

Inverse Bifurcation Analysis of a Model for the Mammalian G_1/S Regulatory Module

James Lu¹, Heinz W. Engl¹, Rainer Machné², and Peter Schuster²

¹ Johann Radon Institute for Computational and Applied Mathematics,
Austrian Academy of Sciences
Altenbergerstrasse 69, A-4040 Linz, Austria
james.lu@oeaw.ac.at, heinz.engl@oeaw.ac.at
² Theoretical Biochemistry Group,
Institute for Theoretical Chemistry, University of Vienna
Währingerstrasse 17, A-1090 Vienna, Austria
raim@tbi.univie.ac.at, pks@tbi.univie.ac.at

Abstract. Given a large, complex ordinary differential equation model of a gene regulatory network, relating its dynamical properties to its network structure is a challenging task. Biologically important questions include: what network components are responsible for the various dynamical behaviors that arise? can the underlying dynamical behavior be essentially attributed to a small number of modules? In this paper, we demonstrate that inverse bifurcation analysis can be used to address such *inverse problems*. We show that sparsity-promoting regularization strategies, in combination with numerical bifurcation analysis, can be used to identify small sets of "influential" submodules and parameters within a given network. In addition, hierarchical strategies can be used to generate parameter solutions of increasing cardinality of non-zero entries. We apply the proposed methods to analyze a model of the mammalian G_1/S regulatory module.

1 Biological Background

Properties emerging from complex dynamical systems such as *bistability* (the existence of two stable, steady states) or *oscillations* are getting increasing attention as potential design principles of cellular networks of metabolic, gene regulatory or signal transducing systems [1]. While such biological systems form large networks that are difficult to grasp, the theory of dynamical systems provides the means to categorize and reduce them to tangible and understandable core modules governing the overall system properties.

One of the most studied cell biological systems in this regard is the cell cycle. Via iterations of theoretical and experimental analysis it has been successfully analyzed as a modular system consisting of a core oscillator controlled by a series of bistable switches mediating the transition through checkpoints between the different cell cycle phases [2, 3, 4, 5]. While the individual players differ between

species, the overall design seems conserved enough to allow researchers to propose a generic cell cycle model that can be mapped onto diverse species [6].

The individual modules are responsible for coordinating each phase of the cell cycle with external resources, such as nutrients or - in a multicellular context - the diverse signaling factors supplied by the organism. The deciding factors include: has the current phase been successfully completed? do external resources still allow the onset of proliferation? The gene regulatory switches have to integrate diverse halt or go signals to form the appropriate decisions. In this methodological paper, we focus on a small part of this vast complexity: a simplified model of the G_1/S transition of the mammalian cell cycle that has recently been proposed by Swat *et al.* [7].

2 A Model of the G_1/S Transition

The model of the mammalian G_1/S transition consists of 9 chemical species, 25 reactions and 40 parameters, representing the transcription factor families AP-1, E2F, pRB, and cyclin/cyclin-dependent kinase complexes cyclin D/Cdk4,6 and cyclin E/Cdk2 [7]. Please see the original publication [7] and the thesis by M. Swat [8] for a description of the model; an SBML version of this model is available for download from [9]. The model may be mathematically expressed as a system of ordinary differential equations (ODEs). Denoting by x and α the biochemical concentrations and parameters respectively, the instantaneous change in $x \in \mathbb{R}^n$ is described by the parametrized vector field $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$:

$$\dot{x} = f(x, \alpha). \quad (1)$$

Depending on the parameter α , the parameterized ODE system (1) can exhibit different dynamical behavior. In performing a (forward) *bifurcation analysis*, one examines what different qualitative behavior can arise when various parameters are varied. The computed bifurcation diagrams contain information on the location of parameter regions where the solution behavior stays qualitatively the same, as well as parameter locations where the solution changes its character (see [10] for a mathematical overview of bifurcation analysis).

To understand the significance of bifurcation diagrams for the model system necessitates both a short explanation of the parameters and a rough outline of the biological context, especially the mitogenic stimulation of the G_1/S transition represented here by the bifurcation parameter F_m which is the main focus of our analysis. See Figure 1 for the correspondence between bifurcation points and cell phases. Further details of the model are discussed together with the results of inverse analysis in Section 5.

2.1 Bistable Core

While the model of Swat *et al.* is of a qualitative nature - trying to capture both the basic architecture of interactions and experimentally known dynamics - its underlying equations are frequently used in modeling of gene regulatory

networks allowing for straight-forward biological interpretation of results. The model is based on extensive usage of the Hill-Langmuir equation (see e.g. [11,12]) to describe binding equilibria of both the transcription factor E2F1 and its specific DNA binding sites, as well as the transcriptional repressor pRB and this E2F1:DNA complex. The parameters K_{mx} , and J_x (where x is a variable subscript) thus represent dissociation constants in units of concentration, except for the parameter K_{m2} , which falls outside this definition because of an additional factor a used in the core module encoding for the bistability. The parameters k_x represent simple kinetic rates in units of 1/time or 1/(concentration*time) and parameters ϕ_x represent degradation rates in units of 1/time. Maximal transcription rates k_x are then scaled by the temporal averages of the activated ($E2F1/(K_{mx} + E2F1)$) and uninhibited ($J_x/(J_x + pRB)$) gene regulatory sites, according to the Hill-Langmuir equation. The same equations are also used (although with less mechanistic justification) for cyclin autocatalytic activation processes.

The core feature of the model by Swat *et al.* is the autocatalytic activation of the transcription factor E2F1 by binding to its own regulatory site. Two independent binding sites are assumed, both of which need to be occupied for autocatalytic activation. This assumption leads to the necessary nonlinearity in the equation which is the main source of bistability in the model.

2.2 Mitogenic Stimulation

Swat *et al.* realize this classic module of transition through the restriction point and into the S phase of the cell cycle by assuming the following: E2F1 mediated activation of the transcription of all involved genes; a strong inhibition of this process by unphosphorylated pRB; a weakened inhibition by half-phosphorylated pRB (pRB_p); finally no inhibition by fully phosphorylated pRB (pRB_pp). These phosphorylations are sequentially catalyzed by cyclin D/Cdk4,6 and cyclin E/Cdk2 complexes.

The model starts in G_0/G_1 phase and transcription of cyclin D is initiated by a simple linear dependence on the transcription factor AP-1, which itself is modeled as a continuous responder to an input parameter F_m . This latter parameter represents the mitogenic stimulation and the whole model is designed to show bistable dependence on a continuous variation of this mitogenic stimulation parameter. The parameter is used here as the bifurcation parameter. How can biologists relate this to a real life context?

Diverse mitogenic stimuli, usually the growth factors, act on receptors in the cell membrane to activate a complex network of signal transduction. Receptors activate membrane and cytoskeletal modifications in feedback-based switches, with the Ras G protein family as central elements [13]. The pathways at the membrane branch into several intracellular processes, of which two are considered central for cell cycle regulation: intracellular calcium release [14] and the mitogen-activated protein kinase (MAPK) cascades [15] both of which show complex and diverse spatio-temporal variations of intracellular activity but

ultimately should converge again in the cellular nucleus to activate further gene transcription via the transcription factors NFAT and AP-1 [16,17].

Both pathways, calcium release and MAPK activation cascades, have also been intensively studied by theoretical modeling. While calcium increase can show complex oscillatory behavior [18], of which both the frequency and the amplitude might be important information carriers [19, 20, 21, 22, 23, 24], not so much is known about consequences of these spatio-temporal variations for the activation of the cell cycle [14, 25]. In contrast, temporal modulation of MAPK activity is increasingly recognized as the determining signal for cell fate decision [26, 15]. Transient, sustained weak or sustained strong activation have different consequences, either resulting in cell cycle arrest to induce cell differentiation or the onset of the cell cycle. Nuclear feedforward sensors for temporal activity of MAPK that mediate these differences have been identified [27]. From a theoretical perspective, the MAPK cascade has been found again to show bistable dependence on the strength of mitogenic stimulation [28, 29, 30]. Such a bistable, all-or-none switch can, for example, create a discrete border of responding and non-responding cells lined up within continuous concentration gradient of a signal and thus be important for the processes of developmental morphogenesis [31, 32].

Thus, the pathway of a mitogenic stimulus to the activation of nuclear transcription of cell cycle genes, as represented in this model by Swat *et al.* by the simple parameter F_m , involves again a series of bistable switches. However, diverse interacting pathways, be it metabolic or structural states of the cell or opposing signals, can modulate this behavior. In the case of the MAPK cascade's bistability it is known that increasing levels of counteracting phosphatases can render this bistable switch into a continuous responder to a mitogenic signal [33, 34], a process which is often regulated via negative feedback by the MAPK themselves [35]. In such a situation the downstream bistability of the G_1/S would receive increased control.

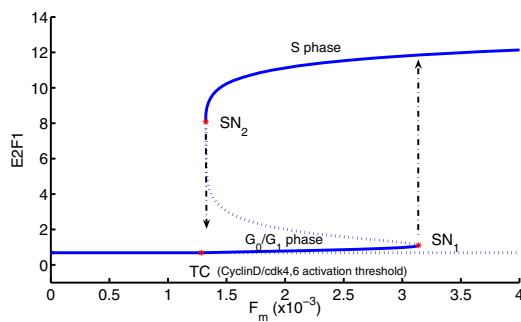


Fig. 1. Bifurcation diagram with respect to mitogenic stimulation F_m , showing various cell phases

3 Inverse Bifurcation Analysis

In the context of molecular biology, the aim of *inverse* bifurcation analysis is to map the geometry of bifurcation diagrams (back) to biochemical parameters (see [36, 37] for applications in biology; refer to [38, 39] for various applications of bifurcation control in engineering). An example of a geometric property important in the study of biological systems is the distance (in parameter space) from reference regions to bifurcation manifolds; this geometric property may be used to quantify the robustness of biological systems. Inverse bifurcation problems may be broadly divided into two classes: those of *design* type, such as finding parameter configurations so that the distance to bifurcation is as large as possible; those of *identification* type, such as finding parameter and network configurations such that the bifurcation diagram exhibits some observed behavior. In this paper, we focus on the latter type: given an ODE model, we would like to infer its properties by computationally mapping bifurcation diagrams of various shapes back to the parameter space.

For gene regulatory systems where the underlying "function" can be related to its bifurcation diagram, the proposed inverse bifurcation analysis can be useful in yielding information on how the regulation process arises from network properties. In Section 3.1, we propose a method to map bifurcation diagrams of different geometric shapes to *sparse* parameters (i.e., parameters with few components of non-zeros). That is, out of the possibly many solutions that result in the desired bifurcation diagram, we select the sparsest ones by adding a sparsity-promoting penalty term to the objective function being minimized. For our model system as described in Section 2, this allows one to identify (minimal) ways to manipulate the qualitative response of the G_1/S transition to varying mitogenic stimulation. In general, even with sparsity constraints the parameter solutions may not be uniquely determined. Hence, in Section 3.2 we propose a *hierarchical* strategy that can be used to identify a sequence of parameters that solve the inverse problem. The combination of sparsity regularization and hierarchical strategy allows for the identification of parameter sets (and the associated submodules of the system) that are influential in generating a variety of dynamical behaviors, thereby providing a way to infer the central non-linear mechanisms. We remark that there exist related methods for model analysis and reduction. For instance, a method based on linear feedback analysis to obtain destabilizing components and interactions has been used to analyze the sources of network behaviors in cell cycle and circadian rhythm models [40]. Model reduction techniques based on Proper Orthogonal Decomposition (POD) have been applied to obtain lower dimensional representations of circadian rhythm models [41]; time-scale decomposition method based on singular perturbation analysis has been used to obtain reduced-order non-linear models of metabolism [42].

3.1 Sparsity-Promoting Regularization

Inverse problems are typically ill-posed, in particular unstable. As a forward problem, bifurcation analysis is typically well-posed: the computed bifurcation

diagram exists, is unique and depends smoothly on the problem data (i.e., the vector field in some topology). In contrast, inverse bifurcation analysis is usually ill-posed, in particular: there may be no parameter configurations that can give rise to the desired bifurcation diagram; if solutions exist, they may not be unique or may not depend continuously on the problem data (i.e., the geometric description of the bifurcation diagram). Mathematical techniques, called *regularization methods*, have to be used to cope with this ill-posedness (see [43] for their mathematical theory and applications). While stabilizing ill-posed problems, regularization methods typically bias the solution to some desired behavior. Depending on the mathematical properties of the problem and the application of interest, different regularization techniques may be appropriate. Does one want to obtain a solution of the minimum Euclidean norm? Or is it more desirable to obtain a solution that is sparse, i.e., has as few non-zeros as possible? For biological applications, sparsity is often useful: for instance, one might want to find a network that is as small as possible yet consistent with the experimental data; or, one might wish to identify a small number of parameters whose variation can give rise to a wide range of system behavior. Below, we describe a sparsity-promoting regularization method and apply it in the context of inverse bifurcation analysis.

Consider the following functional, mapping vectors $x = (x_1, \dots, x_m)$ to \mathbb{R} via:

$$l_p(x) = \sum_i x_i^p.$$

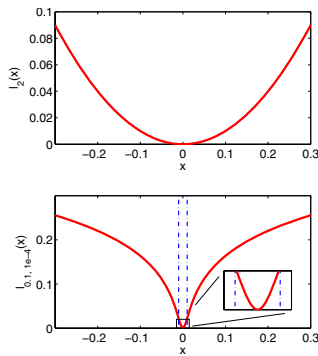


Fig. 2. l_2 and $l_{p,\epsilon}$ functionals

For $p = 2$, the above is the square of the Euclidean norm. For the (limiting) case $p = 0$, it measures the number of non-zeros in the vector x ; therefore, such a penalty term could be used to obtain sparse solutions [44]. However, $l_0(x)$ is not even a continuous function and hence significant computational effort is needed for its minimization (e.g., using combinatorial methods). In the setting of linearly constrained problems, Donoho and Elad showed that provided the solution allows for a sufficiently sparse representation, the

solution of minimum l_0 value can be obtained efficiently by solving the problem with the l_1 penalty term [44]. In our context the problems are nonlinear; to obtain sparse solutions via gradient-based methods, we consider differentiable approximations to $l_0(x)$ of the form

$$l_{p,\epsilon}(x) = \sum_i (x_i^2 + \epsilon)^{p/2}, \quad 1 \geq p > 0, \epsilon > 0,$$

where ϵ is used to remove non-differentiability at $x_i = 0$. We note that for $p < 1$, the above functional is not convex. In particular, the second derivative $d^2/dx_i^2(l_{p,\epsilon}(x))$ changes sign from being positive in the range $|x_i| < \sqrt{\epsilon/(1-p)}$ to being negative for $|x_i| > \sqrt{\epsilon/(1-p)}$. See Figure 2 for a comparison between

l_2 and $l_{p,\epsilon}$ functions; the dotted lines in the latter denote $x = \pm\sqrt{\epsilon}$. We remark that mathematically, l_p with $p \leq 1$ as a regularization method is much less understood than the case of l_2 . In the setting of (infinite-dimensional) *linear* inverse problems, Daubechies *et al.* have shown that l_1 is indeed a regularization method [45]. To enforce sparsity constraints using convex penalty terms, Ramlau and Teschke [46] have developed a method for solving *nonlinear* inverse problems using replacement functionals. However, it appears that for $p < 1$, many mathematical questions are still open.

To study the mammalian G_1/S regulatory module, we consider inverse bifurcation problems formulated as constrained optimization problems of the following form: minimize over the set of system parameters α_s ,

$$\begin{aligned} \text{ConMin}(\alpha_s, p, \epsilon) : \min_{\alpha_s} l_{p,\epsilon} \left(\frac{\alpha_s - \alpha_s^*}{\alpha_s^*} \right) \\ \text{subject to } \text{SN}_1(\alpha_s) = \text{SN}_1^* \\ \text{SN}_2(\alpha_s) = \text{SN}_2^*, \end{aligned} \quad (2)$$

where we consider placing the abscissa of saddle nodes at location SN_1^* , SN_2^* different from those for the nominal parameter value α_s^* . In particular, we consider mapping the following 3 modes of bifurcation diagram variation into the parameter space:

- Elongating saddle-node nose: $\text{SN}_1^* \leftarrow \text{SN}_1 + d$
- Moving both saddle-nodes to the right: $\text{SN}_1^* \leftarrow \text{SN}_1 + d$, $\text{SN}_2^* \leftarrow \text{SN}_2 + d$
- Decreasing range of bistability: $\text{SN}_2^* \leftarrow \text{SN}_2 + d$

Using an adjoint-based method for computing the gradient of the saddle-nodes $\text{SN}_{1,2}$ (see [38, 39, 36]), the constrained minimization problem $\text{ConMin}(\alpha_s, p, \epsilon)$ can then be solved using methods such as Sequential Quadratic Programming (SQP) (see [47, 48] for a general overview).

We solve the optimization problem (2) with all parameters being input variables (except, of course, the bifurcation parameter F_m) and taking as α_s^* the nominal parameter values given in the paper [7]. The routine `fmincon` from the MATLAB [49] Optimization Toolbox is used to solve the problem (2); the underlying algorithm is SQP with approximate Hessian obtained by BFGS-update and line-search is performed (refer to [47] for an overview of optimization algorithms). As the termination criteria, absolute tolerance of 10^{-5} is taken for function value, the constraint violation as well as the optimization variable ¹. Using regularization terms $l_{p,\epsilon}$ with $p = 0.1$, $\epsilon = 10^{-4}$ and l_2 , the top parts of Figure 3 to 5 show the initial and computed bifurcation diagrams in light and dark curves respectively; Table 1 summarizes the number of (forward) bifurcation analyses needed to obtain the results shown in the figures; the large values reflect the tight optimization tolerance used in this study and can be decreased by relaxing the optimality condition. Note that while the bifurcation diagrams shown in Figure 3 and 5 turn out to be qualitatively the same irrespective of

¹ Corresponding to the MATLAB settings `TolFun`, `TolCon`, `TolX`.

the choice of regularization term, the respective bifurcation diagrams shown in Figure 4 are qualitatively different; in particular, the transcritical bifurcation point is unconstrained, and it happens that for the parameters obtained using different regularization terms it occurs at different values of the bifurcation parameter F_m .

The bottom part of Figure 3 to 5 compare the computed changes in the parameters using different penalty terms. It can be observed that while l_2 penalty gives rise to maximum parameter changes of smaller magnitude, the computed parameters have many non-zero components. In contrast, the parameters identified using $l_{0.1,10^{-4}}$ regularization are much sparser: for all cases, only 1 or 2 parameters are classified as being identified (i.e., lie outside the range $[-\sqrt{\epsilon}, \sqrt{\epsilon}] = [-0.01, 0.01]$ as denoted by the dotted vertical lines in the figures). The remaining parameters lie within the narrow range about zero where the penalty function is locally convex owing to the smoothing term ϵ . Refer to column 1 of Table 2 for the parameters identified by the algorithm. See Sections 3.2 and 5 for the equations corresponding to the parameters and the biological interpretation of the results respectively.

3.2 Hierarchical Identification Strategy

In general, there are multiple distinct solutions satisfying the constraints in (2). It is useful to obtain a sequence of parameter solutions, allowing for non-zeros of increasing cardinality. An approach towards this goal is to identify parameters in a *hierarchical* manner.

Algorithm 1. HIER-PARAM-IDENT($\alpha_s^0 \in \mathbb{R}^m$, MaxLev, p, ϵ)

- Initialize: $s \leftarrow \{1, \dots, m\}$, $I_{\text{identified}} \leftarrow \emptyset$
- FOR $j = 1, \dots, \text{MaxLev}$
 - $I_{\text{rem}} \leftarrow s \setminus I_{\text{identified}}$
 - Solve $\alpha_{I_{\text{rem}}}^j \leftarrow \text{ConMin}(\alpha_{I_{\text{rem}}}^0, p, \epsilon)$
 - $I_j \leftarrow \{i : |(\alpha_{I_{\text{rem}}}^j)_i| > \sqrt{\epsilon}\}$
 - $I_{\text{identified}} \leftarrow I_{\text{identified}} \cup I_j$
- END
- **Return** $\{\alpha_{I_1}^1, \alpha_{I_2}^2, \alpha_{I_3}^3 \dots\}$

Table 1. Number of bifurcation runs carried out for inverse analysis

Modification Case \ Regularization Term	$l_{0.1,10^{-4}}$ -penalty	l_2 -penalty
Elongating SN_1 nose	211	68
Moving $SN_{1,2}$ to right	216	469
Decreasing bistability	197	463

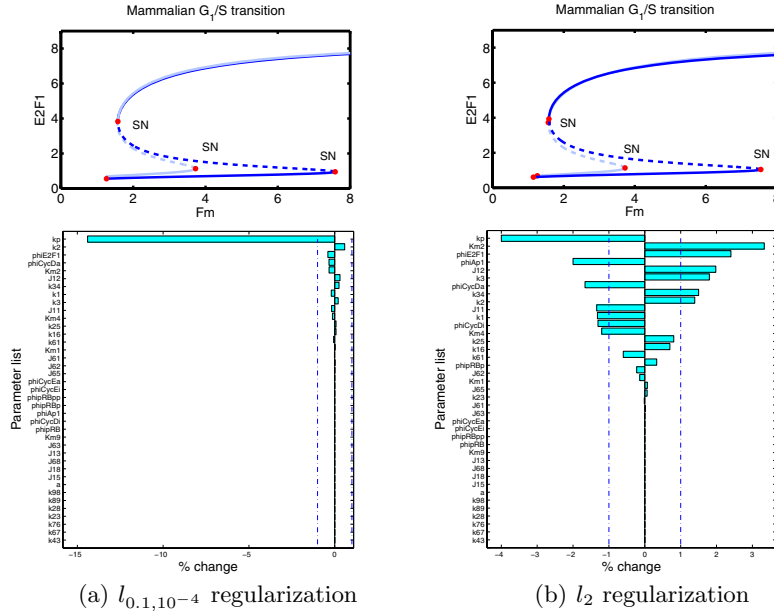


Fig. 3. Elongating saddle-node nose: bifurcation diagrams and solutions obtained using different regularization terms

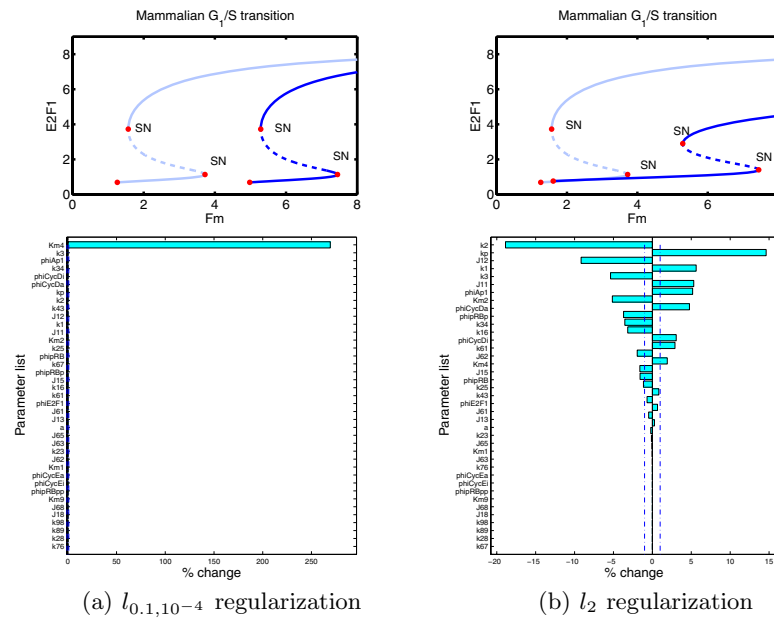


Fig. 4. Moving saddle-nodes to the right: bifurcation diagrams and solutions obtained using different regularization terms

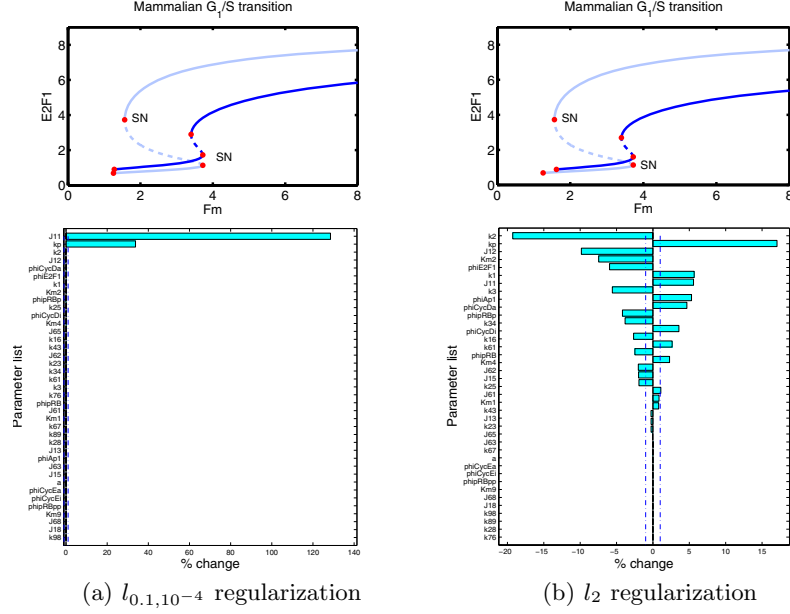


Fig. 5. Decreasing the range of bistability: bifurcation diagrams and solutions obtained using different regularization terms

In the hierarchical approach (see Algorithm 1), parameters are identified in multiple levels. Once a parameter has been identified (using the greater-than- $\sqrt{\epsilon}$ rule), in the subsequent levels it is not permitted to vary from its initial (nominal) value. Thus, it allows for the identification of distinct parameter combinations that all satisfy the same geometric constraints on the bifurcation diagram. Table 2 shows the parameters identified by the hierarchical approach for the same test cases as carried out in Section 3.1. As an illustration, let us look at the case of elongating the nose of saddle-node: the parameter k_p has been identified at level 1; subsequently, it is fixed at its nominal value of 0.05 and only the remaining parameter values $\alpha_s \setminus \{k_p\}$ are allowed to be varied in level 2. It can be observed from Table 2 that the cardinality of the parameter solution exhibits an increasing trend with the level number. Another observation is that of the parameters identified, all except the parameter ϕ_{AP-1} are associated with the following species: the core module,

$$\begin{aligned} \frac{d}{dt}[\text{pRB}] &= k_1 \frac{[\text{E2F1}]}{K_{m1} + [\text{E2F1}]} \frac{J_{11}}{J_{11} + [\text{pRB}]} \frac{J_{61}}{J_{61} + [\text{pRB}_p]} \\ &\quad - k_{16}[\text{pRB}][\text{CycD}_a] + k_{61}[\text{pRB}_p] - \phi_{\text{pRB}}[\text{pRB}], \\ \frac{d}{dt}[\text{E2F1}] &= k_p + k_2 \frac{a^2 + [\text{E2F1}]^2}{K_{m2}^2 + [\text{E2F1}]^2} \frac{J_{12}}{J_{12} + [\text{pRB}]} \frac{J_{62}}{J_{62} + [\text{pRB}_p]} - \phi_{\text{E2F1}}[\text{E2F1}] \end{aligned}$$

as well as $[\text{CycD}_i], [\text{CycD}_a]$:

$$\begin{aligned}\frac{d}{dt}[\text{CycD}_i] &= -k_{34}[\text{CycD}_i] \frac{[\text{CycD}_a]}{K_{m4} + [\text{CycD}_a]} + \dots \\ \frac{d}{dt}[\text{CycD}_a] &= k_{34}[\text{CycD}_i] \frac{[\text{CycD}_a]}{K_{m4} + [\text{CycD}_a]} + \dots\end{aligned}$$

4 Algorithm Implementation

The numerical results shown here are obtained using the Inverse Bifurcation Toolbox [36]. It is a MATLAB based implementation that utilizes `CL_MATCONT` to perform the underlying bifurcation analysis [50, 51]. Biological models in the Systems Biology Markup Language (SBML) format [52] is read in using the Mathematica [53] package `MathSBML` [54]. The vector fields are symbolically differentiated within Mathematica to obtain the derivatives $f_x, f_\alpha, f_{xx}, f_{x\alpha}$. We remark that while the current implementation is expected to be applicable to systems with dozens of parameters and state variables, it is not geared towards large-scale problems with dimensionality on the order of hundreds or thousands. To up-scale the method, further work on the mathematical as well as informatics aspects of the algorithm are needed. For the mathematical aspects, the regularization aspects needs to be studied in more detail since ill-posedness increases with parameter dimension; given the regularized problem, solution procedures can be accelerated by developing appropriate preconditioning strategies for both the inverse problem as well as the bifurcation analysis. For the informatics side, the use of compiled programming languages and efficient data management are necessary to achieve high performance: although high-level, platform-independent languages such as MATLAB and Mathematica allow for rapid algorithm prototyping, the required interpretation can result in significant slow-down of computational speed.

Table 2. Result of hierarchical algorithm with $p = 0.1, \epsilon = 10^{-4}$

Modification Case \ Param. Ident. ($\alpha_{I_j}^j$)	Level $j = 1$	Level $j = 2$	Level $j = 3$
Elongating SN_1 nose	$k_p \downarrow 14.3\%$	$k_{34} \uparrow 31.7\%$ $K_{m2} \uparrow 6.4\%$	$\phi_{AP-1} \downarrow 20.9\%$ $\phi_{E2F1} \uparrow 7.3\%$
Moving $SN_{1,2}$ to right	$K_{m4} \uparrow 269.3\%$	$J_{11} \uparrow 191.7\%$ $k_p \uparrow 17.3\%$	$k_2 \downarrow 39.9\%$ $\phi_{E2F1} \downarrow 11.7\%$ $K_{m2} \downarrow 10.3\%$
Decreasing bistability	$J_{11} \uparrow 128.5\%$ $k_p \uparrow 33.8\%$	$k_1 \uparrow 169.1\%$ $K_{m2} \downarrow 21.7\%$ $J_{12} \downarrow 20.1\%$	$k_2 \downarrow 43.7\%$ $\phi_{E2F1} \downarrow 28.3\%$

5 Discussion

In this study we attempt to manipulate specific bifurcation diagrams of a simple model of the mammalian G_1/S cell cycle transition in three different ways by

inverse bifurcation analysis. Of the parameters identified by sparsity-promoting regularization techniques, most are (perhaps not surprisingly) associated with the reactions of the core module, E2F1 autocatalysis and inhibition by pRB. After all, the model has been designed around a bistable core module [7]. But how can these manipulations and the identified parameters be interpreted in a biological context? While this is only a methodological paper and a simplified model has been chosen on purpose, we can still attempt a biological interpretation of the results. These (admittedly speculative) considerations merely serve to exemplify the general applicability of this approach to situations of specific interest and mechanistically more profound models.

Both the elongation of the nose of the SN_1 saddle-node bifurcation and the shift of the whole bifurcation to the right side would correspond to a generally decreased sensitivity of the cell to growth factors. Such an unresponsiveness is a common situation in embryonic development where cells undergo repeated transitions towards their final differentiation state [55, 56]. Intermediately differentiated cells should thus not become fully insensitive. Fully differentiated cells should not undergo anymore cell divisions and choose a different response to growth factor signals. Cell biologists would call this situation a cell cycle arrest, which is often mediated by the E2F1/pRB pair interlinked in another interesting positive feedback cycle with p53/Mdm2 pair [57]. Below, we consider the three modification cases used in the inverse analysis.

Modification 1. Elongating only the right nose of the bifurcation also implies that once activated, the system would stay active also for a subsequent decrease of the input stimulus. Once engaged in the S phase, the cell would continue to do so unless a large decrease in mitogen concentration precedes the activation of the subsequent bistable system in the cell cycle (which is not included in this model). The hierarchical identification strategy reveals three alternative ways to induce such a cell cycle arrest (see Table 2). The easiest (i.e., minimal) way to accomplish this is to increase the basal rate of E2F1 expression k_p and indeed repression of E2F1 expression was one of the two artificial manipulations inducing the expression of differentiation markers of a squamous cell carcinoma cell line [58]. Alternatively, the affinity of E2F1 to its own regulatory site can be decreased (corresponding to an increase of K_{m2}) when at the same time the maximal rate of cyclin D transcription (k_{34}) is increased (Level $j = 2$). In a similar fashion, the AP-1 transcription factor could be stabilized and E2F1 destabilized via modifications of their degradation rates (Level $j = 3$).

Modification 2. In this modification, the whole bifurcation is moved to the right, i.e., to higher mitogen concentrations. This would move the whole window of mitogen responsiveness of the G_1/S transition towards a differentiated state. Level 2 of the hierarchical approach reveals that this can be achieved in a similar manner to the first modification and Level 1, but in opposite direction: increasing k_p , but additionally increasing J_{11} , the dissociation rate of pRB from its own regulatory complex (thus decreasing self-inhibition of pRB). A minimal way to move the bifurcation (computed for Level 1) is to decrease the autocatalysis of

cyclin D activation by increasing K_{m4} . As this autocatalysis has been introduced without mechanistic justification (Swat M., personal communication) this could be interpreted as an additional input to the cyclin D activation such as the p21cip1 cyclin inhibitor, which is known to act as an activator of cyclin D/Cdk4,6 at low concentrations and is modulated by MAPK pathways as well [15]. Level 3 of this modification indicates that strengthening the dependence of E2F1 on autocatalytic activation would also yield the same results.

Modification 3. The removal of the bistability of this module, which in some sense corresponds to removing nonlinearity from this feedback system, would render the G_1/S module into a continuous responder to mitogenic signal. This would, for instance, result in the loss of threshold concentration of mitogen. For example, in the MAPK pathways it is known that an initial signal is not enough to enter the cell cycle but can yield opposing responses, depending on whether it is followed by sustained signals or not [26, 15]. A cell population with such a continuous response could loosen the ability to first check for compatibility of the cell state before starting the cell cycle. Thus, one could imagine that removing this bistability at the G_1/S transition could favor the development of cancer, as cells might start cell cycle even in presence of only little mitogenic signal. Let us look at the results of inverse bifurcation and see which parameters could be affected in such pathological conditions. Level 1 of the hierarchical strategy, i.e. the minimal number in the variation of parameters, already reveals a parameter combination that we have met previously but with different magnitudes: increasing both J_{11} (i.e., weakening pRB self-inhibition) and k_p (i.e., increasing the basal rate of E2F1 transcription) would be such deregulation that leads to loss of bistability. In Level 2 we found that an increase in E2F1 autocatalysis (lower K_{m2}), counteracted by higher pRB maximal expression (k_1) and E2F1 inhibition will also yield this result. Level 3 suggests that decreasing the E2F1 degradation rate as well as the maximal rate of auto-activation (k_2) also helps to decrease the non-linear nature of the model. This is actually the same situation as in Level 3 of the second modification case, but without additional enhancement in the E2F1 autocatalysis.

A recent, elegant knock-out study of genes involved in the yeast G_1/S transitions sheds some light on these results. Different yeast knockout strains of G_1/S genes (similar to E2F1 and cyclin D genes in mammals) show either an increased stochasticity in the onset of cell cycle or a significant lengthening of the G_1 phase [59, 60]. Clearly, these results cannot be directly mapped onto the above outlined analysis, but they show that above interpretations could well correlate with such situations. Knock-out of the *swi4* gene, which plays a similar role as our E2F1 transcription factor leads to an increased stochasticity of the G_1/S transition. Could this be a result of a decreased bistability of this control module? A similar result can at least be expected for the third modification, the conversion into a continuous responder to mitogen. Another multiple knock-out involves both the yeast's cyclin D homolog (*cln2*) and another transcription factor with a similar role to E2F1, *mpb1*. In this case, while the G_1 phase of the cells is significantly elongated, the stochasticity of the transition actually

decreased. Such a situation might be expected for the first two modifications of this analysis, the shift of the transition point to higher mitogenic stimulation.

To conclude, we remark that the relationship of a specific phenotypic response to perturbation of a cell biological system can be quite complex and often difficult to interpret. Even more challenging is the task of identifying "good" parameters for manipulation to yield a desired phenotype, such as the optimization of a metabolic process or a pharmacological intervention in pathological states. Here we considered the bifurcation diagram of a differential equation model of the mammalian *G1/S* transition as a representation of the *phenotype*. Rather than intuitive reasoning, we employed a systematic approach to modify this *bifurcation phenotype*. Comparing, for instance, the results of Level 2 of the second modification and Level 3 of the third modification, or Level 3 of the latter two modifications shows how different the consequences can be for very similar manipulations.

Thus our study exemplifies, on a simplified dynamic model of a rather qualitative nature, how the methodology of inverse bifurcation could be used as a tool in the iterative process of experimentation and modeling. A cell biological experiment can usually neither measure nor manipulate all parameters and variables of a system in parallel. Given a mathematical model of a biological system, the inverse bifurcation algorithm helps to gather information about potentially influential species and interactions. The hierarchical application of sparsity promoting constraints further allows to identify minimal sets of core parameters that a cell biologist could change to yield complex manipulations of a phenotype. Subsequently, experiments could concentrate specifically on these parameters to verify or disprove the model's assumptions. This procedure should then be carried out several times until the mathematical analysis yields a result which stands the test of experiment.

Acknowledgements

We gratefully acknowledge Lukas Endler, Stefan Müller, Maciej Swat, Philipp Kügler and Andreas Hofinger for helpful discussions. This work is supported by the WWTF, project number MA05.

References

1. Tyson JJ, Chen KC, Novak B: **Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell.** *Curr Opin Cell Biol* 2003, **15**(2):221–31.
2. Novak B, Tyson JJ: **Numerical analysis of a comprehensive model of M-phase control in *Xenopus* oocyte extracts and intact embryos.** *J Cell Sci* 1993, **106** (Pt 4):1153–68.
3. Solomon MJ: **Hysteresis meets the cell cycle.** *Proc Natl Acad Sci U S A* 2003, **100**(3):771–2.
4. Pomerenig JR, Sontag ED: **Building a cell cycle oscillator: hysteresis and bistability in the activation of *Cdc2*.** *Nat Cell Biol* 2003, **5**(4):346–51.

5. Pomerening JR, and SYK: **Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations.** *Cell* 2005, **122**(4):565–78.
6. Csikasz-Nagy A, Battogtokh D, Chen KC, Novak B, Tyson JJ: **Analysis of a generic model of eukaryotic cell-cycle regulation.** *Biophys J* 2006, **90**(12):4361–79.
7. Swat M, Kel A, Herzog H: **Bifurcation analysis of the regulatory modules of the mammalian G₁/S transition.** *Bioinformatics* 2004, **20**(10):1506–1511.
8. Swat MJ: **Bifurcation analysis of regulatory modules in cell biology.** *PhD dissertation*, Humboldt-Universität Berlin 2005, [<http://edoc.hu-berlin.de/dissertationen/swat-maciej-2005-11-03/PDF/swat.pdf>].
9. **SBML file of the mammalian G₁/S regulatory module proposed by Swat et al. (2004)** [<http://www.tbi.univie.ac.at/~raim/models/swat04/>].
10. Kuznetsov YA: *Elements of Applied Bifurcation Theory*. New York, USA: Springer-Verlag 2004.
11. Smolen P, Baxter DA, Byrne JH: **Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems.** *Am J Physiol* 1998, **274**(2 Pt 1):C531–42.
12. Hofer T, Nathansen H, Lohning M, Radbruch A, Heinrich R: **GATA-3 transcriptional imprinting in Th2 lymphocytes: a mathematical model.** *Proc Natl Acad Sci U S A* 2002, **99**(14):9364–8.
13. Coleman ML, Marshall CJ, Olson MF: **RAS and RHO GTPases in G₁-phase cell-cycle regulation.** *Nat Rev Mol Cell Biol* 2004, **5**(5):355–66.
14. Santella L, Ercolano E, Nusco GA: **The cell cycle: a new entry in the field of Ca²⁺ signaling.** *Cell Mol Life Sci* 2005, **62**(21):2405–13.
15. Roovers K, Assoian RK: **Integrating the MAP kinase signal into the G₁ phase cell cycle machinery.** *Bioessays* 2000, **22**(9):818–26.
16. Macian F, Lopez-Rodriguez C, Rao A: **Partners in transcription: NFAT and AP-1.** *Oncogene* 2001, **20**(19):2476–89.
17. Macian F, Garcia-Cozar F, Im SH, Horton HF, Byrne MC, Rao A: **Transcriptional mechanisms underlying lymphocyte tolerance.** *Cell* 2002, **109**(6):719–31.
18. Schuster S, Marhl M, Hofer T: **Modelling of simple and complex calcium oscillations. From single-cell responses to intercellular signalling.** *Eur J Biochem* 2002, **269**(5):1333–55.
19. Walker SA, Kupzig S, Bouyoucef D, Davies LC, Tsuboi T, Bivona TG, Cozier GE, Lockyer PJ, Buckler A, Rutter GA, Allen MJ, Philips MR, Cullen PJ: **Identification of a Ras GTPase-activating protein regulated by receptor-mediated Ca²⁺ oscillations.** *EMBO J* 2004, **23**(8):1749–60.
20. Dolmetsch RE, Xu K, Lewis RS: **Calcium oscillations increase the efficiency and specificity of gene expression.** *Nature* 1998, **392**(6679):933–6.
21. Li W, Llopis J, Whitney M, Zlokarnik G, Tsien RY: **Cell-permeant caged InsP₃ ester shows that Ca²⁺ spike frequency can optimize gene expression.** *Nature* 1998, **392**(6679):936–41.
22. Hajnoczky G, Robb-Gaspers LD, Seitz MB, Thomas AP: **Decoding of cytosolic calcium oscillations in the mitochondria.** *Cell* 1995, **82**(3):415–24.
23. Dupont G, Houart G, De Koninck P: **Sensitivity of CaM kinase II to the frequency of Ca²⁺ oscillations: a simple model.** *Cell Calcium* 2003, **34**(6):485–97.

24. Tomida T, Hirose K, Takizawa A, Shibasaki F, Iino M: **NFAT functions as a working memory of Ca²⁺ signals in decoding Ca²⁺ oscillation.** *EMBO J* 2003, **22**(15):3825–32.
25. Nixon VL, McDougall A, Jones KT: **Ca²⁺ oscillations and the cell cycle at fertilisation of mammalian and ascidian eggs.** *Biol Cell* 2000, **92**(3-4):187–96.
26. Marshall CJ: **Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation.** *Cell* 1995, **80**(2):179–85.
27. Murphy LO, Smith S, Chen RH, Fingar DC, Blenis J: **Molecular interpretation of ERK signal duration by immediate early gene products.** *Nat Cell Biol* 2002, **4**(8):556–64.
28. Huang CY: **Ultrasensitivity in the mitogen-activated protein kinase cascade.** *Proc Natl Acad Sci U S A* 1996, **93**(19):10078–83.
29. Ferrell JEJ, Machleder EM: **The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes.** *Science* 1998, **280**(5365):895–8.
30. Stefanova I, Hemmer B, Vergelli M, Martin R, Biddison WE, Germain RN: **TCR ligand discrimination is enforced by competing ERK positive and SHP-1 negative feedback pathways.** *Nat Immunol* 2003, **4**(3):248–54.
31. Gurdon JB, Bourillot PY: **Morphogen gradient interpretation.** *Nature* 2001, **413**(6858):797–803.
32. Hazzalin CA, Mahadevan LC: **MAPK-regulated transcription: a continuously variable gene switch?** *Nat Rev Mol Cell Biol* 2002, **3**:30–40.
33. Markevich NI, Hoek JB, Kholodenko BN: **Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades.** *J Cell Biol* 2004, **164**(3):353–9.
34. Hornberg JJ, Bruggeman FJ, Binder B, Geest CR, de Vaate AJ, Lankelma J, Heinrich R, Westerhoff HV: **Principles behind the multifarious control of signal transduction. ERK phosphorylation and kinase/phosphatase control.** *FEBS J* 2005, **272**:244–58.
35. Bhalla US, Ram PT, Iyengar R: **MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network.** *Science* 2002, **297**(5583):1018–23.
36. Lu J, Engl HW, Schuster P: **Inverse bifurcation analysis: application to simple gene systems.** *Algorithms for Molecular Biology* 2006, **1**(11).
37. Conrad E: **Bifurcation analysis and qualitative optimization of models in molecular cell biology with applications to the circadian clock.** *PhD dissertation*, Virginia Polytechnic Institute and State University 2006, [http://scholar.lib.vt.edu/theses/available/etd-04272006-1104%09/unrestricted/phd_20060510.pdf].
38. Dobson I: **Computing a closest bifurcation instability in multidimensional parameter space.** *J. Nonlinear Sci.* 1993, **3**(3):307–327.
39. Mönnigmann M, Marquardt W: **Normal vectors on manifolds of critical points for parametric robustness of equilibrium solutions of ODE systems.** *J. Nonlinear Sci.* 2002, **12**(2):85–112.
40. Schmidt H, Jacobsen EW: **Linear systems approach to analysis of complex dynamic behaviours in biochemical networks.** *Systems Biology* 2004, **1**:149–158.
41. Indic P, Gurdziel K, Kronauer RE, Klerman EB: **Development of a two-dimensional manifold to represent high dimension mathematical models of the intracellular mammalian circadian clock.** *Journal of Biological Rhythms* 2006, **21**(3):222–232.

42. Gerdtzen ZP, Daoutidis P, Hu WS: **Non-linear reduction for kinetic models of metabolic reaction networks**. *Metabolic Engineering* 2004, **6**:140–154.
43. Engl HW, Hanke M, Neubauer A: *Regularization of inverse problems, Volume 375 of Mathematics and its Applications*. Dordrecht: Kluwer Academic Publishers Group 1996.
44. Donoho DL, Elad M: **Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization**. *Proc. Natl. Acad. Sci. USA* 2003, **100**(5):2197–2202 (electronic).
45. Daubechies I, Defrise M, De Mol C: **An iterative thresholding algorithm for linear inverse problems with a sparsity constraint**. *Comm. Pure Appl. Math.* 2004, **57**(11):1413–1457.
46. Ramlau R, Teschke G: **Tikhonov replacement functionals for iteratively solving nonlinear operator equations**. *Inverse Problems* 2005, **21**(5):1571–1592.
47. Gill PE, Murray W, Wright MH: *Practical Optimization*. London: Academic Press 1981.
48. Conn AT, Gould NIM, Toint PL: *Trust-Region Methods*. MPS-SIAM Series on Optimization, Philadelphia, USA: SIAM 2000.
49. **MATLAB** [<http://www.mathworks.com/products/matlab/>].
50. Dhooze A, Govaerts W, Kuznetsov YA: **MATCONT: a MATLAB package for numerical bifurcation analysis of ODEs**. *ACM Trans. Math. Software* 2003, **29**(2):141–164.
51. **MATCONT: continuation software in MATLAB** [<http://www.matcont.ugent.be/>].
52. **Systems Biology Markup Language** [<http://sbml.org/>].
53. **Mathematica** [<http://www.wolfram.com/products/mathematica/>].
54. Shapiro BE, Hucka M, Finney A, Doyle J: **MathSBML: a package for manipulating SBML-based biological models**. *Bioinformatics* 2004, **20**:2829–2831.
55. Perez-Pomares JM, Munoz-Chapuli R: **Epithelial-mesenchymal transitions: a mesodermal cell strategy for evolutive innovation in Metazoans**. *Anat Rec* 2002, **268**(3):343–51.
56. Korenjak M, Brehm A: **E2F-Rb complexes regulating transcription of genes important for differentiation and development**. *Curr Opin Genet Dev* 2005, **15**(5):520–7.
57. Godefroy N, Lemaire C, Mignotte B, Vayssiere JL: **p53 and Retinoblastoma protein (pRb): a complex network of interactions**. *Apoptosis* 2006, **11**(5):659–61.
58. Wong CF, Barnes LM, Dahler AL, Smith L, Popa C, Serewko-Auret MM, Saunders NA: **E2F suppression and Sp1 overexpression are sufficient to induce the differentiation-specific marker, transglutaminase type 1, in a squamous cell carcinoma cell line**. *Oncogene* 2005, **24**(21):3525–34.
59. Bean JM, Siggia ED, Cross FR: **Coherence and timing of cell cycle start examined at single-cell resolution**. *Mol Cell* 2006, **21**:3–14.
60. Ubersax JA: **A noisy 'Start' to the cell cycle**. *Mol Syst Biol* 2006, **2**:2006.0014.