

Robustness vs. identifiability of regulatory modules?

The case of mitotic control in budding yeast cell cycle regulation

Jörg Stelling

Max-Planck-Institute for Dynamics
of Complex Technical Systems
Leipziger Str. 44
Magdeburg, Germany

stelling@mpi-magdeburg.mpg.de

Ernst Dieter Gilles

Max-Planck-Institute for Dynamics
of Complex Technical Systems
Leipziger Str. 44
Magdeburg, Germany

gilles@mpi-magdeburg.mpg.de

ABSTRACT

Invariant input-output behavior and pathway redundancy provide a competitive advantage for cellular survival. Therefore robustness of cellular control circuits seems to be a general, essential feature. The same property however provides a challenge to the investigator because it hinders the estimation of kinetic parameters and thus the quantitative understanding of cellular regulation. The current study investigated the value of mathematical modeling in the elucidation of cellular control circuits under these constraints. A subsystem responsible for mitotic control in budding yeast cell cycle regulation was chosen as example. On the basis of scarce quantitative experimental data we were able to develop a detailed mechanistic model. The model showed desirable descriptive and predictive character. For instance, model predictions agreed well with experimental observations without additional parameter adaptation. Furthermore, the model allowed for specifying barely characterized regulatory interactions. Determination of parameter estimation accuracy subsequently showed that the information content of the data available was much higher than generally expected. Combined data for wild type and four variants with only five measured variables lead to acceptable estimates for about half of the 114 kinetic parameters in this complex model. Analysis also proved, that the combination of data from the unperturbed and the disturbed control circuit was essential for this outcome. Parameter estimation accuracy additionally gives a quantitative measure for robustness, which in our case strongly supports the concept of highly optimized tolerance, i.e. of the co-existence of robustness and fragility in cellular control. A differentiated view on robustness and identifiability is thus required, but based on our experience, realistic models of cellular control are possible despite limited quantitative data.

1. INTRODUCTION

Cellular regulatory networks have for a long time been seen as overwhelmingly complex, highly interacting assemblies of genes and their products. Only recently, it has been suggested, that the architecture of these networks is in fact modular: different tasks are fulfilled by specialized quasi-autonomous subsystems (modules), which perform functions analogous to controllers in technical processes. The necessity of fault tolerance then seems to be one explanation for the complexity of cellular control circuits [21, 31]. A way to

achieve fault tolerance is to evolve robust regulatory modules. Here, robustness of a (biochemical) network means the relative insensitivity of the systems' qualitative behavior towards perturbances in its kinetic parameters [4]. From a biological point of view, robustness results in "buffering" of cellular control in case of small mutations or noisy environmental conditions. It is supposed to be an essential feature of for example proliferation control and development, as here robust control is closely linked to the organism's survival.

This aspect of biological regulation has been examined by mathematical modeling of chemotaxis in *Escherichia coli* [4, 2] and of the segment polarity network in *Drosophila* [35]. In both cases, the models displayed the qualitative behavior observed *in vivo* and robustness of this behavior, for instance perfect adaptation in chemotaxis. These features turned out to be properties of the model structure, and only to a very limited degree of the parameter values. If this finding would hold true in general, it would have important implications for the understanding of cellular systems as a whole: The key characteristics of modularity and robustness could greatly facilitate the mathematical modeling of larger regulatory networks.

However, for a quantitative understanding at the system level, finding the right model structures for small networks and then combining them will prove insufficient. Also the quantitative aspects of cellular regulation have to be described in an appropriate way first of all to come to a tight coupling of model and experiment and finally to enable the purpose-driven manipulation of the system. For mathematical models this task implies identifying the model parameters as well as the structure. This can in so far be complicated, as robustness and identifiability are contradictory system characteristics: For a robust module, the input-output behavior gives little information on the internal working principles, as it is almost invariant. The quantitative (internal) behavior also of robust systems nevertheless varies if perturbed. The most eminent difficulties are therefore thought to arise from the combination of robustness and today's prevailing qualitative biological knowledge [16]. Under these circumstances, the correct determination of internal parameters can be impeded, especially if robustness depends on the cooperation of multiple, redundant pathways in order to achieve the control objectives [15].

As a more differentiated view on (cellular) systems, Doyle and co-workers recently introduced the theory of highly optimized tolerance (HOT). It points out the combination of generally robust input-output behavior, but extreme sensitivity or fragility in the case of selective disturbances [8, 25]. As a consequence, regulatory modules obeying these principles could be more easily captured in quantitative terms. Theoretical concepts on robustness are thus directly linked with the development of acceptable mathematical descriptions in cell biology. Nevertheless, both a preliminary distinction of the competing theories, as well as an answer to the question in how far quantitative mathematical modeling is possible given the current status of biological knowledge, can only be obtained by studying concrete biological examples.

The aim of this study is firstly to establish a mathematical model able to describe the system under consideration - the control of mitotic events in budding yeast cell cycle regulation - in sufficient quantitative accuracy. Extending the pioneering work of Novak and Tyson in this field [9, e.g.], we aim at providing a more detailed picture of the regulatory processes, both by regarding an enlarged network and by using parameter estimation to adjust the model to experimental data. The second part of our contribution deals with a differentiated study of model robustness and model identifiability by means of sensitivity analysis and determination of parameter estimation accuracy.

2. CELL CYCLE REGULATION IN YEAST

2.1 The general picture

In every eukaryotic cell, passage through the cell cycle is a tightly controlled process. Cell cycle regulation, the control system, is essential for the coordination of cell growth and division. The complexity of this task can be estimated from the fact, that in budding yeast the transcription of approximately 800 out of 6300 genes is cell-cycle regulated [30].

At the physiological level, the cell division cycle is characterized by a fixed sequence of cell cycle phases. Cells starting the cycle in G1 phase have to reach a critical size for all subsequent processes to begin. Once this aim is accomplished, a checkpoint called 'START' enables DNA replication in S phase. The next important cell cycle phase is mitosis (M), during which chromosomes have to be aligned and attached to the spindle apparatus in metaphase, before in anaphase they are distributed to mother and daughter cell and finally cytokinesis can take place. Each of these steps is controlled by checkpoints for e.g. correct spindle position.

At the molecular level, cyclin dependent kinases (CDKs), their phase-specific activators called cyclins and their inhibitors (CKIs) form the core of the regulatory network. In *Saccharomyces cerevisiae*, one catalytic subunit (Cdc28), nine cyclins - the G1 cyclins Cln1-3, the S phase cyclins Clb5-6 and the mitotic cyclins Clb1-4 - and the inhibitor Sic1, which specifically inactivates B-type cyclins, are involved in cell cycle regulation [24]. This enables the cell to form distinct kinase complexes with different functionality. The phase-specific cyclin fluctuation in yeast - as in all eukaryotic cells - relies upon such diverse processes as regulated transcription of cyclin genes, constitutive or controlled

protein degradation and specific inactivation of Clb-CDKs [24]. To focus on the control mechanisms, we will now describe in detail the system governing mitosis as shown in figure 1.

2.2 Controlling mitosis

The two main regulatory tasks to be accomplished at the end of the cell cycle are first to activate late mitotic CDK, especially Clb2-CDK to drive mitotic events. As this activity also blocks the exit of mitosis, the cell secondly has to get rid of the mitotic CDK activities for a new cell cycle round to begin. These processes are mainly regulated by transcriptional control and by selective protein degradation: Transcriptional activation of the CLB2 gene cluster including the genes *CLB1*, *CLB2*, *CDC20* and *SWI5* [30] requires cyclins like Clb5 present in early mitosis. The associated kinases presumably initiate transcription via phosphorylation of the Mcm1 / Fkh2 / Ndd1 transcriptional effector complex [7]. This seems to be an auto-regulatory loop as *CLB2* transcription is abolished in cells lacking Clb1-4/CDK activity [3]. S phase CDKs furthermore keep up the phosphorylation of Sic1 thereby targeting it for degradation by the SCF (Skp1 / Cdc53 / F-box complex) proteasome pathway. They also inactivate the proteolytic machinery responsible for Clb2 degradation: Hct1 recognizes Clb2 protein and enables its ubiquitylation by the anaphase promoting complex (APC) and thus degradation by the proteasome. As Hct1 phosphorylation carried out by Clb5-CDK abolishes the Hct1-APC interaction [38], the kinase inhibits the pathway. All three mechanisms together allow for the accumulation of Clb2 protein in mitosis. All three functions are also fulfilled by late mitotic kinases, which furthermore are supposed to suppress SIC1 transcription via modification of transcription factor Swi5. Thereby, a self-sustaining state of high mitotic kinase activity establishes.

In anaphase, the Hct1 homologue termed Cdc20 is responsible for abolishing sister chromatid cohesion via destruction of the anaphase inhibitor Pds1. Concomitantly, active APC^{Cdc20} leads to partial degradation of Clb2 cyclin, which proved a necessary precondition for subsequent total elimination of late mitotic kinase activity [36, 5]. Both processes are subject to control by a checkpoint for assembly of the mitotic spindle. The spindle depolymerizing agent nocodazole (noc) accordingly prevents anaphase entry.

The next step of Clb2 inactivation begins after the chromosomes have been distributed to mother and daughter cell, which is equivalent to a correct positioning of the elongated spindle. The phosphatase Cdc14 plays a prominent role in the exit from mitosis, as it reverses inhibitory phosphorylation caused by mitotic CDK activity [32, 18]. The competitive inhibitor Net1 retains Cdc14 in the nucleolus until a poorly understood signal for correct spindle position transduced by the mitotic exit network (MEN) enables its release during telophase and early G1 phase [29]. Cell cycle progression thus depends on a further checkpoint and can not occur in case of non-functional MEN proteins. The exact mechanism leading to release of Cdc14 from the nucleolus has not been determined yet. Nevertheless, experimental evidence suggests that Net1 phosphorylation by mitotic exit network kinases plays a role in this process [23]. The counter-current reaction(s) can currently only be speculated upon. The ef-

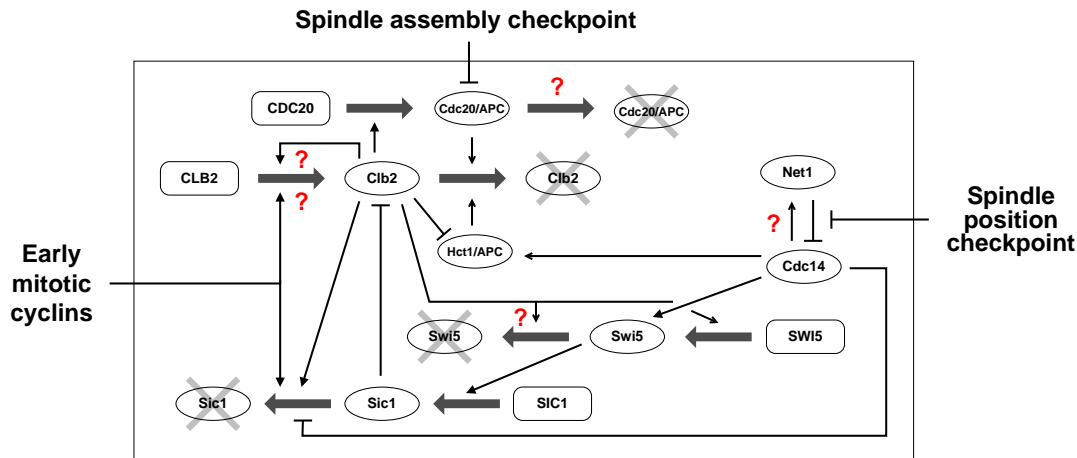


Figure 1: Molecular interactions governing mitosis. Regulatory interactions between genes (rectangles) and proteins (ellipses) are symbolized by thin lines for either positive (arrowhead) or negative (block) influences. Bold arrows stand for the processes of transcription / translation and protein degradation. Regulatory input signals from cell cycle checkpoints or from G1 control circuits are indicated accordingly. Question marks point to uncertain or hypothetical interactions. The subsystem for control of sister chromatid cohesion by Cdc20, Pds1 and Esp1 is not included.

fect of Cdc14 activation is at least threefold: Cdc14 removes the inhibitory phosphorylation from Hct1 and as a consequence switches on Clb2 proteolysis. The phosphatase also stabilizes Sic1 protein and allows nuclear entry of Swi5, a transcriptional activator of SIC1 [32]. These mechanisms jointly lead to bringing the cell into a G1 state, which is characterized by low Clb2-CDK activity and high Sic1 levels.

Due to the existence of parallel pathways for controlling late mitotic CDK activity, the system behaves quite robust. Deletion of either of the two major counter-players of CLB2 - HCT1 and SIC1 - has little effect on the qualitative behavior. Yeast cells are still able to exit mitosis, albeit either Clb2 levels stay high in G1, or inactivation relies completely on Clb2 degradation, respectively. Important other effects of perturbations in the system on the phenotype are summarized in table 1. In quantitative terms, however, these changes matter, as for instance *hct1Δ* strains display a delay in completing mitosis and therefore grow more slowly than wild type.

Table 1: Genotype - phenotype relations

| Genotype and additional treatment | Clb2 active in M | Cdc14 release | Clb2 degradation | Sic1 protein |
|-----------------------------------|------------------|---------------|------------------|--------------|
| wild type | + | + | + | + |
| wild type + nocodazole | + | - | - | - |
| men (mitotic exit network mutant) | + | - | - | - |
| <i>hct1Δ</i> | + | + | - | + |
| <i>sic1Δ</i> | + | + | + | - |

In addition to the regulatory interactions for mitotic control already described, several open questions exist. Question marks in figure 1 indicate some of these areas. Besides the

interaction of Net1 and Cdc14, especially the exact mechanisms for transcriptional control of the CLB2 gene cluster are currently unclear. Again, this regulation seems to involve redundancy, namely at least activation by mitotic cyclin-dependent kinases as well as regulated accumulation of the transcription factors themselves [13]. The degradation of some of the proteins like Cdc20 [28] and Swi5 requires more exact knowledge, too. These uncertainties make model development for the system somehow difficult, but in turn mathematical modeling could provide clues on the network function, if being successful.

3. A MITOSIS CONTROL MODULE

3.1 Modular modeling approach

Cellular systems composed of modules offer the possibility to develop mathematical models for each module separately and then to obtain the system model by connecting the modules. Our modular approach furthermore tries to implement re-usable model entities that enable a mathematical description, which is close to the biological processes. These principles are applied at the most elementary level of biological regulation by developing and employing submodels for processes like transcription, translation, enzymatic conversions, protein-protein-interactions and others. These submodels and higher aggregated models are then used to build up a standardized library for the convenient modeling of cellular systems [19, 31].

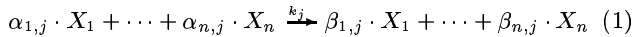
As an example for these submodels, we want to discuss a gene expression unit used for the mitotic control module. The biochemical reaction scheme shown in table 2 represents a simplified description for this functional unit. It comprises the binding of a transcription factor to DNA. This complex serves as the basis for formation of the initiation complex with RNA polymerase and finally for mRNA production. Translation is captured accordingly with ribosomes directly binding to mRNA. Degradation pathways for mRNA and protein (via the proteasome) complete the module. Captur-

| | | | | | |
|--|----------------------------------|---|------------------------------------|--------------------------|---|
| transcription factor binding to promoter | $[TF_j] + [gene_i]$ | $\xrightarrow[\xrightarrow{K_{D,TF_j}}]{K_{XA,TF_j}}$ | $[TF_j \cdot gene_i]$ | | |
| transcription | $[TF_j \cdot gene_i] + [RNAPol]$ | $\xrightarrow[\xrightarrow{K_{DPol}}]{K_{APol}}$ | $[TF_j \cdot gene_i \cdot RNAPol]$ | $\xrightarrow{K_{TK}}$ | $[mRNA_i] + [TF_j \cdot gene_i] + [RNAPol]$ |
| translation | $[mRNA_i] + [ribosome]$ | $\xrightarrow[\xrightarrow{K_{DRib}}]{K_{ARib}}$ | $[mRNA_i \cdot ribosome]$ | $\xrightarrow{K_{TL}}$ | $[protein_i] + [mRNA_i] + [ribosome]$ |
| mRNA degradation | | | $[mRNA_i]$ | $\xrightarrow{K_{DM_i}}$ | $[\]$ |
| | | | $[mRNA_i \cdot ribosome]$ | $\xrightarrow{K_{DM_i}}$ | $[ribosome]$ |
| proteolysis via the proteasome | $[protein_i] + [proteasome]$ | $\xrightarrow[\xrightarrow{K_{D,PR_i}}]{K_{A,PR_i}}$ | $[protein_i \cdot proteasome]$ | $\xrightarrow{K_{P,PR}}$ | $[\] + [proteasome]$ |

Table 2: Simplified biochemical reaction network for the description of a gene expression unit.

ing the processes at this level of detail has the consequence of models, which seem to be more complex. However, it firstly allows to integrate knowledge on well-characterized processes (i.e. experimentally determined transcription and translation rates). Secondly, parameter number and parameter search space are reduced and physically meaningful, as for instance only one parameter for the DNA-transcription factor interaction and the transcription factor-RNA polymerase interaction is used. Both features facilitate estimation of the remaining parameters. Also the interplay of global and local control is considered, which proves important for a system-wide perspective [31].

To derive the mathematical equations, the entire biochemical reaction system with n molecular species X_i is decomposed into r elementary reactions $j = 1 \dots r$ of the form



Here, k_j is the kinetic constant of the reaction, whereas $\alpha_{i,j}$ and $\beta_{i,j}$ are the stoichiometric coefficients for substance X_i 's involvement as educt or as product, respectively. The mathematical equations (ODEs) for the description of the reaction dynamics are derived straightforwardly from this scheme by application of mass action kinetics. In a canonical form like the one proposed in [27], the differential equations for the concentrations of the components c_i ($i = 1 \dots n$) write as:

$$\frac{dc_i}{dt} = \sum_{j=1}^r k_j \cdot (\beta_{i,j} - \alpha_{i,j}) \cdot \prod_{l \in S_j} c_l^{\alpha_{l,j}} \quad (2)$$

with S_j being the set of species actually participating in reaction j as educts, i.e. $\alpha_{l,j} > 0 \forall l \in S_j$.

We refer to the elementary chemical reactions instead of formal kinetics like Michaelis-Menten rate laws for two major reasons: The latter kinetics imply assumptions on relative velocities for the elementary reactions, which beforehand, for a system with largely unknown parameter values, may not be applicable. Additionally, the system of ODEs given in eqn. 2 is easily amendable to automatic model generation and model analysis, for instance to obtain the partial derivatives needed for sensitivity analysis (see section 4.1).

3.2 Model structure and data basis

One aim of this study was to establish a mathematical model for the control of mitosis in budding yeast, which is as close

as possible to the biological mode of regulation. We therefore included the well-established biological facts described in section 2.2 in development of the model structure. However, several assumptions needed to be made regarding the incompletely characterized interactions mentioned above.

The most important of these interactions concern transcription and proteolysis: The Mcm1 / Fkh2 / Ndd1 complex, necessary for expression of genes in the CLB2 cluster, is by itself transcriptionally and post-translationally regulated. For the model, we assume, that an activating phosphorylation has to be carried out by mitotic kinases. For production of the complex, a constant level of mRNA is included in the model, as it has been shown that oscillations of neither Fkh2 mRNA nor Ndd1 mRNA are required for proper transcriptional control of the CLB2 gene cluster [22, 13]. Furthermore, the questions of how exactly Cdc20, Swi5 and (parts of) the Mcm1 / Fkh2 / Ndd1 complex are degraded is unresolved. We tested several hypotheses regarding this subject. We came to the conclusion, that proteolysis could be carried out constitutively and / or via the APC with the help of the recognition factor Hct1 (see section 3.3). These mechanisms were implemented as possible degradation pathways.

Altogether, the model comprises 4 gene expression units for CLB2, CDC20, SWI5 and SIC1. Translation is described additionally for Clb5 mRNA and the mRNA representative of the Mcm1 / Fkh2 / Ndd1 complex. Together with the general components (RNA polymerase, ribosomes, APC and SCF) and proteins at constant level (Hct1 and the kinase Dbf2 representative of the MEN kinases), the dynamics of 13 proteins is described. Due to the high network interconnectivity, the representation of altogether 169 reactions finally results in a model with 171 state variables and 114 kinetic parameters.

Only a very limited number of model parameters like average translation rates [14] and specific mRNA half-lives [17] could be taken directly from literature. Parameter estimation then faced the problem - as encountered by the Tyson study [9] - of rare quantitative data on yeast cell cycle regulators. We therefore used the data from the collection of strains and culture conditions summarized in table 1 as data basis. All the data discussed below can be found in figure 2.

The central quantitative experimental data for Clb2 protein concentration, the corresponding cyclin-dependent kinase activity and MEN activity in yeast cells was taken from

the study by Fesquet *et al.* [12, Fig. 1]. Relative values were used to calculate absolute concentrations by estimating the order of magnitude of the protein concentrations to ≈ 100 nM [9]. As a conservative estimate of the error in these variables, we assumed a uniform standard deviation of $\pm 20\%$ of the peak value. The Clb2 concentration in mitotic exit mutants declines in anaphase to $\approx 50\%$ of wild type maximum [36, 5], and a similar behavior can be expected for the *hct1* Δ strain. However, in cells arrested by nocodazole the level increases with a slope of $\approx 0.05 \text{ h}^{-1}$ [26, Fig. 7]. For all strains capable of Cdc14 activation, it was furthermore demanded, that $50 \pm 10\%$ of the nucleolar Cdc14 be released in telophase [37]. Hct1 should be phosphorylated until the mitotic exit network is active and then be completely unphosphorylated in G1 phase. Additionally, it was assumed that in G1 phase $\approx 50 \pm 20 \text{ nM}$ Sic1 protein is present in wild type and *hct1* Δ cells. This assumption can be justified by the fact that equal amounts of this protein were observed *in vivo* [W. Seufert, personal communication], which have to be able to suppress Clb2-CDK activity in *hct1* Δ .

Simulation furthermore requires definition of the initial conditions. Starting values for mRNA concentrations were calculated based on the relative data given in [20, Fig. 7] using the absolute mRNA amounts in asynchronous cell populations reported in [17]. We further made the simplifying assumption, that in (late) S phase, from which on the model is supposed to be valid, Sic1 protein has already been degraded and Hct1 is completely inactivated. On this data basis, an evolutionary strategy was used to estimate the model parameters. Although this type of optimization algorithm less likely gets stuck in local optima than gradient-based methods, we used different starting guesses for the parameters to enhance convergence toward a globally optimal parameter set.

3.3 Assessing model quality

The first step in assessing the model quality, its descriptive and predictive character, is to determine in how far the simulations fit the experimental data used for estimating the model parameters. This comparison (figure 2) shows that the model is able to describe the qualitative behavior of the wild type, of the genetically modified strains and of the nocodazole-arrested cells. Given the experiences with modeling other complex regulatory networks [35], this gives a first hint to a basically adequate model structure. It can also be seen, that model and experiment quantitatively match well for the wild type and all the variants. Time courses of variables, for which no experimental data could be provided, show a more ambiguous picture. The transcription factor Swi5, for example, is shifted from its phosphorylated to its unphosphorylated form at the end of mitosis in agreement with the observations made *in vivo* [1]. The periodic post-translational activation of the complex responsible for CLB2 transcription appears correct as well. It has to be noted, that in both cases the overall behavior, for instance Sic1 accumulation, could also have been established by alternative mechanisms, e.g. enhanced Swi5 mRNA production. However certain flaws, indicative of a requirement for additional quantitative data, exist in the high Sic1 protein levels (steady state level of $\approx 200 \text{ nM}$, data not shown) for the *hct1* Δ strain in G1.

Additional evidence supporting reasonable model quality is revealed by details of the dynamics. To provide an example, time courses of Clb2-CDK activity, which can be used as indicator for the timing of mitotic events, are reproduced in figure 4. A closer look at the kinetics of Clb2-CDK inactivation at the end of mitosis in wild type, *sic1* Δ and *hct1* Δ reveals that the first two strains reach a low level of kinase activity comparably fast. The deletion of HCT1 in contrast has the consequence of a slower inactivation of late mitotic kinase, also due to a higher Clb2 level during mitosis. This finding agrees well with the delay in mitotic exit observed *in vivo* for this mutant [26]. It has not been pre-determined by the parameter estimation procedure because for the mutants the exact timing of low kinase activity was less constrained than for wild type (see figure 2).

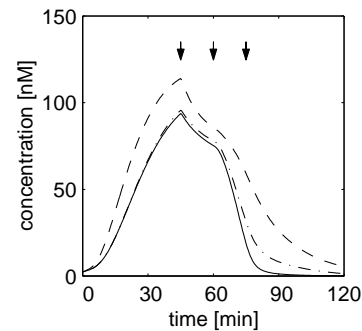


Figure 4: Mitotic timing. Active Clb2-CDK concentration is shown for wild type (—), *hct1* Δ strain (---) and the *sic1* Δ mutant (- · -). For further detail see the legend to figure 2.

Another criterion for model evaluation is, in how far estimated parameter values match with experimentally determined kinetic constants. From the comparison of selected parameters shown in table 3, the following conclusions can be drawn: First of all, the values rarely correspond as well as in the case of the inhibition constant for the Net1-Cdc14 interaction. However, given the range of parameter values opened up in the estimation step, the matching within approx. one order of magnitude seems acceptable. More important, the relative values of parameters in reactions describing counter-acting regulatory mechanisms - for instance the relation between Cdc14 activity and Clb2-CDK activity - fit surprisingly well. A judgment on whether the deviations in absolute terms are due to the robustness of the network, or result for example from inaccurate estimates of absolute values for the experimental data, has to await the quantitative parameter analysis to be described in section 4.2.

Whereas so far the interpretation of model quality was based on the set of experiments used for parameter estimation, we now want to turn to the decisive test-cases of true model predictions. As a selection of simulations corresponding to experiments with known outcome, the behavior of some deletion mutants is compiled in figure 3. The predictions include telophase arrest of a *hct1* Δ *sic1* Δ double mutant [34] and impaired activation of late mitotic kinases in the NET1 deletion strain [33]. Deletion of the phosphatase Cdc14 in contrast allows for mitotic entry, but entry into the next cell cycle phase is impeded by Hct1 staying in its inactive

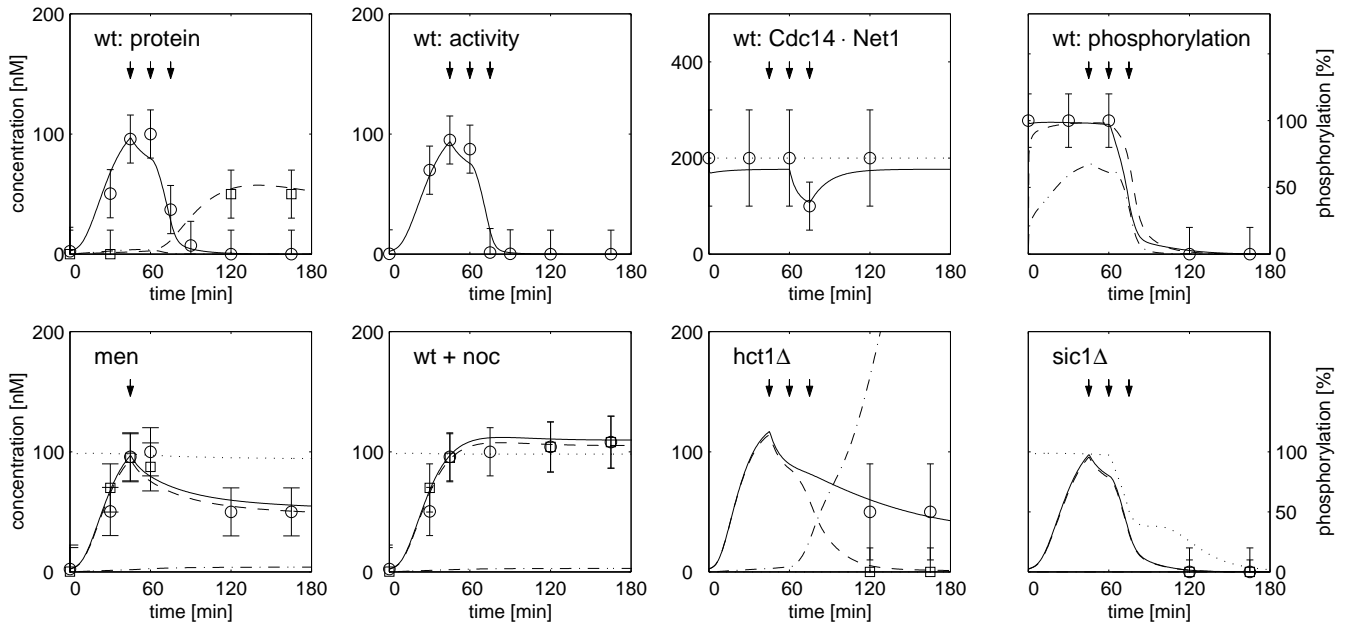


Figure 2: Experimental data and simulation results for wild type and mutants used in parameter estimation. The **upper panels** show experimental data (symbols) and simulation results (lines) for selected variables in wild type. Time $t = 0$ corresponds to (late) S phase. At the time points indicated by the arrows, the spindle assembly checkpoint is inactivated (45 min), and MEN activity is switched on (60 min) or off (75 min), respectively. The variables are from left to right: (1) CLB2 (—,○) and Sic1 (—,□) protein concentration, (2) active Clb2-CDK concentration (—,○), (3) total Cdc14 protein (···) and nucleolar Cdc14 (—,○), plus (4) the phosphorylation states of Hct1 (—,○), Swi5 (—) and Mcm1 / Fkh2 / Ndd1 (—·—). In the **lower panels** the mutant behavior is summarized. For each condition, the time courses of total Clb2 protein (—,○), Clb2-CDK activity (—,□), Sic1 protein concentration (—·—) and Hct1 phosphorylation state (···) are given.

Table 3: Comparison of estimated and experimentally determined kinetic parameters

| Kinetic constant | Estimation range | Estimated | Experimental | Ref. |
|---|----------------------------------|---------------------------------|-------------------------------------|------|
| K_I for Cdc14 inhibition by Net1 | $10^{-2} - 10^3$ nM | 1.32 nM | ≈ 3 nM | [23] |
| K_D for interaction of unphosphorylated Hct1 with APC | $10^{-2} - 10^3$ nM | 0.45 nM | ≈ 5 nM | [18] |
| K_D for interaction of phosphorylated Hct1 with APC | $10^{-2} - 10^3$ nM | 21.54 nM | $\approx 50-100$ nM | [18] |
| K_{cat} of Clb2-Cdc28 kinase ^a | $10^{-3} - 10^3$ s ⁻¹ | 24.5 ± 15.4 s ⁻¹ | ≈ 2.2 s ⁻¹ | [6] |
| K_{cat} of Cdc14 phosphatase ^b | $10^{-3} - 10^3$ s ⁻¹ | 0.44 ± 0.37 s ⁻¹ | $\approx 0.01-0.02$ s ⁻¹ | [32] |

^a Estimated values (mean \pm standard deviation for all substrates included in the model) are compared to values for the co-substrate ATP obtained for histone phosphorylation at 23° C.

^b Estimated values (mean \pm s.d. for all substrates included in the model) are related to Swi5 and Hct1 dephosphorylation activity *in vitro* at 30° C.

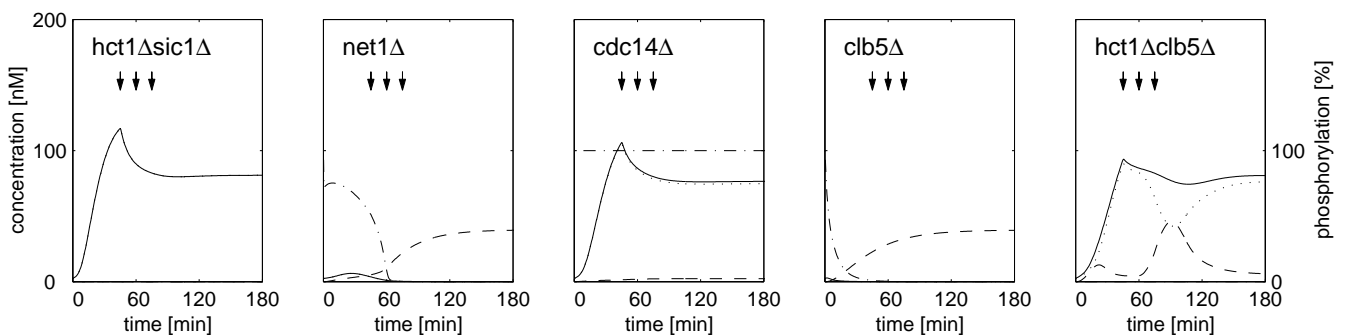


Figure 3: Model predictions for deletion mutants. Symbols and variables for the time courses displayed are the same as those in the lower panels of figure 2.

phosphorylated state and by failure in Sic1 protein accumulation [32]. The influence of Clb5 and related cyclins on the expression of late mitotic cyclin Clb2 was also investigated. The simulation results show no Clb2 accumulation in the *clb5Δ* strain, but functional mitotic entry upon additional deletion of the HCT1 gene. The exit from mitosis is nevertheless impaired in the double mutant. Sic1 accumulates and represses Clb2-CDK activity, but this state turns out to be unstable. The results are supported by experimental observations which describe the cells as able to carry out mitosis but having a reduced viability [37]. These predictions are not obvious since theoretically the low Clb2-CDK activity present in S phase could drive Clb2 protein accumulation via the positive auto-regulation of CLB2 transcription without the initial help of S phase cyclin-dependent kinase.

We finally want to point out that albeit a large mathematical model is considered, the model structure proved to be essential for the ability to describe biological reality correctly. As an example, simulation results for two model structures slightly differing in the mechanisms of Swi5 protein degradation are compared (figure 5). The first structure, used in all previous simulations, provides constitutive proteolysis as well as controlled degradation via APC^{Hct1}. In the second alternative, the APC-dependent pathway is knocked out. Repeated, independent parameter estimations were performed for both model structures. All experiments in figure 2 with the exception of the *hct1Δ* strain could be described with both models in comparable quality. The system however seems to depend on regulated proteolysis, which is likely to be carried out via APC^{Hct1}: Enhanced transcription of Sic1 backs up the failure in Clb2 protein degradation in *hct1Δ* and thus makes the module more fault-tolerant. Mathematical modeling thereby gave a clue on internal working principles of the mitosis control module, which can now be tested experimentally.

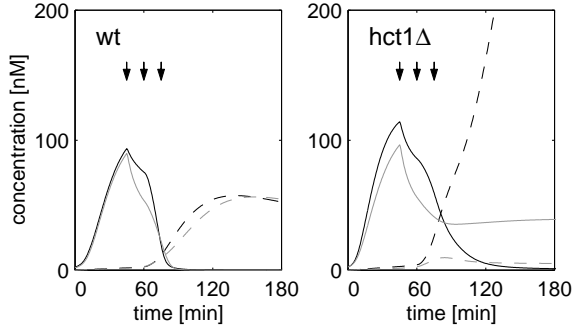


Figure 5: Comparison of model structures. Simulation results - Clb2-CDK activity (—) and Sic1 protein concentration (---) - are compared for two, slightly different model structures: The 'optimal' structure with constitutive Swi5 proteolysis plus Swi5 degradation via APC^{Hct1} (black lines) and a variant with constitutive proteolysis only (grey lines).

Altogether, the model for mitotic control was tested in several, independent ways. Since all tests gave satisfying results, the conclusion of having developed a model with descriptive and predictive character is so far justified. Further, quantitative examination of the models' robustness

and identifiability based on sensitivity analysis is described in the following section.

4. ROBUSTNESS AND IDENTIFIABILITY

4.1 Foundations of parameter analysis

In a general form, the mathematical model is formulated as a system of ODEs $\mathbf{S}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t)$ depending on the state variables \mathbf{x} (corresponding to the concentration vector \mathbf{c} in eqn. 2), the inputs \mathbf{u} , and the model parameters \mathbf{p} as

$$\mathbf{S}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t) = \dot{\mathbf{x}} - \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t) = \mathbf{0} \quad (3)$$

with the time variable $t \geq t_0$, the functional for the right hand sides $\mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t)$, and the initial conditions $\mathbf{x}(t_0) = \mathbf{x}_0$.

The estimation of model parameters basically relies on the comparison of model behavior and the real system's behavior. For this purpose, an identification functional $\Phi(\mathbf{p})$ has to be minimized by variation of parameter values. The formulation of this functional in turn needs the error for each out of N measurements \mathbf{e}_i , which is the difference between measurement $\mathbf{x}^M(t_i, \mathbf{u})$ and model $\mathbf{x}(t_i, \mathbf{u}, \mathbf{p})$ at a discrete time point t_i . The differences are furthermore weighted by the inverse of the standard deviation of the measurement \mathbf{Q}_i . The entire optimization problem thus reads

$$\Phi(\mathbf{p}) = \sum_{i=1}^N \left[\mathbf{e}_i^T \cdot \mathbf{Q}_i \cdot \mathbf{e}_i \right] \rightarrow \min$$

$$\text{with } \mathbf{e}_i = \mathbf{e}_i(t_i, \mathbf{u}, \mathbf{p}) = \mathbf{x}(t_i, \mathbf{u}, \mathbf{p}) - \mathbf{x}^M(t_i, \mathbf{u}). \quad (4)$$

Parameter estimation leads to a set of optimal parameters \mathbf{p}^* . The state sensitivities - the partial derivatives of \mathbf{x} with respect to \mathbf{p} - are derived from the variation equation

$$\frac{\partial}{\partial t} \frac{\partial \mathbf{x}}{\partial \mathbf{p}} \Big|_{\mathbf{p}^*, t_i} + \mathbf{J}_x \cdot \frac{\partial \mathbf{x}}{\partial \mathbf{p}} \Big|_{\mathbf{p}^*, t_i} = \mathbf{J}_p, \quad (5)$$

whereby the Jacobians \mathbf{J}_x for the states and \mathbf{J}_p for the parameters are given by

$$\mathbf{J}_x = \frac{\partial \mathbf{S}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t)}{\partial \mathbf{x}} \Big|_{\mathbf{p}^*, t_i}, \quad \mathbf{J}_p = \frac{\partial \mathbf{S}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t)}{\partial \mathbf{p}} \Big|_{\mathbf{p}^*, t_i}. \quad (6)$$

It has to be underlined that the sensitivities are time dependent and local with respect to parameter space.

The Fisher information matrix $\mathbf{F}(\mathbf{p}^*)$ links model and experiment via the sensitivities and the inverse of the measurement covariance matrix $\mathbf{C}(t_i)^{-1}$, respectively by

$$\mathbf{F}(\mathbf{p}^*) = \sum_{i=1}^N \left[\frac{\partial \mathbf{x}}{\partial \mathbf{p}} \Big|_{\mathbf{p}^*, t_i}^T \mathbf{C}^{-1}(t_i) \frac{\partial \mathbf{x}}{\partial \mathbf{p}} \Big|_{\mathbf{p}^*, t_i} \right] \quad (7)$$

A lower bound for the estimation accuracy of parameter j - σ_j - can then easily be derived from the corresponding eigenvalue λ_j of the Fisher information matrix:

$$\sigma_j \geq 100 \cdot \sqrt{\frac{1}{\lambda_j}} \quad [\%]. \quad (8)$$

A more detailed description of the procedure and its theoretical foundations is for example provided in [10].

4.2 Model analysis

Parameter estimation accuracy σ quantitatively captures the sensitivity of the system's behavior towards variances in the (optimal) parameters. It can therefore be used to determine the system's identifiability on the basis of the given data. The inverses of the estimation accuracies σ^{-1} of the mitotic control module calculated with the same experimental data as in the parameter optimization are shown in figure 6. Bars indicate an estimation accuracy of $\leq 99\%$, whereby higher values stand for better estimates. About half of the parameters could be estimated within a margin of $\pm 90\%$, i.e. within less than two orders of magnitude. These about 60 parameters were determined with experimental data available for only five variables in wild type and for two variables in each of the other conditions without any further experiment. To recall, the model itself has 171 state variables and 114 parameters. The experimental data basis thus carried much more information on the system than is expected from simply comparing data quantity to model complexity. The remaining parameters either require more experiments to be determined, or they can generally not be estimated, because the reactions they are used to describe are insignificant for the system. The latter case could again be identified by the same procedure of calculating the parameter estimation accuracies.

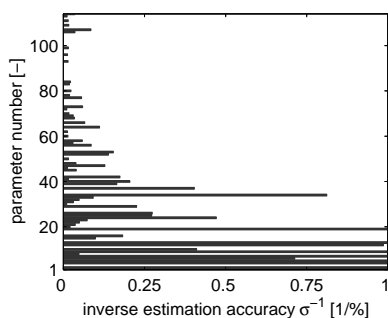


Figure 6: Parameter estimation accuracy. The inverse of the parameter estimation accuracy σ is plotted for each of the model parameters.

It has furthermore to be noted, that within the set of parameters, the estimation accuracies vary to a large extent, but the variation is not random. Some of the parameters reach an estimation accuracy of $\leq 1\%$, i.e. they are highly sensitive. These parameters seem to be clustered, as the parameter numbers are ordered according to the genes and proteins included in the network: Most of the highly sensitive parameters are involved in the description of Cdc14 (parameter numbers 1-21) and of Clb2-CDK (22-57). High sensitivity, however is not linked to all of the interactions of these proteins, but only to specific ones, for example the biologically essential inhibition of Cdc14 by Net1. The same holds true for the parameters, which could not be estimated within the given accuracy range. For the system's robustness these findings provide first evidence for a concept of robustness, in which specific parts of the system are extremely sensitive and thus can lead to overall fragility.

For the practical purpose of conducting experiments to identify a regulatory module, the parameter estimation quality

gives clues on the information content of the experiments. In figure 7, upper panel, the cumulative distribution of the number of parameters with a given accuracy in terms of quantitatively describing the measurements is shown as a function of this (critical) accuracy. Parameters with an accuracy of $> 99\%$ were not considered, so that this distribution does not reach the value of one. The distribution for the case, in which wild type behavior is considered exclusively, shows a maximum of about 30% of the parameters, many of them accumulating towards lower levels of accuracy. If the experimental information for wild type is augmented by the data for mutants and other disturbed operation modes of the module, the number of parameters identified as well as the quality of identification increase considerably, although for the disturbed modes data for less variables was available. For identification of the mitotic control module thus to include more mutants in parameter estimation is a promising way to increase model quality and to allow for complete model verification.

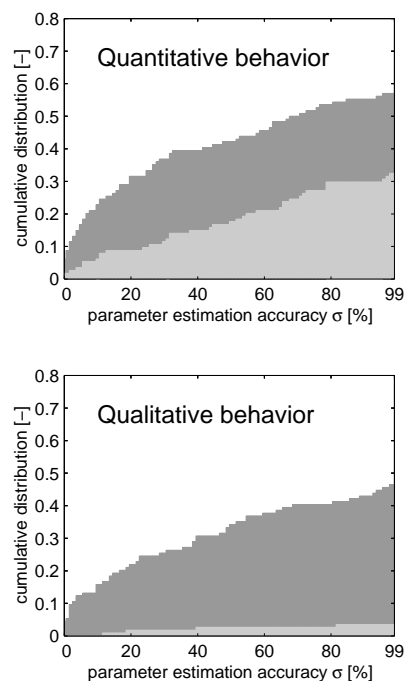


Figure 7: Parameter estimation accuracy as indicator for network robustness. The cumulative distributions of parameter estimation accuracies are given for the quantitative behavior in terms of compliance with the measurement (upper panel) and for the qualitative behavior as defined in the text (lower panel). Dark grey stands for wild type and all the mutants implied in parameter estimation, whereas light grey is for wild type only.

With the same type of mathematical instruments used to examine the model's identifiability, also its robustness can be analyzed. Robustness is defined as the (relative) insensitivity of the system's qualitative behavior in case of changes in the kinetic parameters. An investigation of robustness accordingly requires the specification of the qualitative behavior.

In case of the control of mitosis in budding yeast cell cycle regulation, firstly high mitotic CDK activity is necessary for the initiation of mitosis. A quantitative estimate leads to $\approx 100 \pm 50\%$ of maximal wild type activity. The lower bound can be deduced from the fact that *ndd1* mutants, which display Clb2 oscillation at a lower level, are nevertheless viable [22]. This activity should furthermore appear timely, i.e. at the time of the peak in wild type. Only residual kinase activity can secondly be allowed for the entry into G1 phase. Here we set a threshold of $\approx 0 + 20\%$ of maximal wild type activity for the network to be judged functional in qualitative terms. Again, the suppression is to be in time and additionally to be stable over the time horizon regarded in the simulation, as the model lacks of the G1/S control mechanisms.

The distribution of parameter estimation accuracies for the so defined qualitative behavior of the control circuit can be found in the lower panel of figure 7. If only the wild type is considered, the network behaves very robust. About 4% of the parameters can be identified within two orders of magnitude, which in turn means that the behavior is almost insensitive to changes in the parameters. The opposing features of robustness and identifiability become clear from these data. However, also for the network with all redundant pathways operational, some sensitive spots were identified. Robustness thus co-exists with fragility.

For the qualitative behavior of the wild type and all the disturbed variants, the picture changes completely. Even with the very reduced qualitative behavior, the model identification is comparable to the identification obtained by including all quantitative data. Two conclusions are obvious from this fact: First of all, the robustness of the mitotic control module relies on the existence of redundant pathways, as by elimination of this redundancy, the control circuit becomes sensitive to disturbances even in its qualitative behavior. The quantitative identification of the internal working principles for this control module, secondly, does not require extensive quantitative data: Few measurements for the system in different operation modes provide nearly as much information as an extended collection of time courses.

5. CONCLUSIONS

In order to examine the relation of robustness and identifiability of regulatory modules, we concentrated on the control of mitotic events in budding yeast cell cycle regulation. A mathematical model for this regulatory module was developed using a detailed, modular modeling approach, which aims at a close representation of the mechanisms operating in the biological system. Compared to the model established by Tyson and co-workers [9], our model's resolution is higher for this subsystem, as for instance the network was extended to cover control of mitotic exit via the Cdc14 phosphatase. By estimating the model parameters from experimental data with only few assumptions on absolute concentrations, we intended to narrow the usual gaps between model and experiment [16] as well. Thorough model analysis strongly argues for this to be a successful approach, as according to different criteria, the model matches well with experimental findings. Furthermore, the model allowed for specifying barely characterized regulatory interactions like a probable involvement of APC^{Hct1} in the control of the transcription factor Swi5.

In the Tyson model [9], a direct activation of Swi5 by Cdc20 is assumed, for which to our knowledge neither experimental nor mechanistic evidence exists. This again underlines the need for a more detailed modeling approach.

During model validation we determined parameter sensitivities and parameter estimation accuracy. Sensitivity analysis was up to now applied only for smaller genetic networks [11], whereas calculation of the estimation accuracies was rarely performed. The results obtained for the mitotic control model however suggest a more general value of these methods in mathematical modeling of cellular systems. First of all, determination of parameter estimation accuracy showed that the information content of the scarce quantitative data available was much higher than generally expected. For our complex model, about half of the 114 kinetic parameters could be estimated based on time courses for only five variables. Analysis also proved, that the combination of data from the unperturbed regulatory module in wild type with data from mutants or otherwise disturbed cells was essential for this outcome. As the method enables judgements on the 'best' (combination of) experiments, it is also applicable for experiment design [10]. Rapid, iterative improvement of experiment and model thus becomes possible.

Robustness of biochemical networks has to date mainly been determined via extensive numerical simulation. We propose parameter estimation accuracy additionally as a quantitative measure for robustness. For our model, the network clearly shows robust qualitative behavior, which is due to the operation of redundant pathways. There exist, however few spots of high sensitivity in the model. The analysis therefore strongly supports the concept of highly optimized tolerance, i.e. of the co-existence of robustness and fragility in cellular control as opposed to uniform insensitivity of the network. A differentiated view on robustness and identifiability of regulatory modules is thus required, but based on our experience, realistic models of cellular control are possible despite limited quantitative data. In general, these findings point to an up to now barely considered potential of fragmented, but quantitative knowledge for coming to a quantitative description of cellular systems.

6. ACKNOWLEDGMENTS

Financial support for part of this study by a "Peter und Traudl Engelhorn Stiftung" grant to J.S. is gratefully acknowledged.

7. REFERENCES

- [1] B. Aernie, A. Johnson, J. Toyn, and L. Johnston. Swi5 controls a novel wave of cyclin synthesis in late mitosis. *Mol. Biol. Cell*, 9:945–56, 1998.
- [2] U. Alon, M. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 397:168–71, 1999.
- [3] A. Amon, M. Tyers, B. Futcher, and K. Nasmyth. Mechanism that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins. *Cell*, 74:993–1007, 1993.
- [4] N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387:913–17, 1997.
- [5] M. Bäumer, G. Braus, and S. Irniger. Two different modes of cyclin Clb2 proteolysis during mitosis in *Saccharomyces cerevisiae*. *FEBS Lett.*, 468:142–48, 2000.

- [6] A. Bishop, J. Ubersax, D. Petsch, D. Matheos, N. Gray, J. Blethrow, E. Shimizu, J. Tsien, P. Schultz, M. Rose, J. Wood, D. Morgan, and K. Shokat. A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature*, 407:395–401, 2000.
- [7] L. Breeden. Cyclin transcription: timing is everything. *Curr. Biol.*, 10:R586–88, 2000.
- [8] J. Carlson and J. Doyle. Highly optimized tolerance: robustness and design in complex systems. *Phys. Rev. Lett.*, 84(11):2529–32, 2000.
- [9] K. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, and J. Tyson. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell*, 11:369–91, 2000.
- [10] A. Emery and A. Nenarokomov. Optimal experiment design. *Meas. Sci. Technol.*, 9(6):864–76, 1998.
- [11] R. Erb and G. Michaels. Sensitivity of biological models to errors in parameter estimates. *Pac. Symp. Biocomput.*, 4:53–64, 1999.
- [12] D. Fesquet, P. Fitzpatrick, A. Johnson, K. Kramer, and K. Johnston. A Bub2p-dependent spindle checkpoint pathway regulates the Dbf2p kinase in budding yeast. *EMBO J.*, 18(9):2424–34, 1999.
- [13] B. Futcher. Microarrays and cell cycle transcription in yeast. *Curr. Opin. Cell Biol.*, 12:710–15, 2000.
- [14] B. Futcher, G. Latter, P. Monardo, C. McLaughlin, and J. Garrels. A sampling of the yeast proteome. *Mol. Cell Biol.*, 19(11):7357–68, 1999.
- [15] L. Hartwell, J. Hopfield, S. Leibler, and A. Murray. From molecular to modular cell biology. *Nature*, 402 (Supp.):C47–C52, 1999.
- [16] J. Hastly, D. McMillen, and J. Collins. Computational studies of gene regulatory networks: *in numero* molecular biology. *Nat. Rev. Genet.*, 2:268–79, 2001.
- [17] F. Holstege, E. Jennings, J. Wyrick, T. Lee, C. Hengartner, M. Green, T. Golub, E. Lander, and R. Young. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, 95:717–28, 1998.
- [18] S. Jaspersen, J. Charles, and D. Morgan. Inhibitory phosphorylation of the APC regulator Hct1 is controlled by the kinase Cdc28 and the phosphatase Cdc14. *Curr. Biol.*, 9:227–36, 1999.
- [19] A. Kremling, K. Jahreis, J. Lengeler, and E. Gilles. The organization of metabolic reaction networks: A signal-oriented approach to cellular models. *Metabolic Eng.*, 2(3):190–200, 2000.
- [20] R. Kumar, D. Reynolds, A. Shevchenko, A. Shevchenko, S. Goldstone, and S. Dalton. Forkhead transcription factors, Fkh1p and Fkh2p, collaborate with Mcm1p to control transcription required for M-phase. *Curr. Biol.*, 10:896–906, 2000.
- [21] D. Lauffenburger. Cell signaling pathways as control modules: Complexity for simplicity? *Proc. Natl. Acad. Sci. USA*, 97(10):5031–33, 2000.
- [22] C. Loy, D. Lydall, and U. Surana. NDD1, a high-dosage suppressor of *cdc28-1N*, is essential for expression of a subset of late-S-phase-specific genes in *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, 19(5):3312–27, 1999.
- [23] A. Mah, J. Jang, and R. Deshaies. Protein kinase Cdc15 activates the Dbf2-Mob1 kinase complex. *Proc. Natl. Acad. Sci. USA*, 98(13):7325–30, 2001.
- [24] M. Mendenhall and A. Hodge. Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.*, 62(4):1191–1243, 1998.
- [25] C. Robert, J. Carlson, and J. Doyle. Highly optimized tolerance in epidemic models incorporating local optimization and regrowth. *Phys. Rev. E*, 63:056122, 2001.
- [26] M. Schwab, A. Schulze-Lutum, and W. Seufert. Yeast Hct1 is a regulator of Clb2 cyclin proteolysis. *Cell*, 90(4):683–93, 1997.
- [27] B. Shapiro, A. Levchenko, and E. Mjolsness. Automatic model generation for signal transduction with applications to MAP-kinase pathways. In *Proc. 1st International Conference on Systems Biology*, pages 64–74, Tokyo, 2000.
- [28] M. Shirayama, W. Zachariae, R. Ciosk, and K. Nasmyth. The Polo-like kinase Cdc5p and the WD-repeat protein Cdc20p/fizzy are regulators and substrates of the anaphase promoting complex in *Saccharomyces cerevisiae*. *EMBO J.*, 17(5):1336–49, 1998.
- [29] W. Shou, J. Seol, A. Shevchenko, C. Baskerville, D. Moazed, Z. Chen, J. Jang, A. Shevchenko, H. Charbonneau, and R. Deshaies. Exit from mitosis is triggered by Tem1-dependent release of the protein phosphatase Cdc14 from the nucleolar RENT complex. *Cell*, 97:233–44, 1999.
- [30] P. Spellman, G. Sherlock, M. Zhang, V. Iyer, K. Anders, M. Eisen, P. Brown, D. Botstein, and B. Futcher. Comprehensive identification of cell-cycle regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, 9:3273–97, 1998.
- [31] J. Stelling, A. Kremling, M. Ginkel, K. Bettenbrock, and E. Gilles. Towards a virtual biological laboratory. In H. Kitano, editor, *Foundations of systems biology*, chapter 5. MIT Press, *in press*, 2001.
- [32] R. Visintin, K. Craig, E. Hwang, S. Prinz, M. Tyers, and A. Amon. The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Mol. Cell*, 2:709–18, 1998.
- [33] R. Visintin, E. Hwang, and A. Amon. Cfi1 prevents premature exit from mitosis by anchoring Cdc14 phosphatase in the nucleolus. *Nature*, 398:818–23, 1999.
- [34] R. Visintin, S. Prinz, and A. Amon. CDC20 and CDH1: A family of substrate-specific activators of APC-dependent proteolysis. *Science*, 278(5337):460–63, 1997.
- [35] G. von Dassow, E. Meir, E. Munro, and G. Odell. The segment polarity network is a robust developmental module. *Nature*, 406:188–92, 2000.
- [36] F. Yeong, H. Lim, C. Padmashree, and U. Surana. Exit from mitosis in budding yeast: biphasic inactivation of the Cdc28-Clb2 mitotic kinase and the role of Cdc20. *Mol. Cell*, 5:501–11, 2000.
- [37] F. Yeong, J. Lim, Y. Wang, and U. Surana. Early expressed Clb proteins allow accumulation of mitotic cyclin by inactivating proteolytic machinery during S phase. *Mol. Cell Biol.*, 21(15):5071–81, 2001.
- [38] W. Zachariae, M. Schwab, K. Nasmyth, and W. Seufert. Control of cyclin ubiquitination by CDK-regulated binding of Hct1 to the anaphase promoting complex. *Science*, 282:1721–24, 1998.