Stochastic simulation for the inference of transcriptional control network of yeast cyclins genes

Jiri Vohradsky*

Laboratory of Bioinformatics, Institute of Microbiology ASCR, v.v.i., Videnska 1083, 14220 Prague, Czech Republic

ABSTRACT

Cell cycle is controlled by the activity of protein family of cyclins and cyclin-dependent kinases that are periodically expressed during cell cycle and that are conserved among different species. Genome-wide location analysis found that cyclins are controlled by a small number of transcription factors that form closed network of genes controlling each other. To investigate gene expression dynamics of this network, we developed a general procedure for stochastic simulation of gene expression process. Using the binding data, we simulated gene expression of all genes of the network for all possible combinations of regulatory interactions and by statistical comparison with experimentally measured time series excluded those interactions that formed gene expression temporal profiles significantly different from the measured ones. These experiments led to a new definition of the cyclins regulatory network coherent with the binding experiments which are kinetically plausible. Level of influence of individual regulators in control of the regulated genes is defined. Simulation results indicate particular mechanism of regulatory activity of protein complexes involved in the control of cyclins.

INTRODUCTION

Cell cycle control is one of the most important topics of molecular biology. Large part of the studies on cell cycle was done on budding yeast *Saccharomyces cerevisiae* that have served for years as a model organism for study of eukaryotic cellular processes. The yeast cell cycle consists of four distinct phases G1, S, G2 and M. In G1 phase, biosynthetic activity of the cell resumes from the slowdown in the M phase and genes active in the following S phase are transcribed, mainly those needed for DNA replication. The S phase starts with initiation of DNA replication, during this phase DNA is doubled and cell enters following G2 phase during which microtubules are synthesized, which are required during M phase when cell divides. Systems studies of cell cycle were mainly performed by omics experiments. Microarray analysis identified more than 800 genes that are periodically expressed during cell cycle (1–3). It has been found that regulation of the cell cycle is controlled by the activity of protein family of cyclins and cyclin-dependent kinases (CDK) (4,5) which are also among the periodically expressed genes group. Transcriptional regulation of the cell cycle clock was associated with a relatively small number of cell cycle-dependent transcription factors, namely Fkh1, Fkh2, Mcm1, Mbp1, Swi4, Swi6, Ndd1 and Ace2 (6–9), which are conserved among different saccharomyces species (1). Out of them, two molecular complexes that were found as acting coordinately were identified—SBF (Swi4/Swi6) and MBF (Mbp1/Swi6). Based on these studies, an initial model of transcriptional control of the cell cycle in yeast was suggested: the MBF and SBF complexes control late G1 phase (7,10). Mcm1 is involved in control of some genes of M/G1 phase (11). Mcm1, together with Fkh1 and Fkh2 that associated with Ndd1 controls transcription of G2 and M genes (8). M and early G1 phase are controlled by Swi5 and Ace2.

To investigate deeper this scheme, Simon *et al.* (12) ran a set of genome-wide location experiments to find out how the cell cycle transcription program is controlled by each of the known cell cycle transcription activators. By combination of their binding data and previous studies, Simon *et al.* identified and extended model of transcriptional control of cyclin/CDK genetic network, where each group of transcription factors controls key cell cycle regulators needed for cell cycle progression (Figure 1A). Since then, no systematic experimental study on the control of

---

*To whom correspondence should be addressed. Tel: +420 241062513; Fax: +420 241722257; Email: vohr@biomed.cas.cz
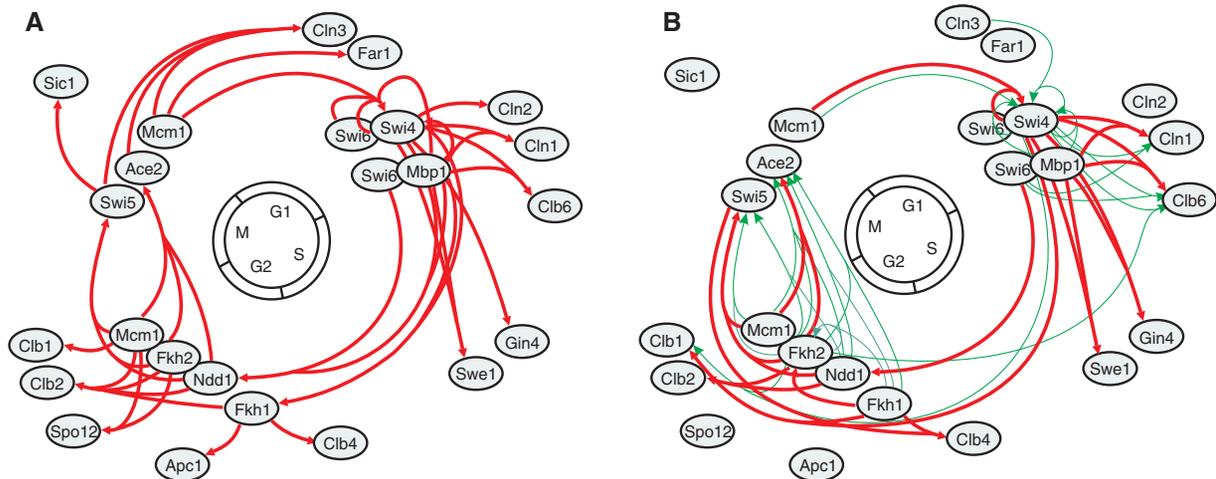
Figure 1. Interactions found by genome-wide location analysis (**A**) and computed using the simulation procedure presented here (**B**). In panel B, red lines represent interactions having the highest correlation with experimental time series, and green lines represent alternative interactions that are still statistically significant but have a correlation coefficient lower than that of the interactions represented by red lines.

cyclins network has been performed. In their paper, Simon *et al.* suggests a new model of transcriptional control of cell cycle regulators that define cyclins genetic network. Based on the location analysis, they found that SBF and MBF complexes control transcription of G1/S cyclin genes, but also regulate expression of the G2/M cyclin Clb2, which promotes entry into mitosis. SBF and MBF also regulate the transcription of the transcription factor Ndd1, which also binds the CLB2 promoter. Thus, SBF, MBF and Ndd1 ultimately collaborate to regulate transcription of the CLB2 gene. SBF and MBF therefore regulate genes necessary for the transition through G1/S, as well as genes whose products set the stage for further progression through the cell cycle. Their data also revealed that the G2/M activators (Mcm1/Fkh2/Ndd1) bind genes whose expression is necessary for both entry into and exit from mitosis. The G2/M activators bind and regulate transcription of CLB2, whose product is necessary to enter mitosis. These genes also set the stage for exit from mitosis by regulating the gene encoding Cdc20, an activator of the APC, which targets the APC to degrade Pds1 and thus initiate chromosome separation. Cdc20-activated APC also degrades Clb5 and thus enables Cdc14 to promote the transcription and activation of Sic1 and to initiate the degradation of Clb2. In addition, the G2/M activators Mcm1/Fkh2/Ndd1 regulate transcription of SPO12, which encodes a protein that also regulates mitotic exit. The M/G1 transcriptional regulators (Mcm1, Ace2 and Swi5) bind genes that are key to entering and progressing through G1. Swi5 binds to the SIC1 promoter, and all three transcriptional regulators bind to the CLN3 promoter. Sic1 inhibits Clb–Cdc28 during mitosis, thus facilitating exit from mitosis. Cln3–Cdc28 activates SBF and MBF in late G1, thus setting the stage for another cell cycle circuit (12).

Later a systematic study dealing with binding of extended number of transcriptional regulators to DNA was performed (13). Using genome-wide location analysis with epitope tagging Harbison *et al.* determined

the genomic occupancy of 203 DNA-binding transcriptional regulators including those identified as members of the cyclins genetic network. Similar work that focused on conservation of transcriptional regulators across different saccharomyces species was performed recently (1).

Based on these studies, we have focused on whether the regulatory interactions within the cyclins control genetic network, derived from static binding, are also reflected in their kinetics. For this purpose, we used microarray kinetic measurements monitoring in discrete time intervals levels of expression of whole yeast transcriptome (2). Using a dedicated simulation model of gene expression and a simple statistics, we investigated all interactions suggested by the binding experiments from the kinetic point of view and adapted the cyclins genetic network to conform to both experiments. Final network included interactions coherent with the binding experiments that were also kinetically plausible.

MATERIALS AND METHODS

Simulation of gene expression

Gene transcription, from the regulatory point of view, starts with a transcription factor binding to the promoter region of a particular gene. After the regulator binds to DNA, the RNA polymerase complex binds to the start codon of the gene and initiates its transcription into messenger RNA. The probability that the transcriptional event will occur depends on the number of regulator molecules around the gene and their strength of binding to the promoter of a gene. Under normal conditions, RNA polymerase and other transcriptional factors involved in the process are considered to be available in excess, and their concentration in the cell does not influence the accumulation of transcribed mRNA molecules. The number of mRNA molecules accumulated in a cell is influenced by its degradation.

In the cell environment, the number of regulators molecules is generally low and the occurrence of the transcriptional process for particular gene is stochastic. Stochasticity of general biochemical reactions has been modeled by Gillespie's algorithm (14,15). The algorithm was originally developed for chemical bimolecular reactions and simulates collisions that lead to the formation of a new molecule. The probability that a bimolecular reaction will take place is dependent on the occurrence of individual molecules. The algorithm starts with the initialization step, in which the number of molecules in the system is defined along with the reaction constants and random number generators. In the following Monte Carlo step, a random number is generated to determine the next reaction to occur, and the time step is updated. This procedure is iterated until either all the input molecules or the simulation time are exhausted. The algorithm relies on the knowledge of reaction constants and concentrations of the acting molecules—the items that are very rarely available for transcriptional processes. Most of the concentration measurements are performed using microarrays in relative units where the exact number of molecules in the cell is hard to estimate. Also Gillespie's algorithm was developed for bimolecular reactions which, in case of more than one regulator, make simulation of transcription difficult to achieve.

For this reasons, it is necessary to find relation between relative values of gene expression levels of regulators, measured by microarrays and the probability of transcriptional event occurrence. Veitia (16) using thermodynamic analysis of the process of transcription derived that the relation between concentration of regulators and the probability of transcription is given by a sigmoidal curve, with number of thermodynamic parameters. Reason for using a sigmoid is supported by two observed features. First is the saturated promoter, when local concentration of the regulator is so high that transcription occurs with the probability close to one. Further increase of regulators concentration does not increase probability of transcription. Second, for initiation of transcription a basal level of regulators concentration is necessary, below which transcription does not occur. Its probability is therefore zero. Based on this analysis, we suggested a simple model where the probability of occurrence of transcriptional event is given by a sigmoid in which the influence of the given regulator is weighted by a constant that can be associated with the strength of the promoter binding of the given regulator. Magnitude of this constant corresponds to the range and units in which the gene expression levels are measured. Probability of transcription event occurrence is thus given by a formula:

$$p(t) = \frac{1}{1 + \exp\{-[wy(t) + b]\}} - B \qquad (1)$$

The affinity constant w represents the level of influence of given regulator's concentration $y(t)$, the delay parameter b is a constant <0 that determines the amount of y for a given w that is high enough to start the transcriptional process. B is a correction constant specifying probability of transcription for $y(t) = 0$. For saturated promoter $p(t) = 1 - B$, for typical range of the parameters $B < 0.001$. As in the low molecular environment in the cell, transcription can never reach probability of its occurrence equal 1, while for $y(t) = 0$ probability of transcription has to be 0. Parameter B corrects for this phenomenon.

The formula can be generalized for n regulators to

$$p(t) = \frac{1}{1 + \exp\{-[\sum\limits_{i=1}^{n} w_i y_i(t) + b]\}} - B \qquad (2)$$

(according to experimental observations, it is reasonable to consider only a maximum of three regulators).

Simulation of gene transcription process

To simulate the process of transcription and accumulation of the transcribed gene mRNA with a computer, the probability of transcription was computed according to Equation 1 or 2, and a random number r from a uniform distribution on the interval (0,1) was generated. When the computed probability of transcription satisfies $p > r$, transcription starts and generates one molecule of mRNA of the regulated gene; accumulated $y_{mRNA}(t + \Delta t) = y_{mRNA}(t) + 1$. Transcription takes time Δt proportional to the length of the gene here set to 50 nt/s (17). Accumulated mRNA is subjected to natural degradation, which is assumed to follow the first-order chemical reaction. Therefore, the amount of mRNA degraded during the transcription process is:

$$y_{mRNAd} = y_{mRNA}(t + \Delta t) \exp(-k_d \Delta t) \qquad (3)$$

k_d is the degradation constant. This amount is subtracted from the accumulated mRNA amount $(y_{mRNA}(t + \Delta t) = y_{mRNA}(t + \Delta t) - y_{mRNAd})$. If $p \leq r$, t is increased by Δt and the degraded mRNA is subtracted from the accumulated one. The simulation is repeated until the end of the established time interval is reached.

Model parameter inference from gene expression data

The model defined in Equations 1, 2 and 3 contains several parameters that must be estimated from gene expression data. The model simulates process of gene expression on the level of transcription and generates an mRNA expression profile, which is a result of regulator activity and a set of parameters. The parameters deduced from Equations 1, 2 and 3 were as follows: the regulator binding affinity constant w_i, the delay parameter b and the degradation rate constant k_d. Two additional constants must be added: the initial amount of mRNA y_{mRNA0} and a constant k, which serves as a multiplier in Equations 1 and 2 and scales the arbitrary units in which the regulators' gene expression profile is measured to the scale into the units used in the model: $y_{mRNAcor} = y_{mRNA0} + k \cdot y_{mRNA}$.

The parameters of a process can be computed by fitting the simulated gene expression profile of the regulated gene to the measured one with the use of regulators expression profile $y(t)$; this is achieved by maximizing the correlation between the measured process $y_{mRNA}(t)$ and the simulated

$\hat{y}_{mRNA}(t)$ profile, i.e. by minimizing the objective function Q given by

$$Q = 1 - \text{cor}(\bar{\hat{y}}_{mRNA}, \bar{y}_{mRNA}) \tag{4}$$

Due to the stochasticity of the simulation process, each simulation generates a slightly different profile; thus, an estimation of the parameters belonging to each individual simulation would imply a different set of parameters. Thus, parameters were computed from the average profile. Experiments showed that the average profile stabilizes after ∼20 runs. Therefore, an iterative optimization procedure, minimizing the objective function Q (Equation 4), was run for an average profile containing at least 20 simulations. Simulated average profile was then transformed linearly to fit to the scale of the measured one.

For each regulated (target) gene, ChIP-on chip data provide information about binding of regulators to the promoters of given regulated gene. These regulators can be considered as potentially controlling expression of this gene. Therefore, for each target gene, the above described optimization was run for all combinations of single and two potential regulators for all regulators suggested by the binding experiments. Control by more than two regulators was analyzed only in case where it was suggested by Simon *et al.* (12).

Potential regulators for the genes of cyclins regulatory network were identified from the binding data of Harbison *et al.* (13) and Simon *et al.* (12).

Inference of regulator–target gene interaction

To find out which potential regulators can regulate given target gene, simulations with optimized parameters were run 20 times for all target genes and all possible combinations of their regulators as described in the previous paragraph. Measured target gene expression profiles had mean $\bar{y}_{mRNA}(t_i)$ and variance $s_{mRNA}(t_i)$ for discrete number of time points $i = 1{:}m$. Each simulation generated a set of target gene temporal profiles with mean $\hat{y}_{mRNA}(t_i)$ and a variance $s_{ymRNA}(t_i)$. The simulated average target gene profiles were compared with measured ones and those simulated profiles that were significantly different from the measured ones were rejected. As each accepted profile was computed from a certain combination of regulators, these regulators were taken as true regulators of the given target gene.

The statistical test had to say that not only the average simulated profile was not significantly different from average measured one as whole, but also that it was not different in any of time points measured. This in principle can be achieved by two-way ANOVA with post-test, testing whether the difference between the simulated and measured experiments was statistically significant in each time point. We used two-way ANOVA test with Tukey's least significant difference procedure to identify time points where the simulated and measured profiles were significantly different. Therefore, those simulated profiles, and consequently, the combinations of regulators for the given target gene, were accepted that passed the

two-way ANOVA test and were not different in more than one time point out of 25 measured.

Measured gene expression profiles were adopted from the work of Pramila *et al.* (2) where three experiments were run, measured in 25 time points in 5 min interval over 120 min.

RESULTS AND DISCUSSION

Genetic network controlling cyclins consists of just a few transcription factors—Swi6, Swi4, Mbp1, Mcm1, Fkh1, Fkh2 and Ace2. It was originally derived by Simon *et al.* (12) on the basis of genome-wide location analysis (Figure 1A). The genes of this network control each other and also cyclins that control cell cycle progression and periodicity. They are: B-type cyclins involved in cell cycle progression Clb1,2,4,6 that activate Cdc28 to promote the transition from G2 to M phase or function in formation of mitotic spindles along with Clb3 and Clb4 (Clb6). Another group of genes they control are Cln1,2,3; cyclins involved in regulation of the cell cycle, activating Cdc28 kinase and promoting G1 to S phase transition. Remaining controlled genes are Apc1 that forms largest subunit of the anaphase-promoting complex/cyclosome (APC/C), and kinase inhibitors Sic1 and Far1.

The above-mentioned simulation procedure was used to infer topology and interactions within this network. For this purpose, microarray experiments were used that covered two cell cycles of yeast that had been made in triplicate (2). Original data were stored as log base 10 values. For further use, the data were exponentiated and individual profiles corresponding to the same gene from the three parallel experiments were averaged and standard deviation for each time point was calculated. Binding interactions among members of this network were identified from ChIP experiments (13) and have been assembled in Table 1. Table 1 served for the design of necessary simulations, modeling both single regulators and two regulators acting together. The number of necessary simulations is given by the Table 1 column labeled 'comb'. In the case of genes for which three regulators were suggested by Simon *et al.* (NDD1, SWI5, ACE2, SWI4, CLN3 and CLB2), we also performed a simulation for three regulators. Altogether 125 simulations were made. For each simulation, model parameter estimates were computed (see 'Materials and Methods' section), and a statistical test was used to compare average simulated profiles with the average experimentally measured profile. Those simulated profiles passing the statistical test were selected, and the values of the parameters w_{ij} were used to infer the regulators of the given gene and their activity (activator or repressor). The results are summarized in Table 2.

A comparison of Tables 1 and 2 shows the differences between the predictions made solely based on genome-wide location data and data computed using the simulation model. A comparison of the two approaches is given in Table 3. Table 3 shows the following: (i) there was a partial overlap between the genome-wide location data analysis and the results of simulation experiments.

Table 1. Regulator binding to the promoters of regulated genes in cyclin/CDK regulators network as identified by ChIP experiments (13)

	FKH1	FKH2	NDD1	MCM1	SWI5	ACE2	SWI6	SWI4	MBP1	CLN3	SUM	comb
FKH1		1									1	1
FKH2	1	1									2	3
NDD1		1					1	1	1		4	10
MCM1											0	0
SWI5	1	1	1	1							4	10
ACE2	1	1	1	1							4	10
SWI6											0	0
SWI4				1			1	1	1	1	5	15
MBP1										1	1	1
CLN3				1	1	1		1			4	10
CLB4	1				1						2	3
CLB1	1	1		1				1			4	10
CLB2	1	1	1	1				1	1		6	21
SIC1					1	1					2	3
FAR1				1			1				2	3
CLN2				1			1	1	1		4	10
CLN1	1	1		1			1	1			5	15
CLB6		1					1	1	1		4	10
GIN4				1			1	1	1		4	10
SWE1							1	1	1		3	6
SPO12		1		1							2	3
APC1	1										1	1
SUM	8	10	3	11	3	2	8	10	7	2	61	125

The digit 1 in a cell indicates that the regulator from that column binds the promoter of the regulated gene from the corresponding row. Shaded cells label interactions deduced from the work of Simon *et al.* (Figure 1). The SUM of a given row represents the potential number of regulators controlling the gene assigned to that row, whereas the SUM of a column represents the number of genes that can be controlled by the regulator assigned to that column. The row statistic 'comb' represents the total number of possible combinations of one and two regulators for the gene in that row.

(ii) In some cases, the statistical test for the similarity between the simulated and measured target gene expression profiles was satisfied by more than one combination of regulators. (iii) The 'cc' column indicates a high correlation between simulated and measured mean expression profiles, a correlation that is always >0.9. The alternative combinations of regulators must be considered equal and cannot be excluded based on differences in the correlation. All are statistically significant, and all are very highly correlated. Interactions suggested by genome-wide location analysis (12) and those found by the simulations are depicted in Figure 1. Out of 17 total regulated genes, six of them were found to have the same regulators by genome-wide location analysis and simulations, seven were found to have different regulators and the remaining four genes indicated partial regulator overlap between the two types of analyses. In the following paragraph, individual regulatory interactions and differences between the results of simulation and genome-wide location analysis are discussed.

FKH1—Suggested control by the Swi4, Mbp1 pair was not confirmed. Control by Fkh2 suggested from ChIP experiments was not able to simulate Fkh1 profile with sufficient accuracy. Although correlation between measured and simulated expression profiles was quite high (0.90), it failed to pass our statistical test in 5 points (approximately at 30 h and above 95 h, see Supplementary Dataset).

FKH2—This interaction was not predicted by Simon *et al.* ChIP data suggested control by Fkh1/Fkh2. Control by Fkh1 or Fkh1/Fkh2 was confirmed by simulation.

NDD1—Suggested control by Swi6/Swi4/Mbp1 was not confirmed; instead, control only by Swi6 was found.

SWI5—Suggested control by triplet Fkh2/Ndd1/Mcm1 was confirmed but only for the case in which Ndd1 and Mcm1 acted as repressors. Because in genome-wide location analysis, the type of control was not discussed, so this result can be considered as matching. The same phenomenon was observed for the remaining combinations (Fkh2/Mcm1 or Fkh2/Ndd1 or Fkh1/Ndd1), where Mcm1 and Ndd1 always acted as repressors.

ACE2—Suggested control by triplet Fkh2/Ndd1/Mcm1 was confirmed with Mcm1 acting as repressor. There was a relatively low influence of Ndd1 (5 times lower than that of Fkh2; see Supplementary Dataset). Other pair-vice combinations for which Fkh2 or Fkh1 acted as activators and the other genes as repressors were identified.

SWI4—Suggested control by Mcm1/Swi6/Swi4/Mbp1 was confirmed for Mcm1/Swi4/Mbp1. In all known cases, self control of Swi4 was confirmed, accompanied by a second regulator (Mcm1, Cln3 or Mbp1).

CLN3—No combination of regulators was found to be possible for this gene, including suggested control by Mcm1/Swi5/Ace2.

CLB4—Suggested control by Fkh1 was confirmed, accompanied by Swi5, which was suggested to act as repressor.

CLB1—Instead of Mcm1, Clb1 indicated control by Fkh1 or Fkh2, together with Swi5 as repressor.

CLB2—From the suggested control by Fkh1/Fkh2/Ndd1/Mcm1, only the pair Fkh2/Ndd1 was sufficient to control Clb2, with Ndd1 acting as repressor.

Table 2. The results of the inference of regulatory interactions within the cyclin/CDK regulatory network using our simulation procedure

	FKH1	FKH2	NDD1	MCM1	SWI5	ACE2	SWI6	SWI4	MBP1	CLN3
FKH1		1								
FKH2	1	1								
FKH2	1	1								
NDD1		1					1	1	1	
MCM1										
SWI5	1	1	1	1						
SWI5	1	1	1	1						
SWI5	1	1	1	1						
ACE2	1	1	1	1						
ACE2	1	1	1	1						
ACE2	1	1	1	1						
ACE2		1	1	1						
SWI6										
SWI4				1			1	1	1	1
SWI4				1			1	1	1	1
SWI4				1			1	1	1	1
SWI4				1			1	1	1	1
MBP1										1
CLN3				1	1	1		1		
CLB4	1				1					
CLB1	1		1	1				1		
CLB1	1	1		1				1		
CLB2	1	1	1	1				1	1	
SIC1					1	1				
FAR1				1			1			
CLN2				1			1	1	1	
CLN1	1	1		1			1	1		
CLN1	1	1		1			1	1		
CLB6		1					1	1	1	
CLB6		1					1	1	1	
CLB6		1					1	1	1	
GIN4				1			1	1	1	
GIN4				1			1	1	1	
SWE1							1	1	1	
SPO12			1	1						
APC1	1									

The digit 1 in a cell indicates that the regulator from the corresponding column binds to the promoter of the regulated gene from the corresponding row (as in Table 1). Shading represents regulatory interactions inferred from the simulation process. Multiple appearances of the same gene name (see the first column) indicate significant alternative control for that gene.

SIC1, FAR1 and CLN2—No kinetically plausible combination of regulators was found for these genes.

CLN1—Control by Swi4/Mbp1 was confirmed together with the alternative pair Swi6/Swi4.

CLB6—Swi4/Mbp1 pair was confirmed together with alternative pair Swi4/Swi6 as well as Swi4/Fkh2.

GIN4—Suggested control by Mbp1 was confirmed by combination with Swi4 or alternatively Swi6/Swi4 (SBF complex).

SWE1—Control by Swi4/Mbp1 was confirmed.

The kinetic behavior of transcription of the selected genes was modeled with high accuracy, typically yielding a correlation coefficient between measured and simulated profiles higher than 0.98 (Table 3), with statistically significant similarity between measured and simulated expression profiles for 24 out of 25 total measured points.

In several cases, regulation suggested by genome-wide location analysis for two or more regulators proved redundant from the kinetic point of view. For NDD1, genome-wide location analysis suggested three regulators (Swi6/Swi4/Mbp1), whereas for the simulation of Ndd1

kinetics, Swi6 alone was sufficient. Similarly, for SWI4, suggested control is by Mcm1/Swi6/Swi4/Mbp1, whereas for the simulation, the regulators Mcm1/Swi4/Mbp1 were sufficient (Table 3). When simulation and experimental variance were used to confirm or reject control by different combinations of regulators for a given target gene, several alternative control patterns for one target gene could be discerned; e.g. for the case of SWI5, control by a triplet Fkh2/Ndd1/Mcm1 was statistically equivalent to control by pairs Fkh2/Mcm1 and Fkh2/Ndd1 (a similar result was obtained for CLB6, SWI4 and CLN1; Table 3). These individual regulator combinations were indistinguishable from each other, and they must be considered equivalent. Such results constitute alternative hypotheses which have to be confirmed independently.

Two complexes SBF (Swi4/Swi6) and MBF (Mbp1/Swi6) are supposed to control cyclins CLN1,2 and CLB5,6 respectively (18) and NDD1 together with Mbp1 (12). CLN1,2 was indeed found to be regulated by SBF, alternatively by Swi4/Mbp1 pair. In accordance with the binding experiments SWI4/MBP1 (or alternatively SBF) was found to control also CLB6. Simulation

Table 3. Comparison of cyclin/CDK control network revealed by genome-wide location analysis and the simulations performed in this article

	Simon *et al.*	Simulation	cc
FKH1	SWI4, MBP1	NULL	
FKH2	NULL	FKH1	0.98
FKH2		FKH1, FKH2	0.97
NDD1	SWI6, SWI4, MBP1	SWI6	0.93
MCM1	NULL	NULL	
SWI5	FKH2, NDD1, MCM1	FKH2, NDD1, MCM1	0.98
SWI5		FKH2, MCM1	0.98
SWI5		FKH2, NDD1	0.98
ACE2		FKH2, MCM1	0.96
ACE2	FKH2, NDD1, MCM1	FKH1, NDD1	0.94
ACE2		FKH2, NDD1, MCM1	0.94
ACE2		FKH1, FKH2	0.94
SWI6	NULL	NULL	
SWI4		MCM1, SWI4	0.99
SWI4		SWI4, CLN3	0.99
SWI4		SWI4, MBP1	0.99
SWI4	MCM1, SWI6, SWI4, MBP1	MCM1, SWI4, MBP1	0.98
MBP1	NULL	NULL	
CLN3	MCM1, SWI5, ACE2	NULL	
CLB4	FKH1	FKH1, SWI5	0.89
CLB1	MCM1	FKH1, SWI4	0.98
CLB1		FKH2, SWI4	0.98
CLB2	FKH1, FKH2, NDD1, MCM1	FKH2, NDD1	0.99
SIC1	SWI5	NULL	
FAR1	MCM1	NULL	
CLN2	SWI4	NULL	
CLN1	SWI4, MBP1	SWI4, MBP1	0.94
CLN1		SWI6, SWI4	0.94
CLB6	SWI4, MBP1	SWI4, MBP1	0.99
CLB6		SWI4, SWI6	0.98
CLB6		SWI4, FKH2	0.94
GIN4	MBP1	SWI4, MBP1	0.95
GIN4		SWI6, SWI4	0.95
SWE1	SWI4, MBP1	SWI4, MBP1	0.94
SPO12	MCM1, FKH2		
APC1	FKH1		

If more than one combination of regulators satisfied the statistical test for a given target gene, then they are all listed. Shaded regions indicate interactions matching exactly in both types of analyses. The column denoted by 'cc' contains Pearson correlation coefficients between simulated and measured expression profiles of the target genes (first column). Multiple occurrences of the same gene in the first column indicate alternative control for that gene.

of CLB6 control by MBF suggested by literature showed low correlation with experimental time series (Pearson correlation = 0.34). For the control of NDD1, Swi6 only was found to be sufficient although the simulation of the control by MBF showed to be highly correlated with experimental time series as well (Pearson correlation = 0.93, see Supplementary Dataset).

When the same set of regulators was identified by both experiments (simulation versus genome-wide location analysis), some regulators located by simulation were required to serve as repressors. For example, take ACE2. Both experiments suggest three regulators (Fkh2/Ndd1/Mcm1), but simulation found that Mcm1 must act as repressor, or else the Ace2 expression profile could not be modeled successfully. Similarly for SWI5 (regulators Fkh2/Ndd1/Mcm1), Ndd1 and Mcm1 had to act as repressors to simulate the expression profile

of Swi5. Genome-wide location analysis does not address this issue. After searching the literature, we found that for this case, no one has investigated the regulators' activity as an activator or repressor.

For four particular genes (SIC1, FAR1, CLN2 and CLN3; Table 3), no combination of regulators suggested by ChIP-chip experiments could simulate the kinetics of the target genes with sufficient confidence. Control suggested by genome-wide location analysis was thus not confirmed by simulation. This observation indicates that the control of these genes may follow different pathway than purely transcriptional.

The differences between binding and simulation experiments could be caused principally by two items: measurement inaccuracy and a type of control different than the transcriptional one. Time series of gene expression are usually measured using microarrays, and number of replicates usually does not exceed three. When the measurement covers whole cell cycle or two (as for the data used here), the synchrony of the population decays with time, leading to increase of variance with time. As a result, the average measured expression profile may differ from the real one. As the simulation is designed to fit the experimental data, any error in them distorts final conclusion. Increase of the measurement reproducibility will also increase reliability of the predictions based on the simulations.

The simulation used here, suppose transcriptional control and investigates that this type of control is possible from the kinetic point of view. When other type of control is involved, then the simulation cannot find the correct solution. The same argument holds for interpretation of the binding experiments which say that the regulator binds the promoter, but cannot discover posttranscriptional or other events and cannot even say whether the binding results in transcription. If, e.g. phosphorylation is involved in transcriptional control, the binding experiments' results, combined with the simulation results, will fail to discover the real control mechanism. The additional information, the phosphorylation, is missing. Control of CLB2 can serve as an example. CLB2 was found to be controlled by Mcm1 together with Fkh2 and co-activator Ndd1. But Clb2-Cdc28 phosphorylates Ndd1, which is important for recruitment to the CLB2 gene promoter, and phosphorylates Fkh2, which enhances the interaction of Fkh2 with Ndd1 (18). The simulation found that only the pair Fkh2, Ndd1 was sufficient to control CLB2 gene, with Ndd1 acting as repressor. Therefore, the multiple phosphorylations involved in the control were replaced by kinetically acceptable alternative. If the current knowledge indicates that the control of the given gene is transcriptional and the results of kinetic simulation differ from those measured statically (as the binding experiments) it is an indication that the process follows different pathway then suggested, i.e. most probably some information is missing. Therefore, such situations have to be carefully checked and additional experiments have to be designed that would investigate involvement of other possible control mechanisms. As the binding experiments usually indicate several possible combinations of regulators, simulation can choose

those combinations which are kinetically possible. Role of the modeling in these cases is in questioning or confirmation of the current knowledge.

Role of the simulation of transcriptional process presented here is in investigation of kinetic plausibility of regulatory interactions suggested by other experiments. The simulation of molecular interactions between transcription factors and promoter of the controlled gene is closest possible approach to the transcriptional process which allows simulating natural variance of the process given by inevitably small number of interacting molecules. When tested against experimentally measured time series, all kinetically plausible regulatory interactions can be revealed. Modeling, in this case, serves as additional confidence level of data interpretation and a means for the design of new experiments leading to discovery of, so far, unknown control mechanisms. Second role of the simulation experiments is in modeling. Once the correct pathway is known, the output caused by changes to the given system can be tested without making expensive experiments. Simulation allows studying the dynamic features of the whole network of regulatory interactions that goes beyond purely experimental observations.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Dataset.

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Petricka,J.J. and Benfey,P.N. (2011) Reconstructing regulatory network transitions. *Trends Cell Biol*., **21**, 442–451.
2. Pramila,T., Wu,W., Miles,S., Noble,W.S. and Breeden,L.L. (2006) The Forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle. *Genes Dev*., **20**, 2266–2278.
3. 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.
4. Orlando,D.A., Lin,C.Y., Bernard,A., Wang,J.Y., Socolar,J.E., Iversen,E.S., Hartemink,A.J. and Haase,S.B. (2008) Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature*, **453**, 944–947.
5. Nigg,E.A. (1995) Cyclin-dependent protein kinases: key regulators of the eukaryotic cell cycle. *Bioessays*, **17**, 471–480.
6. Bahler,J. (2005) Cell-cycle control of gene expression in budding and fission yeast. *Annu. Rev. Genet*., **39**, 69–94.
7. Koch,C. and Nasmyth,K. (1994) Cell cycle regulated transcription in yeast. *Curr. Opin. Cell Biol*., **6**, 451–459.
8. Koranda,M., Schleiffer,A., Endler,L. and Ammerer,G. (2000) Forkhead-like transcription factors recruit Ndd1 to the chromatin of G2/M-specific promoters. *Nature*, **406**, 94–98.
9. Lu,Y., Mahony,S., Benos,P.V., Rosenfeld,R., Simon,I., Breeden,L.L. and Bar-Joseph,Z. (2007) Combined analysis reveals a core set of cycling genes. *Genome Biol*., **8**, R146.
10. Iyer,V.R., Horak,C.E., Scafe,C.S., Botstein,D., Snyder,M. and Brown,P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, **409**, 533–538.
11. MacKay,V.L., Mai,B., Waters,L. and Breeden,L.L. (2001) Early cell cycle box-mediated transcription of CLN3 and SWI4 contributes to the proper timing of the G(1)-to-S transition in budding yeast. *Mol. Cell Biol*., **21**, 4140–4148.
12. Simon,I., Barnett,J., Hannett,N., Harbison,C.T., Rinaldi,N.J., Volkert,T.L., Wyrick,J.J., Zeitlinger,J., Gifford,D.K., Jaakkola,T.S. *et al.* (2001) Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, **106**, 697–708.
13. Harbison,C.T., Gordon,D.B., Lee,T.I., Rinaldi,N.J., Macisaac,K.D., Danford,T.W., Hannett,N.M., Tagne,J.B., Reynolds,D.B., Yoo,J. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99–104.
14. Gillespie,D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem*., **81**, 2340–2361.
15. Gillespie,D.T. (2007) Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem*., **58**, 35–55.
16. Veitia,R.A. (2003) A sigmoidal transcriptional response: cooperativity, synergy and dosage effects. *Biol. Rev. Camb. Philos. Soc*., **78**, 149–170.
17. Lodish,L., Berk,A., Zipurski,S.L., Matsudaira,P., Baltimore,D. and Darnell,J.E. Jr (2000) Molecular Cell Biology. W. H. Freeman and Company, New York.
18. Bloom,J. and Cross,F.R. (2007) Multiple levels of cyclin specificity in cell-cycle control. *Nat. Rev. Mol. Cell Biol*., **8**, 149–160.