few hundred constraints were strongly upregulated, meaning their value of $G_{i\alpha}$ was atypically negative. (A negative value moves the transcript up because of the negative sign in the exponent in eq 1. For more than a century, distributions of maximal entropy have been written with a minus sign in the exponent,²⁷ and we followed this practice.) For the same constraint, a few hundred transcripts were downregulated, meaning that their value of $G_{i\alpha}$ was atypically positive (see Figure S3 of the Supporting Information for $\alpha = 1$). When we sought to determine the average participation in constraint α by averaging the value of constraint $G_{i\alpha}$ over all the transcripts [$\langle G_\alpha \rangle = \sum_i X_i(t) G_{i\alpha}$], the resulting number was not large because of the opposing contributions from transcripts with positive and negative values of $G_{i\alpha}$. Overexpressed transcripts, those with negative $G_{i\alpha}$ values, had somewhat higher levels, X_i , so the sum was not quite zero; however, without a special reason, the mean participation value $\langle G_\alpha \rangle$ will not stand out (Figure S4 of the Supporting Information).

A closely related issue is the participation of transcripts in the baseline. From the equation $X_i^o = \exp(-\lambda_0 G_{i0})$, those transcripts that are strongly expressed must have a very negative value of G_{i0} . It follows that the transcripts that are very stable have a particularly large negative value of G_{i0} . Thus, the mean participation value in the baseline [$\langle G_0 \rangle = \sum_i X_i(t) G_{i0}$] must be negative (Figure S4 of the Supporting Information). The more negative this value, the more energetically stable the baseline distribution of transcripts will be.

These features of transcript participation can also be viewed by plotting a histogram of the values of $G_{i\alpha}$.¹⁹ The histogram is bell-shaped and centered about zero for constraints $\alpha = 1, 2, \dots$, and also bell-shaped but centered at some negative value for the baseline ($\alpha = 0$). This was found for almost all constraints that we identified in the data of either strain. However, there were exceptions as shown in Figure 2 for the dominant constraint that characterizes the respiratory phase.

Exceptionally Overexpressed Transcripts. The histograms for the respiratory pattern have a somewhat fat tail at high negative values [Figure 2; $G_{i1} < -0.02$ IFO (panel A), and $G_{i1} < -0.025$ CEN.PK (panel B)], comprising 299 and 285 genes, respectively (Table S1 of the Supporting Information lists 285

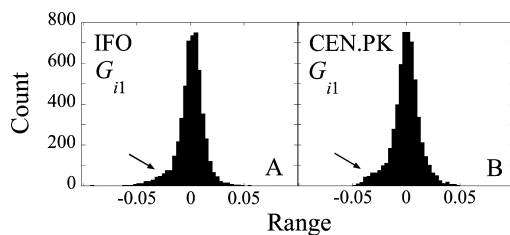


Figure 2. Histograms of weights of different transcripts in the main expression pattern, respiratory, of the two strains of *S. cerevisiae* as indicated. Both histograms show a shoulder, indicated by an arrow, as compared to a distribution symmetric about zero,¹⁹ at large negative values. One could say that the distributions are “fat-tailed”. The data for IFO0233 were taken from refs 3 and 4 and span several cycles. The data for CEN.PK113-7D were reported in ref 15 and have a period of 4 h.

for CEN.PK). To identify the genes associated with the shoulders, we used the AmiGo gene ontology database.²⁸ Most of the genes identified were associated with metabolic processes, but as we further discuss below, 98 genes (a third of the 299 genes for IFO) and 142 (half for CEN.PK) were ribosomal transcripts (GO terms 0005840 and 0042254) and were localized in the shoulder.

The mean participation value of pattern α , $\langle G_\alpha \rangle$, depends on the values of $G_{i\alpha}$. The mean values of the transcript patterns with symmetric histograms of the $G_{i\alpha}$ values are expected to oscillate about zero (Figure S4 of the Supporting Information). However, when the distribution of the $G_{i\alpha}$ values has a shoulder in the direction of negative values, the mean participation value was more negative than the rest. Figure S4 of the Supporting Information shows results of the mean value versus time for transcripts $\alpha = 0, 1$, and 2.

■ COMPUTING FREE ENERGY CHANGES

Work Done by Transcription Dynamics. The work available from the nonequilibrium transcription levels or the work done on the transcription system to drive it away from equilibrium can immediately be computed from the output of surprisal analysis. The energy available to the transcription system to do work at time t is computed as a sum over all transcripts as the mean of the surprisal of the transcription level value.²⁴ The work is evaluated by using eq 1 to compute the surprisal $-\ln[X_i(t)/X_i^o]$

$$\begin{aligned}
 F(t) &= \sum_i X_i(t) \ln[X_i(t)/X_i^o] \\
 &= \sum_i X_i(t) \left[- \sum_{\alpha=1} \lambda_\alpha(t) G_{i\alpha} \right] \\
 &= - \sum_{\alpha=1} \lambda_\alpha(t) \left[\sum_i X_i(t) G_{i\alpha} \right] \\
 &= - \sum_{\alpha=1} \lambda_\alpha(t) \langle G_\alpha \rangle(t)
 \end{aligned}
 \tag{2}$$

All quantities that are needed to compute eq 2 are provided by the surprisal analysis of the data. To make it clear that both contributions in eq 2 depend on time, we write the mean participation value as $\langle G_\alpha \rangle(t)$. The total available work from the transcription system is a sum of contributions of all the transcript patterns. It is important to note that it is only the deviation from the baseline that contributes to the work available. When the transcripts are at a steady balance, their expression level is X_i^o . As is expected and shown by eq 2, in such a case no work is available [$\ln(1) = 0$].

For a given transcription pattern, the greater the product $\lambda_\alpha(t) \langle G_\alpha \rangle(t)$, the more work is available. So, as a simple rule of thumb, because the weights $\lambda_\alpha(t)$ of the patterns decrease with an increasing value of α , the more dominant constraints, those with a low value of α , do most of the work. The work done by the respiratory process, $F_1(t)$, is the dominant contribution as shown in Figure 3. By comparison, the reductive process, $F_2(t)$, required significantly less work.

Free Energy Is Measured in Units of Thermal Energy. Equation 2 computes the free energy as a pure number (meaning in dimensionless units), but the thermodynamic free energy has the dimensions of energy (or of work). To convert eq 2 to these units, the result must be multiplied by $k_B T$, where k_B is Boltzmann’s constant and T is the temperature of the

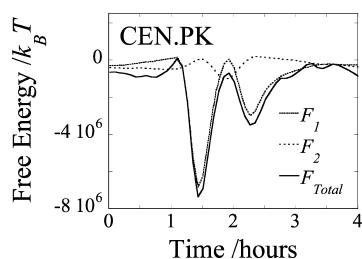


Figure 3. Contribution of the respiratory and reductive transcription patterns ($\alpha = 1$ and 2 , respectively) to the work available vs time. Work is shown in units of thermal energy ($k_B T$). The total over all patterns (see eq 2), $F(t)$, is shown as a solid line. Note the high contribution by the respiratory pattern. The plot is for the data of the CEN.PK strain,¹⁵ because the validation experiment was conducted for this strain.

environment. As also seen in eq 2, the free energy is directly proportional to the extent of the expression level of the transcripts. The greater the level of expression of the transcripts, the more work that is available and/or needed. In the data we use, the values reported by microarray analysis are such that the concentration of RNA from the different samples was adjusted to $1 \mu\text{g}/\mu\text{L}$ (see Experimental and Theoretical Procedures).

Work Done by the Transcripts. The total work done by the transcription patterns in CEN.PK and IFO is shown in Figures 3 and 4, respectively. Also indicated in Figure 4 are the different

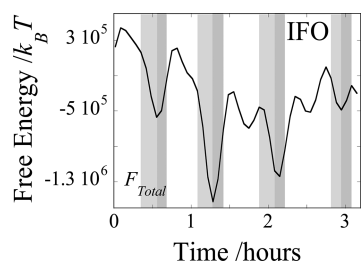


Figure 4. Total work done by the transcription patterns vs time in several cycles from measured microarray data for the IFO strain.⁴ The color shading shows respiratory (light gray), early reductive (dark gray), and late reductive (white) phases.

phases. This shows that most of the work is done during the respiratory stage and is released during the early reductive phase. The list of transcripts implicated in the work being done is provided in Table S1 of the Supporting Information.

Standard Free Energy. The standard free energy of the transcription pattern is defined as in chemical thermodynamics as the change in free energy between the steady-state baseline values and a reference state where all chemical elements are in their standard state. It is computed from the results of surprisal analysis as $F_0(t) = -\sum_i X_i \ln X_i^0 = \lambda_0 \langle G_0 \rangle$. The standard free energy of the transcription pattern should be a negative number because the individual entries G_{i0} were all negative, $\langle G_0 \rangle$ was negative (Figure S4 of the Supporting Information). λ_0 was positive and significantly larger than all the Lagrange multipliers (λ_α , where $\alpha = 1, 2, \dots$) of the constraints shown in Figure 1. The standard free energy of the transcription pattern was therefore very negative and also far larger in absolute value than the free energy changes caused by the transcription dynamics. The baseline reference remained stable (negative) throughout the dynamics. The standard free energy could be somewhat perturbed but could not be made positive, meaning unstable. These conclusions are

shown in a graphical manner in Figure S5 of the Supporting Information.

■ VALIDATION EXPERIMENT

On the basis of the analysis, we hypothesized that cells in the respiratory phase likely build ribosomes, an energy intense process, in preparation for protein production needed for cell growth during the S phase of the cell cycle. Therefore, we performed a validation experiment showing that high levels of a small (40S) subunit of ribosomal protein, Rps6p, are made during the respiratory phase and degrade during the early reductive phase (Figure 5). Relative quantitation was based on

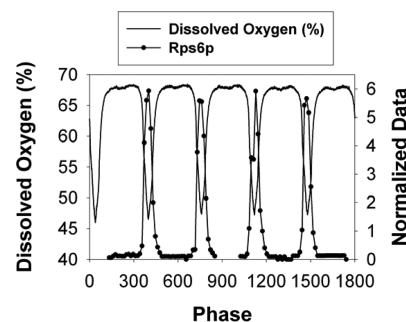


Figure 5. Rps6 protein levels relative to the level of DO in the media. Solid lines represent dissolved oxygen levels in the fermentor. Circles represent normalized Rps6 protein levels normalized against total protein (right axis). Measured for several cycles of the CEN.PK113-7D strain. The relation between time and phase angle is discussed in Experimental and Theoretical Procedures.

Western blot analysis (Figure S6 of the Supporting Information) and normalized to the total protein concentration (Figure S7 of the Supporting Information). We provide new experimental evidence that building of ribosomes, an energy intensive process, is in preparation for protein production needed for cell growth during the S phase of the cell cycle.

Energy Requirements during Ribosomal Building. Because cells collectively take up DO in the media, the respiratory phase is when mitochondrial respiration is active. Electron microscopy studies show the energized conformation of the mitochondria during the respiratory phase and a resting conformation at the reductive phase.²⁹ Previous reports showing oscillations in ATP concentration^{30,31} and ATP/ADP concentration ratio³² reflect the cell's deterministic organization of energy production with energy intensive processes. One possible process is ATP-dependent nucleosome remodeling.³² Our experiments support ribosomal assembly at the respiratory phase, a highly energy demanding process,^{32–34} that coordinates with energy production.

■ CONCLUSION

The cell populations in the yeast continuous culture system display energetic oscillations. These oscillations allow efficient coordination with the cell cycle and cellular growth. Surprisal analysis can be used to resolve the cyclic variations in the levels of transcripts into the contributions from different processes. For each process, there is a group of transcripts that respond collectively. We identified dominant groups, which corresponded to the previously identified respiratory, early, and late reductive phases.

The free energy change during each cycle could be readily computed using as input the results of the thermodynamic-like

surprisal analysis. The most energy intensive process was the biogenesis of ribosomes, very likely in preparation for protein production during the synthesis phase. This reflects the cell's deterministic coordination of energy production with energy intensive processes.

In the longer run, the results reported in this study suggest that surprisal analysis can be used to compute the temporal changes in the free energy of an entire biological process, e.g., transcription. In physicochemical systems, the free energy gradient determines the direction of change. It remains to be shown that this can be used to advantage also in a system-biological context.

■ ASSOCIATED CONTENT

📄 Supporting Information

Details of surprisal analysis (section S1), error analysis (section S2), eigenvalues of the covariance matrix (Figure S1), point-by-point fit of the surprisal (Figure S2), sorted contributions of transcripts to the dominant constraint (Figure S3), mean participation value of the transcripts (Figure S4), free energy of the transcription pattern (Figure S5), Western blot probed with an antibody recognizing Rps6p (Figure S6), protein levels relative to the level of DO in the media (Figure S7), and a list of transcripts comprising the tail of the dominant (respiratory) pattern in the CEN.PK data as indicated in Figure 2 (Table S1). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Department of Chemistry and Biochemistry, Crump Institute for Molecular Imaging, and Department of Molecular and Medical Pharmacology, University of California, Los Angeles, CA 90095. Phone: (310) 206-0476. E-mail: rafi@chem.ucla.edu.

Author Contributions

A.G. and C.M.L. contributed equally to this work.

Funding

C.M.L. was supported by Grant R01 GM81757 from the National Institute of General Medical Sciences. F.R. is director of Fonds National de la Recherches Scientifique, Belgium. We acknowledge the support of the Prostate cancer Foundation via a Creativity Award.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Lily Nan for careful data collection and quantification of Western blot and Coomassie blue staining results and Shwe Chin for phase analysis determinations and compilation of total protein scaled data. We thank Drs. Sohila Zadrán, Keely Walker, and Ian Blaby for their comments on the manuscript.

■ REFERENCES

- (1) Atkinson, D. E. (1977) *Cellular Energy Metabolism and its Regulation*, Academic Press, New York.
- (2) Lloyd, D., Murray, D. B., Klevecz, R. R., Wolf, J., and Kuriyama, H. (2008) *The Ultradian Clock (~40 min) in Yeast (Saccharomyces cerevisiae)*, Springer, Berlin.
- (3) Klevecz, R. R., Bolen, J., Forrest, G., and Murray, D. B. (2004) A genomewide oscillation in transcription gates DNA replication and cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1200–1205.
- (4) Li, C. M., and Klevecz, R. R. (2006) A rapid genome-scale response of the transcriptional oscillator to perturbation reveals a period-doubling

path to phenotypic change. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16254–16259.

- (5) Murray, D. B., Beckmann, M., and Kitano, H. (2007) Regulation of yeast oscillatory dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2241–2246.
- (6) Tu, B. P., Kudlicki, A., Rowicka, M., and McKnight, S. L. (2005) Logic of the yeast metabolic cycle: Temporal compartmentalization of cellular processes. *Science* 310, 1152–1158.
- (7) Tu, B. P., Mohler, R. E., Liu, J. C., Dombek, K. M., Young, E. T., Synovec, R. E., and McKnight, S. L. (2007) Cyclic changes in metabolic state during the life of a yeast cell. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16886–16891.
- (8) Murray, D. B., Klevecz, R. R., and Lloyd, D. (2003) Generation and maintenance of synchrony in *Saccharomyces cerevisiae* continuous culture. *Exp. Cell Res.* 287, 10–15.
- (9) Sohn, H. Y., Murray, D. B., and Kuriyama, H. (2000) Ultradian oscillation of *Saccharomyces cerevisiae* during aerobic continuous culture: Hydrogen sulphide mediates population synchrony. *Yeast* 16, 1185–1190.
- (10) Porro, D., Martegani, E., Ranzi, B. M., and Alberghina, L. (1988) Oscillations in continuous cultures of budding yeast: A segregated parameter analysis. *Biotechnol. Bioeng.* 32, 411–417.
- (11) Adams, C. A., Kuriyama, H., Lloyd, D., and Murray, D. B. (2003) The Gts1 protein stabilizes the autonomous oscillator in yeast. *Yeast* 20, 463–470.
- (12) Salgado, E., Murray, D. B., and Lloyd, D. (2002) Some Antidepressant Agents (Li, Monoamine Oxidase Type A Inhibitors) Perturb the Ultradian Clock in *Saccharomyces cerevisiae*. *Biol. Rhythm Res.* 33, 351–361.
- (13) Xu, Z., and Tsurugi, K. (2007) Destabilization of energy-metabolism oscillation in the absence of trehalose synthesis in the chemostat culture of yeast. *Arch. Biochem. Biophys.* 464, 350–358.
- (14) Chen, Z., Odstroil, E. A., Tu, B. P., and McKnight, S. L. (2007) Restriction of DNA replication to the reductive phase of the metabolic cycle protects genome integrity. *Science* 316, 1916–1919.
- (15) Chin, S. L., Marcus, I. M., Klevecz, R. R., and Li, C. M. (2012) Dynamics of oscillatory phenotypes in *Saccharomyces cerevisiae* reveal a network of genome-wide transcriptional oscillators. *FEBS J.* 279, 1119–1130.
- (16) Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* 9, 3273–3297.
- (17) Alter, O. (2006) Discovery of principles of nature from mathematical modeling of DNA microarray data. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16063–16064.
- (18) Alter, O. (2007) Genomic Signal Processing: From Matrix Algebra to Genetic Networks. In *Microarray Data Analysis: Methods and Applications* (Korenberg, M. J., Ed.) Humana Press, Totowa, NJ.
- (19) Shin, Y. S., Remacle, F., Fan, R., Hwang, K., Wei, W., Ahmad, H., Levine, R. D., and Heath, J. R. (2011) Protein Signaling Networks from Single Cell Fluctuations and Information Theory Profiling. *Biophys. J.* 100, 2378–2386.
- (20) Kravchenko-Balasha, N., Levitzki, A., Goldstein, A., Rotter, V., Gross, A., Remacle, F., and Levine, R. D. (2012) On a fundamental structure of gene networks in living cells. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4702–4707.
- (21) Remacle, F., Kravchenko-Balasha, N., Levitzki, A., and Levine, R. D. (2010) Information-theoretic analysis of phenotype changes in early stages of carcinogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10324–10329.
- (22) Kravchenko-Balasha, N., Remacle, F., Gross, A., Rotter, V., Levitzki, A., and Levine, R. D. (2011) Convergence of logic of cellular regulation in different premalignant cells by an information theoretic approach. *BMC Syst. Biol.* 5, 42.
- (23) Levine, R. D., and Bernstein, R. B. (1974) Energy disposal and energy consumption in elementary chemical reactions. Information theoretic approach. *Acc. Chem. Res.* 7, 393–400.

(24) Procaccia, I., and Levine, R. D. (1976) Potential work: A statistical-mechanical approach for systems in disequilibrium. *J. Chem. Phys.* 65, 3357–3364.

(25) Levine, R. D. (2005) *Molecular Reaction Dynamics*, The University Press, Cambridge, U.K.

(26) Kinsey, J. L., and Levine, R. D. (1979) Performance Criterion for Information Theoretic Data-Analysis. *Chem. Phys. Lett.* 65, 413–416.

(27) Gibbs, J. W. (1902) *Elementary Principles in Statistical Mechanics*, Charles Scribners & Sons, New York.

(28) Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., and Lewis, S. (2009) AmiGO: Online access to ontology and annotation data. In *AmiGO Hub*, pp 288–289, Bioinformatics., Web Presence Working Group (January 2009).

(29) Lloyd, D., Salgado, L. E., Turner, M. P., Suller, M. T., and Murray, D. (2002) Cycles of mitochondrial energization driven by the ultradian clock in a continuous culture of *Saccharomyces cerevisiae*. *Microbiology* 148, 3715–3724.

(30) Satroutdinov, A. D., Kuriyama, H., and Kobayashi, H. (1992) Oscillatory metabolism of *Saccharomyces cerevisiae* in continuous culture. *FEMS Microbiol. Lett.* 77, 261–267.

(31) Xu, Z., Yaguchi, S., and Tsurugi, K. (2004) Gts1p stabilizes oscillations in energy metabolism by activating the transcription of TPS1 encoding trehalose-6-phosphate synthase 1 in the yeast *Saccharomyces cerevisiae*. *Biochem. J.* 383, 171–178.

(32) Strunk, B. S., and Karbstein, K. (2009) Powering through ribosome assembly. *RNA* 15, 2083–2104.

(33) Kressler, D., Hurt, E., and Bassler, J. (2010) Driving ribosome assembly. *Biochim. Biophys. Acta* 1803, 673–683.

(34) Ju, Q. D., and Warner, J. R. (1994) Ribosome Synthesis during the Growth Cycle of *Saccharomyces cerevisiae*. *Yeast* 10, 151–157.