Tuning in the transcriptome: basins of attraction in the yeast cell cycle

R. R. Klevecz* and H. B. Dowse†

**Dynamics Group, Department of Biology, Beckman Research Institute of the City of Hope, Duarte, CA, USA, †Departments of Biological Sciences and Mathematics and Statistics, University of Maine, Orono, ME, USA*

(Received 8 November 1999; revision accepted 20 January 2000)

Abstract. Image processing techniques and wavelet analyses have been applied to the yeast cell cycle expression microchip data to reveal large-scale temporally coherent structures and high frequency oscillations in mRNA levels through the cycle. Because transitions in expression frequently occur in phase, they appear as peaks or troughs in colour maps and contour plots of expression levels. Although apparent in the untreated data, these transitions were identified and enhanced by convolution of a Laplacian kernel with the expression arrays of the first 4096 genes. Transitions associated with maximum up- or down-regulation of mRNA levels appear as bands at 30–40 min intervals through two cell cycles. Time-frequency analyses using wavelet transforms support these visualization techniques and lead to the conclusion that, with respect to gene expression, the dominant period is not the cell cycle (90–120 min) but, more commonly, the higher frequency 30–40 minute submultiple of the cycle period.

INTRODUCTION

Expression microarray technologies promise to resolve an issue that arose experimentally more than three decades ago when oscillations in enzyme levels and changes in the rates of RNA synthesis were first observed in synchronized mammalian cells (Klevecz & Stubblefield 1967; Klevecz & Ruddle 1968; Klevecz, Kauffman & Shymko 1984). Later, computer simulations and indirect studies involving manipulations of the cell cycle phase were interpreted as evidence for control of the cycle by a noisily periodic or chaotic attractor (Shymko, Klevecz & Kauffman 1984; Klevecz & Shymko 1985). In the absence of an ability to measure changes through the cell cycle for the entire transcriptome, molecular genetic approaches emphasized those genes showing high amplitude cycle-stage specific expression maxima, i.e. the cyclins, and sought to genetically dissect the cycle based on the idea that a relatively few genes were necessary and sufficient for cell cycle progression. Central to this reductionist research tradition is the isolation of mutants that appear to block cells at specific points in the division cycle, coupled with attempts to resolve these down-stream events into simple causal sequences (Hunt 1991; Paulovich, Toczyski & Hartwell 1997).

Correspondence: Robert R. Klevecz, Dynamics Group, Department of Biology, Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte, CA 91010, USA. E-mail: neumann@earthlink.net

Now, it is possible to compare, in a single assay, changes in gene expression among all the genes of a cell and thus directly examine the evidence for a large scale temporal order in the cell cycle. If coupling between genes is dense, then the trajectories followed by message concentration may collapse into a relatively few basins of attraction and these might be evident in the global patterns of message synthesis and degradation through the cell cycle. If the cell cycle is an attractor, it can be intentionally tuned or inadvertently perturbed and, in either case, gene expression would be expected to follow a regular, albeit sometimes noisy, trajectory away from the steady state. Therefore, the entire system, and not just the 'cell cycle regulated' genes, should oscillate. Based on the development of techniques for reproducible measurement of gene expression (De Risi, Iyer & Brown 1997; Velculescu *et al.* 1997), as well as commercially available microarray systems, Cho *et al.* (1998) and Spellman *et al.* (1998) reported on the changes in expression levels of nearly all 6178 gene through two cell cycles in yeast cells synchronized by several methods. The generous provision of this data on the Stanford University Web Site ([HTTP://genome-www.stanford.edu/cellcycle/](http://genome-www.stanford.edu/cellcycle/)) permitted the application of image processing techniques and modern methods of signal analysis to this important work. Two independent strategies were developed to examine the yeast database. One treated the entire 6178 gene array as a grey-scale or colour mapped image and sought to find evidence of regular patterns using image processing algorithms. The second treated the arrays as a time-frequency signal and examined the data using wavelet transforms, and wavelet decomposition signal processing techniques. These two very different approaches yielded the same conclusion; most genes of the yeast show regular periodic fluctuations in expression through the cell cycle with a 30–40 min period that is an integral submultiple of the cell cycle.

METHODS

Data sets from the four synchrony experiments reported on by Cho *et al.* (1998) and Spellman *et al.* (1998), made available on the Stanford University Web Site ([HTTP://genome-www.stanford.edu/cellcycle/](http://genome-www.stanford.edu/cellcycle/)), were downloaded as tab delimited text files and saved in Excel format. These files were opened and treated as matrices in Mathcad 8 Professional (Mathsoft, Inc; Cambridge, MA, USA). Cell cycle duration in the experiments varied between 90 and 120 min.

Edge detection

Laplacian edge detection is based on the equation of Laplace and finds regions in a matrix *M* where the second derivative is zero. The difference between the central pixel in the kernel and each pixel or matrix element of the matrix *M* is taken and then averaged. The Laplacian kernel as well as detailed sample calculations are available at [HTTP://www.talandic.com](http://www.talandic.com). As controls, files of identical shape were constructed with random numbers and analysed using the same edge detection methods. In addition the elu experiment served as an internal control. Laplace 4 returns the convolution of matrix *M* with a 3×3 Laplacian kernel. The kernel's centre is 4. In the colour mapped contour plot shown in Fig. 2, the array was scaled and equalized prior to the application of the kernel. Equalize returns matrix *M* with elements adjusted to form a linear cumulative histogram. Scale returns matrix *M* with elements scaled between 0 and 255.

Wavelet analysis

Prior to beginning the time-frequency analysis, wavelet approximations were performed on the time series data from a number of genes to determine how well each wavelet type approximated

the data. To scale the analysis multi-resolution analysis was performed on the same series of single genes and the results from the *cdc28* and alpha factor experiments compared. In this work, time frequency analysis and multi-resolution decomposition was performed using Daubechies wavelet. Comparisons were made between these results and those obtained using the cardinal spline or Battle–Lemarie and Coifflet wavelet. The results were also compared for analysis using Mathcad and the same wavelets analysis using in-house software. Results were consistent with a dominant 30–40 min period in both applications. Detailed calculations can be found at [HTTP://www.talandic.com/pubs/recent.html](http://www.talandic.com/pubs/recent.html).

Controls

To insure that there was no artifactual alignment of phase in the genome wide array analysis, the data from a single gene was duplicated and phase shifted by one sample interval in each duplication prior to wavelet multi-resolution decomposition. As would be expected, the analysis showed a diagonal pattern of peaks and troughs through time, but no coherence or synchronization in time. Several 6178 by 16 arrays of random numbers having the same range and standard deviation as the sample data were generated and analyzed. The contour and surface mapping presentations showed apparently random patterns. To test the possibility that some peculiar and unanticipated aspect of the data was leading to the pattern seen, several wavelet analyses were performed along the length of the chromosomes rather than through time. While a number of clusters of maximum or minimum expression were noted in this analysis, none of the clusters repeated at regular intervals.

RESULTS AND DISCUSSION

The original studies, as well as more recent re-analysis (Tayamo *et al.* 1999) examined closely those genes whose expression levels changed most dramatically through the cell cycle, and further restricted their analysis of ‘cell cycle regulated’ genes to those that showed only a single burst in expression in each cycle (Cho *et al.* 1998; Spellman *et al.* 1998). In part, this selection reflects a signal-to-noise limitation and the uncertainty that expression changes less than twofold reflect more than noise. Image processing and feature extraction techniques can sharpen the appearance of small clusters of elements whose individual behaviour is noisy and re-represent them to reveal large-scale patterns. Wavelet filtering can detect periodicities in the time series that would be overlooked by Fourier analysis. If the expression levels in many genes change at the same point in time, they may be visible even though the changes in any single gene are too low to be resolved statistically.

Wave detection and contour plots

If the entire data set of 6178 genes is represented as an array or matrix, it is possible to treat the resulting grid as an image directly or to remap the expression levels as a black and white image scaled from 0 to 255. The large scale order is apparent in the raw data when the synchrony experiments are presented as matrices to which a contouring algorithm has been applied (Fig. 1). Due to the high density of data, 6178 lines, single maxima in expression, out of phase with the majority of genes, are emphasized by this plotting routine. A somewhat more distinct picture is gained by the application of an edge detection algorithm. Laplacian kernels, which are used routinely in image processing applications as edge detection or edge enhancement algorithms, were used here to find inflection points in the image where the second derivative of the grey scale value is zero (Fig. 2). It is worth noting that such an algorithm will find downshifts as well as upward inflections in expression.

Figure 1. Contour plots of gene expression through the yeast cell cycle. Matrices were created for each of the arrays that resulted from the four synchrony experiments. The 6178 gene matrices were transposed so that the genes were aligned as rows from bottom (YAL00c) to top (YPR 204w) and a series of contour plots of the experiments: (a) *cdc15*; (b) *alpha* factor; (c) *cdc28*; (d) elutriation; are shown. Dark bands represent peaks or troughs in message levels through the cycle.

Data sets from all four of the synchrony methods described in the work of Cho *et al.* (1998) and Spellman *et al.* (1998), were analysed using edge detection and contour plotting routines. A contour plot of the data from all four synchronization procedures are shown in Fig. 1. The dark bands in these figures represent times through the experiment when

Figure 2. Basins of attraction in the yeast cell cycle. Colour mapped contour plot of transitions in mRNA levels through the yeast cell cycle for a field of 16 genes in yeast cells synchronized by alpha factor manipulation. Transitions in mRNA levels occur during restricted periods in the cell cycle. These transition points were enhanced by the use of an edge detection algorithm as described in the methods. The 16 genes shown here were assayed at 16 seven minute intervals through the cycle and then convolved with a Laplacian kernel.

expression levels are at a maximum or minimum. Rows represent the individual genes while columns represent sampling times through two cell cycles. In three of the four synchrony methods, there is evidence for peaks or troughs in expression at 30–40 min intervals. The high frequency components are most evident in the *cdc15* (Fig. 1a) and alpha factor (Fig. 1b) experiments. In Fig. 1(b), the alpha factor synchrony, the sampling frequency was every 7 min and this may have aided the resolution of these bands. In the *cdc28* experiment, high frequency bands are evident, but are not so constant in time as in Fig. 1(a, b). The exception to this high frequency banding is seen in the elutriation synchrony experiment where samples were taken at 30 min intervals and thus could not be expected to permit resolution of a 30 minute period.

One might have expected that peaks or troughs in expression would be distributed more or less uniformly throughout the cell cycle, reflecting the elaboration of a sequential genetic program that would act independently of collective temporal coordination, but this does not appear to be the case. Rather, in Fig. 1(a, b, c), expression highs and lows are restricted in time and occur at 30–40 minute intervals through the cycle. Broad dark bands in the *cdc28* experiment occur at 90 min intervals in the early G1 in two cycles, but are not so evident in the *cdc15* experiment. The appealing aspect of treating the array as a surface or grey scale image is the ease of application and the immediate visual impression of multiple expression peaks per cell cycle for a large fraction of the genome. More rigorous and quantitative support for this impression was sought through the application of signal processing techniques.

Wavelet analysis

Wavelet functions and wavelet analysis represent a growing class of signal processing techniques and transforms that use wavelets to efficiently measure and decompose a complex waveform into a set of detailed waveforms so that the component events can be localized in time and space and each component studied with a resolution matched to its scale. Their self-similarity and scalability make wavelets among the best new tools for discovering recurrent spatial or temporal patterns in noisily periodic or chaotic data. They are particularly suited to data sets that are short and contain both high and low amplitude signals. Such analyses may find features that are overlooked by Fourier time-frequency analysis (Alt *et al.* 1998).

Matrices derived from the four data sets, *cdc28*, *cdc15*, *elu* and *alpha*, were transposed and subjected to analysis by discrete wavelet transforms (Daubechies 1992). Shown here are the results of the discrete wavelet transform (dwt) using the Daubechies wavelet filter. Gene expression was analysed sequentially through 16 time points for each of the 6178 genes. This analysis decomposed the signal into a series of bands of decreasing frequency in which gene expression power appeared as a function of time. These data are plotted here with the genes arrayed as before, on the *y*-axis, with time-period along the *x*-axis and with spectral density in the *z*-axis appearing as a pattern of dark bands. In the *cdc28* and *alpha* factor experiments (Fig. 3a, b) the maximum spectral density appears at a period of 30–40 min and a second region of approximately 50–60 min, in good agreement with the results shown in Fig. 1. Additionally, there is a band in a smaller fraction of the genes with maximal expression at periods of 90–110 min. These would presumably represent the ‘cell cycle regulated’ genes previously identified. The same analysis was performed for the *cdc15* experiment, but the data were represented as a simple Cartesian plot in which the mean of the absolute values of the array is shown. Only the approximate 30-min period and the less clearly represented cell cycle period are evident (Fig. 4).

The experiment in which the cells were synchronized by elutriation had samples taken every 30 min, or one-third to one-fourth the sampling rate of the other experiments (see *elu* panel in Fig. 3). One would expect that spectral analysis of this experiment should fail to find significant power at 30 min simply because the sampling interval is below the Nyquist limit for detection of this high frequency oscillation. This expectation is borne out by the results of Fig. 3c. More important is the finding in this figure that little power is found anywhere on the interval analysed. This may suggest that with respect to message synthesis the yeast cell cycle does not exist as a period independent of the high frequency subcycles. Alternatively sparse sampling may have missed the mRNA synthesis event associated with the early G1 period of the cell cycle, or synchronization may not have been sufficiently tight by this method to permit spectral analysis. The requirement that the vector length be a power of 2

Figure 3. Wavelet transforms and filtering of expression profiles of the yeast cell cycle. Matrices derived from the three data sets, *cdc28*, *alpha*, and *elu* were subjected to analysis by the discrete wavelet transform (dwt) using the Daubechies wavelet filter as described. Gene expression was analysed sequentially for each of 4096 genes and the resulting transforms plotted to show maximum spectral density or power at integral intervals of the sampling interval (10 min in *cdc15* and *cdc28* and 7 min in *alpha* factor; dark bands in 3a–b). Note that the elutriation experiment *elu*, which was sampled at 30 min intervals, does not show evidence of high frequency oscillations. This sampling frequency difference serves as an internal control for the wavelet analysis.

Figure 4. Periodicity in message levels through the yeast cell cycle. Matrices derived from the *cdc15* data set were subjected to analysis by the discrete wavelet transform (dwt) using the Daubechies wavelet filter as described. The filter window for which maximum power was found was determined for each of the first 4096 genes and these were summed to yield a histogram of the frequency of maximum power. The dominant period in this synchrony appears to be in the 30–40 minute range.

further limits this conclusion. As always, this study would have benefited from more dense sampling over a greater number of cell cycles.

By these two very different methods of analysis, the regulatory loop that times gene expression has a period of 30–40 min. Moreover, there appear to be distinct periods within the cycle when relatively greater numbers of genes are expressed.

If some genes display high amplitude oscillations in expression there will be a tendency, either by inspection or by application of conventional signal processing techniques to find these genes and to regard other changes as noise. Fourier transforms have long been regarded as inadequate to the task of resolving periodicity in short under-sampled data sets. Our application of fast Fourier transforms (FFTs) and complex transforms CFFTs to the data confirm this expectation.

If many genes show maximal expression with the same oscillatory periods, it would seem that a greater emphasis on the dynamic aspects of the cell cycle is required. As has been discussed in detail previously (Klevecz, Bolen & Durán 1992; Bolen, Durán & Klevecz 1993; Klevecz 1998), it is possible, even likely, that a gene regulated so as to be coordinated with the cell cycle, need not have a single period of expression in that cell cycle. The regulatory loop for any gene can be represented by a simple system of reaction-diffusion equations that display both positive and negative feedback, and in which the set of all solutions forms a thick surface. When large numbers of such genetic regulatory loops are coupled together by facilitated diffusion of one or more of the reactants or their products, the trajectories are tuned by this interaction into differing modes of oscillation. Differentiation among the modes of expression of each gene is dependent upon the strength of coupling, the time lag between interactions, and the gene's spatio-temporal position in the ensemble. Tuning of this noisily periodic generic regulatory loop can lead to a variety of oscillatory modes for different genes including very low amplitude oscillations that are below the level of experimental detection. A gene may be tuned by these interactions to display no discernable period, or one, two, or three or more periods of synthesis and degradation in each cycle. If it should be the case that phenotype and regulation in the yeast cell cycle are emergent properties of an attractor, then novel experimental approaches can be entertained for yeast and for the mammalian cells it models. If so, and if we can track and reproduce in simulations the path taken by the system through concentration phase space, it may be possible to tune the system by small, regularly timed perturbations to synchronous division, sporulation or other desired modes of behaviour (Ott, Grebogi & Yorke 1990; Shinbrot *et al.* 1993). Such an approach might allow a more expedient experimental path to accurate prediction and effective control of cellular processes.

REFERENCES

- ALT S, RINGO J, TALYN B, BRAY W, DOWSE H (1998) The period gene controls courtship song cycles in *Drosophila melanogaster*. *Anim. Behav.* **56**, 87–98.
- BOLEN JL, DURÁN O, KLEVECZ RR (1993) Amplification and damping of deterministic noise in coupled cellular arrays. *Physica D* **67**, 245–256.
- CHO RJ, CAMPBELL MJ, WINZELER EA *et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell.* **2**, 65–73.
- DAUBECHIES I (1992) Ten Lectures on Wavelets. Philadelphia: Society for Industrial and Applied Mathematics.
- DE RISI JL, IYER VR, BROWN PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686.
- HUNT T (1991) Cyclins and their partners: from a simple idea to complicated reality. *Semin. Cell Biol.* **2**, 213–222.
- KLEVECZ RR (1998) Phenotypic heterogeneity and genotypic instability in coupled cellular arrays. *Physica D* **124**, 1–10.
- KLEVECZ RR, BOLEN J, DURÁN O (1992) Self-organization in biological tissues: analysis of asynchronous and synchronous periodicity, turbulence and synchronous chaos emergent in coupled chaotic arrays. *Int. J. Bifurcation Chaos* **2**, 941–953.
- KLEVECZ RR, KAUFFMAN SA, SHYMKO RM (1984) Cellular clocks and oscillators. In: *International Review of Cytology*, **86**, 97–128. (J. DANIELLI, ed.) Academic Press, New York.
- KLEVECZ RR, RUDDLE FH (1968) Cyclic changes in enzyme activity in synchronized mammalian cell cultures. *Science* **159**, 634.
- KLEVECZ RR, SHYMKO RM (1985) Quasi-exponential generation time distributions from a limit cycle oscillator. *Cell Tissue Kinet.* **18**, 263–271.
- KLEVECZ RR, STUBBLEFIELD E (1967) RNA synthesis in relation to DNA replication in synchronized Chinese hamster cell cultures. *J. Exp. Zool.* **265**, 259.

- OTT E, GREBOGI C, YORKE JA (1990) Controlling chaos. *Phys. Rev. Lett.* **64**, 1196–1199.
- PAULOVICH AG, TOCZYSKI DP, HARTWELL LH (1997) When checkpoints fail. *Cell* **88**, 315–321.
- SHINBROT T, GREBOGI C, OTT E, YORKE JA (1993) Using small perturbations to control chaos. *Nature* **363**, 441–447.
- SHYMKO RM, KLEVECZ RR, KAUFFMAN SA (1984) The cell cycle as an oscillatory system. In: EDMUNDS, L, ed. *Cell Cycle Clocks*, New York: Marcel Dekker, Inc., 273–293.
- SPELLMAN PT, SHERLOCK G, ZHANG MQ *et al.* (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell.* **9**, 3273–3297.
- TAYAMO P, SLONIM D, MESIROV J *et al.* (1999) Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci.* **96**, 2907–2912.
- VELCULESCU VE, ZHANG L, ZHOU W *et al.* (1997) Characterization of the yeast transcriptome. *Cell* **88**, 243–251.