State-Space Approach to Modeling Dynamics of Gene Regulation in Networks

Momiao Xiong[1,2] and Li Jin[3]
[1]*Laboratory of Theoretic Systems Biology,* [3]*School of Life Science, Fudan University, China*
[2]*Human Genetic Center, University of Texas, Health Science Center at Houston,*
E-mail: momiao@gmail.com

ABSTRACT: Genetic networks are a key to unraveling dynamic properties of biological processes and regulation of genes plays an essential role in dynamic behavior of the genetic networks. A popular characterization of regulation of the gene is a kinetic model. However, many kinetic parameters in the genetic regulation have not been available. To overcome this difficulty, in this report, state-space approach to modeling gene regulation is presented. Second-order systems are used to characterize gene regulation. Interpretation of coefficients in the second order systems as resistance, capacitance and inductance is studied. The mathematical methods for transient response analysis of gene regulation to external perturbation are investigated. Criterion for classifying gene into three categories: underdamped, overdamped and critical damped is discussed. The proposed models are applied to yeast cell cycle gene expression data.

1 INTRODUCION

As the genome of many organisms including humans has been sequenced, it is time to develop genome-scale models for elucidating biological systems [1]. DNA mutations result in the subsequent transcriptional and translational alternations. Dynamic changes in gene expressions are closely related to the phenotypes of the cells and diseases. To study dynamic properties of biological systems is very important not only for gaining into deep understanding of biological processes, but also for the development of efficient treatments of the diseases. However, the primary step toward investigation of dynamics of biological systems is to develop their accurate and complete description [10].

Genetic networks are a key to a detailed understanding dynamic behavior of biological processes [9]. Since the operon model of Jacob and Monod was proposed, many mathematical models describing gene regulation have been developed [12], which have demonstrated that naturally occurring genetic networks are "too complex for qualitative descriptions devoid of mathematics" [11].

An essential part of mathematical models of genetic networks is to characterize dynamic behavior of regulation of a single gene in the network. A popular characterization of regulation of the gene is a kinetic model. However, Simulation results showed that many kinetic parameters in the genetic regulation could not be identified from the expression data, which indicated that to reduce the number of parameters in the model is necessary for modeling genetic regulation using expression data. To develop a mathematical framework for modeling gene regulation using expression data is indispensable.

Gene regulation is a complex system that requires accomplishing complex tasks with high accuracy. A very powerful approach to modeling complex systems is the state-space approach [6]. A key idea behind this approach is development of the concept of the state. The state variables that determine the dynamic behavior of the systems may be hidden and cannot directly be observed. The state-space approach is a good choice for modeling regulation.

The number of variables to describe regulation is very flexible, depending on the available data and purpose of applications. The number of parameters in the models can be chosen much less than the number of parameters in the kinetic models of the gene regulation, but larger than that in the most reverse engineering models of genetic regulatory networks.

Development of RNA interference for repression of gene expression [4,13] and synthetic biology in designing artificial regulatory circuits [3,5,7,11,12] will open a new avenue for target gene therapy [8]. Mechanisms for using RNA interference and artificial genetic networks as target gene therapy are systematical control of the state of the cell to minimize malfunctions of the cells. Their successful applications to the development of "smart drugs" require state-space representation of dynamics of gene regulation and genetic networks and applications of modern control theory to design and analysis of regulatory systems.

The purpose of this report is to develop state-space models for gene regulation, study transient response of gene expression under environment stimuli and classify genes according to their transient response to external perturbation. The proposed methods will be applied to cell cycle of yeast gene expression data. Classification of genes in terms of their transient response behavior for the yeast gene expression data set will be performed.

2. STATE SPACE MODELS FOR GENE REGULATION

The regulation of a gene can be modeled by the state-space equations with two state variables:

$$\dot{X} = AX + Bu, \quad y = CX + Du$$

where

$$\dot{X} = \begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \end{bmatrix}, \qquad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \end{bmatrix}$, $C = [c_1 \quad c_2]$, and $D = [d_{11}, d_{12}, \ldots, d_{1m}]$.

The state variables X_1 and X_2 that determine the states of the regulatory process of the gene are hidden variables and unobservable. All state variables are hypothetical variables. State variables represent biological forces to regulate transcription of genes, which describe the behavior of gene transcription, but may be hidden in the system. The output variable y denotes the expression level of the gene and hence is observable. The expression level of the gene is determined by the state variables which describe states of regulation of the gene expressions. The input variables u include expression levels of other connected genes which activate or repress the expression of the gene and external stimuli such as chemicals and drugs. The matrices *A, B, C,* and *D* are called state matrix, input matrix, output matrix and direct transmission matrix.

As a special case of the state-space mode, the gene expression can be modeled by a second-order differential equation:

$$a\frac{d^2X}{dt^2} + d\frac{dX}{dt} + pX = bu \qquad (3.1)$$

Let us define $X_1 = X$ and $X_2 = aX$. Then equation (3.1) can be written as

$$\dot{X}_1 = \frac{1}{a}X_2, \dot{X}_2 = -pX_1 - \frac{d}{a}X_1 + bu,$$

or $\dot{X} = AX + Bu$, $Y = cX$

where

$\dot{X} = \begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \end{bmatrix}$, $A = \begin{bmatrix} 0 & \frac{1}{a} \\ -p & \frac{-d}{a} \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ b \end{bmatrix}$, and

$C = [1 \quad 0]$.

Second-order model for the gene regulation is similar to the models for the components in the mechanical and electrical systems:

$$m\frac{d^2X}{dt^2} + b\frac{dX}{dt} + kX = P \qquad \text{and}$$

$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = e$, respectively, where *P* is a force, *X* is a displacement, *m* is mass, *b* is a viscous-friction coefficient, *K* is a spring constant, e is a voltage, *q* is charge, *R* is resistance, *L* is inductance, and *C* is capacitance. The equations describe the causal and effect relationship which widely exists in the nature.

The parameters in the second-order model for the gene regulation have some biological interpretation. The transcriptional complex causes the changes of the concentration of the transcribed gene. Like dynamics of the mechanical and electrical systems, we assume that the dynamics of the gene transcription involves three elements: inertia coefficient element, damper elements and spring (or proportion) element. The coefficient of the second-order derivative in the equation (3.1) is referred to as a transcriptional inertia, which measures the change in the concentration of the regulator required to make a unit change in acceleration of the concentration change of the transcribed gene. The transcriptional damper coefficient is referred to as resistance defined as the change in concentration of the regulator required to make a unit change rate of the concentration of the transcribed gene which provides resistance to transcription. Transcriptional spring constant or proportionality constant can be defined as the change in concentration of the regulator required to make a unit change of concentration of the transcribed gene which indicates stiffness. The reciprocal of spring constant P is called transcriptional capacitance. The transcriptional inertia, resistance and capacitance are functions of the DNA sequence and the kinetic parameters of the transcription process, and to measure the performance of transcription.

3. TRANSIENT-RESPONSE ANALYSIS OF GENE REGULATION

Response of a dynamic system has two parts: the transient and the steady state response. The process generated in going from the initial state to the final state is called transient response. Steady-state response studies the system behavior at infinite time. Transient-response analysis of gene regulation can be sued to quantify the dynamics of the gene regulation. It can reveal how fast the gene regulation responds to external changes and how accurately the gene expressions can finally achieve the desired steady-state values. It can also be used to study damped vibration behavior and stability of the gene regulation. The equation which determines the complementary solution of equation (3.1) is given by

$$a\frac{d^2X}{dt^2} + d\frac{dX}{dt} + pX = 0 \qquad (3.2).$$

Then, the characteristic equation of the second-order differential equation (3.2) is then given by [6]

$$as^2 + ds + p = 0 \qquad (3.3).$$

If the second-order system for gene regulation is stable, then all coefficients of the characteristic equation must be positive. If we assume that the second-order system of gene regulation is stable, for the convenience of presentation we define

$$\omega_n = \sqrt{\frac{p}{a}} \quad \text{and} \quad \xi = \frac{d}{2a\sqrt{ap}}.$$

The equation (3.3) then can be rewritten as

$$\ddot{x} + 2\xi\omega_n\dot{x} + \omega_n^2 x = 0. \qquad (3.4)$$

Now we consider the following three cases: underdamped case ($0 < \xi < 1$), overdamped case ($\xi > 1$), and critically damped case

$(\xi = 1)$.

Case 1. Underdamped case $((0 < \xi < 1)$.

Let $\omega_d = \omega_n\sqrt{1-\xi^2}$ = damped natural frequency. Then, the response *x(t)* of gene regulation is given by

$$x(t) = e^{-\xi\omega_n t}\{[\frac{\xi}{\sqrt{1-\xi^2}}x(0) + \frac{1}{\omega}\dot{x}(0)\sin\omega_d t + x(0)\cos\omega_d t\}$$

This shows that the natural response in this case is an exponentially decaying and the system is said to be underdamped.

Case 2. Overdamped case $(\xi > 1)$. The response *x(t)* of gene regulation in case is given by

$$x(t) = \left[\frac{(-\xi + \sqrt{\xi^2-1})x(0)}{2\sqrt{\xi^2-1}} - \frac{\dot{x}(0)}{2\omega_n\sqrt{\xi^2-1}}\right] e^{-(\xi\omega_n + \omega_n\sqrt{1-\xi^2})t} + \left[\frac{(\xi + \sqrt{\xi^2-1})x(0)}{2\sqrt{\xi^2-1}} + \frac{\dot{x}(0)}{2\omega_n\sqrt{\xi^2-1}}\right] e^{-(\xi\omega_n - \omega_n\sqrt{\xi^2-1})t}$$

Both terms in the solution decrease exponentially. The system of gene regulation is said to be overdamped.

Case 3. Critically damped $(\xi = 1)$. In this case, the two solutions of characteristic equation (3.4) are the same and can be written as

$$x(t) = x(0)e^{-\omega_n t} + [\omega_n x(0) + \dot{x}(0)]te^{-\omega_n t}.$$

In this case, the system of gene regulation is referred to as critically damped.

The transient response of the dynamic systems is also characterized by step response and impulse response of the systems. The step response and impulse response are defined as follows.

(1) Unit-step response

The step function is defined as

$$u(t) = \begin{cases} 0 & for \quad t < 0 \\ 1 & for \quad t > 0 \end{cases}$$

The Laplace transform of the unit-step function is given by

$$U(s) = \frac{1}{S}.$$

Therefore, the transfer function of the response of the genetic network to unit-step input signal is

$Y(s) = \frac{G(s)}{S}$, where $G(s)$ is the transfer function of the genetic regulatory system.

(2) Unit-impulse response

Consider the impulse function:

$$u(t) = \begin{cases} \infty & for \quad t = 0 \\ 0 & for \quad t \neq 0 \end{cases}$$

Its Laplace transform is equal to $U(s) = 1$. The transfer function of the response of the genetic network to unit-impulse input signal is the simply equal to $Y(s) = G(s)$.

4. RESULTS

To study validity of the proposed model for gene regulation, we apply the state-space model to cell cycle gene expression time course data of yeast (Cho et al. 1998). Cells were collected at every 10 minutes. The total time points were 17. To quantify gene expression levels, the extracted total mRNA of collected yeast cells were hybridized with oligonucleotide array. To evaluate the performance of the proposed state-space models for analysis of regulation, we plot Figures 1 and 2, which show the observed and predicted expression profiles of genes Cdc20 and SWI4. Figures 1 and 2 demonstrate that the accuracy of the second-order systems for modeling regulation of genes Cdc20 and SWI4 is very high. This has great implications. Second-order systems have broad applications in electric and mechanical systems. Knowledge and information in analysis of electric and mechanical systems can be borrowed for analysis of gene expression and regulation data.

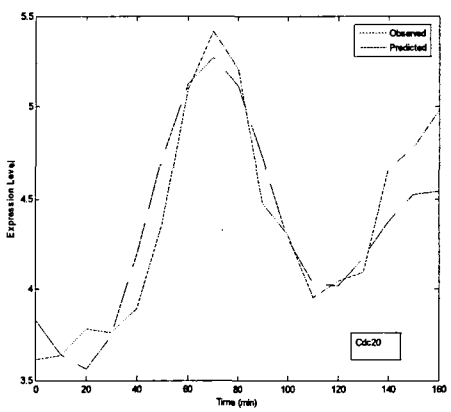

Figure 1 Observed and predicted expression Profiles of gene Cdc20.

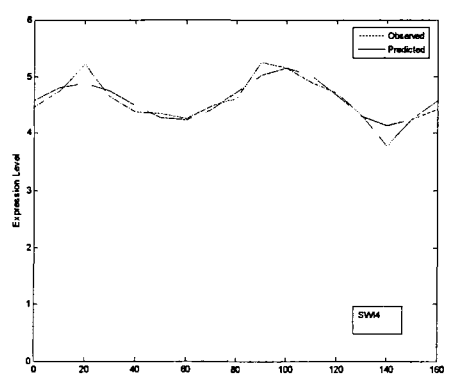

Figure 2 Observed and predicted expression profiles of gene SWI4.

Response of a dynamic system has two parts: the transient and the steady state response. The process generated in going from the initial state to the final state

is called transient response Steady-state response studies the system behavior at infinite time. Transient-response analysis of gene regulation can be used to quantify the dynamics of the gene regulation. It can reveal how fast the gene regulation responds to external changes and how accurately gene regulation can finally achieve the desired steady-state values. It can also be used to study damped vibration behavior and stability of the gene regulation.

To investigate the transient-response behavior of the gene regulation in cell cycles of yeast, we plot Figures 3 and 4 showing unit-step response and unit-impulse response of genes Cdc20 and Cdc6, respectively. The damping ratio of expressions of genes Cdc20 and Cdc6 are equal to $\xi_1 = 0.1389$ and $\xi_2 = 273$, respectively. Thus, regulation of gene Cdc20 is underdamped and regulation of gene Cdc6, are overdamped. Figure 3 showed that both step and impulse response of regulation of gene Cdc20 is quick and amplitude of harmonic motion of response decreases to steady value of gene expression, but Figure 4 showed that both step and impulse response of regulation of gene Cdc6 is slow and decreases exponentially.

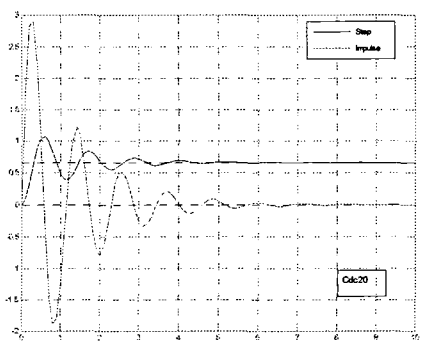

Figure 3 Step and impulse response of regulation of gene Cdc20

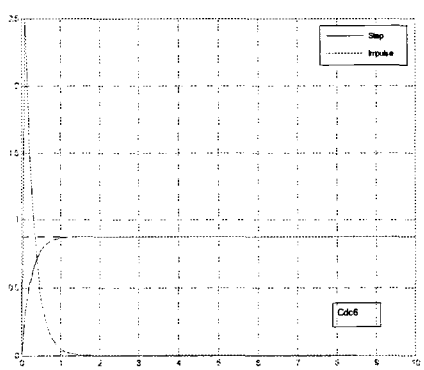

Figure 4 Step and impulse response of regulation of gene Cdc6.

Stability is an important notion of dynamic systems. Stability of gene regulation will determine cell functions. Expression time course profiles of most genes in yeast cell cycle are stable. However, as Figures 5 and 6 shows that expression of gene RLM1 is unstable. Sign of the coefficients a and d is different. Applying Routh's stability criterion [6] can mathematically show that the system of regulation of gene RLM1 is unstable.

Tables 1, 2 and 3 summarized the damping ratios and errors of model fitting for the genes in the cell cycle control of FKS2, pheromone signal transduction and DNA replication pathways. The genes can be classified into three categories: Underdamped, critically damped and overdamped. Most of the genes in the tables are underdamped. Since the expression data are stochastic, the above classifications of the genes are not precisely correct. However, the results demonstrate that pattern of response of regulation of the genes in yeast cell cycle is not the same.

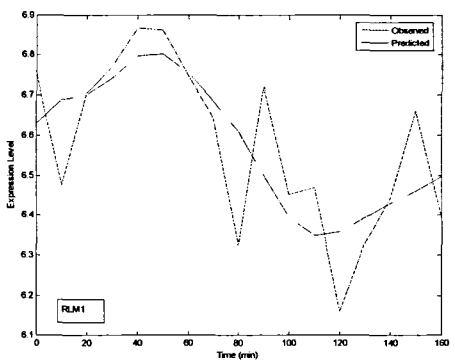

Figure 5: Observed and predicted expression profiles of gene RLM1.

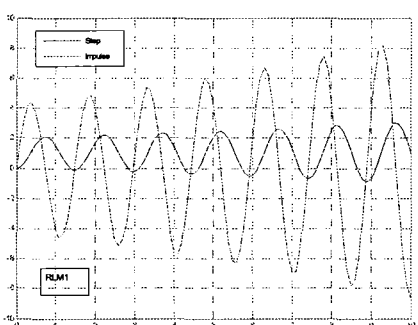

Figure 6: Step and impulse of regulation of gene RLM1

Gene	Damping Ratio	error
YPR165W/RHO1	-0.05667	0.779567
YPL089C/RLM1	-0.04784	0.420926
YOR231W/MKK1	0	0.375668
YDR477w/SNF1	0	0.526823
YPL140C/MKK2	0.409043	0.237493
YJL095W/BCK1	0.586783	0.563527
YGL035C/MIG1	3.57388	0.414533
YBL105c/PKC1	20.25406	0.2625

Table 1: Estimated coefficients and damping ratio of second-order model for regulation of the genes in FKS2 pathway of the yeast.

Gene	Damping Ratio	error
YBR200w/BEM1	0.067896	0.382593
YHR005c/GPA1	0.095949	0.605646
YHL007c/STE20	0.216989	0.256382
YJL157C/FAR1	0.384274	1.355324
YLR362W/STE11	0.435011	0.269565
YDR103w/STE5	0.731119	0.29178
YJR086W/STE18	1.280914	0.31124
YLR229C/CDC42	2.262786	0.282246
YOR212W/STE4	2.996461	0.185345
YFL026W/STE2	3.068568	0.705242
YAL041W/CDC24	14.624614	0.294137
YBL016w/FUS3	15.462556	0.192331
YDL159W/STE7	34.576792	0.239336
YGR040W/KSS1	138.972895	0.377493

Table 2: Estimated coefficients and damping ratio of second-order model for regulation of the genes in pheromone signal transduction pathway of the yeast.

Gene	Damping ratio	error
YER111c/SWI4	-0.014937	0.398371
YJL187C/SWE1	0.020327	0.433615
YGL116W/CDC20	0.138914	0.512818
YHR166c/CDC23	0.17347	0.335368
YOR373W/NUD1	0.284766	0.542291
YGR108W/CLB1	0.289181	0.80225
YDL155W/CLB3	0.308586	0.250906
YDL155W/CLB3	0.308586	0.250906
YNL289W/PCL1	0.361578	1.040091
YMR001C/CDC5	0.392486	0.314102
YFL009W/CDC4	0.438115	0.156234
YLR103c/CDC45	0.481895	0.719066
YGR109C/CLB6	0.553791	1.288552
YBL084c/CDC27	0.6241	0.136391
YLR079w/SIC1	0.642222	1.075208
YLR079w/SIC1	0.642222	1.075208
YMR199W/CLN1	0.658201	1.172824
YFR028C/CDC14	0.676608	0.155801
YDL127w/PCL2	0.698367	0.53357
YAL040C/CLN3	0.774119	0.851464
YLR210W/CLB4	0.878328	0.540543
YDR052C/DBF4	1.013155	0.108905
YDL017W/CDC7	1.471897	0.237697
YKL022C/CDC16	1.500951	0.035274
YBR160w/CDC28	2.538795	0.377918
YPL256C/CLN2	139.7371	1.059588
YJL194W/CDC6	273.0880	0.771104

Table 3: Estimated coefficients and damping ratio of second-order model for regulation of the genes in cell cycle control of DNA replication pathway of the yeast.

DISCUSSION

Many important cell functions are largely determined by dynamic processes of biochemical networks, such as genetic and metabolic networks. A critic issue for studying dynamic behavior of biochemical networks is how to model biochemical networks. Regulation of single gene is an essential element of genetic networks. To develop simple, accurate and applicable model of gene regulation, to study the transient response of gene regulation based on the developed model of gene regulation and to classify genes according to their dynamic behaviors of regulation are indispensable.

A popular characterization of regulation of the gene is a kinetic model. However, most kinetic parameters are not available and recent simulation results showed that many kinetic parameters in the genetic regulation could not be identified from the expression data. This implies that to develop models for gene regulation and to estimate parameters in the models from experimental data sets are only ways to ensure the success of research in dynamic studies of biological systems and systems biology.

The state-space approach has been successfully applied to solving various engineering problems. The purpose of this report is to attempt to borrow state-space methods to model gene regulation. Since the mechanisms of gene regulation have still not been fully discovered, many variables (factors) which determine gene regulation are unknown and hidden in the regulatory processes. Therefore, the concept of state variable is very suitable for description of the regulatory processes. Application of state-space models and second-order systems to gene regulation will allow us to use well developed modern theories of systems science and controls, and corresponding software as well. In this report, state-space models and second-order systems for gene regulation were presented and applied to yeast cell cycle time course gene expression data. Interestingly, the results showed that the accuracy of second-order systems for modeling gene regulation is high. High precision of prediction of gene expressions by second-order model will provide a powerful tool to analyze regulatory systems from gene expression profiles.

Transient response analyses of gene regulation using theoretic models coupled with fitting of theoretic models to gene expression data have not been documented in literatures. Transient step and impulse response analysis of gene regulation in yeast cell cycle by second-order models in this study reveal many unknown features of dynamic processes of gene regulation. I found that there

exist some genes whose regulation is unstable. The reasons for presence of unstable gene regulation in yeast cell cycle are unknown. It is possible that some identified genes with unstable gene expressions are due to experimental errors. But, it is difficult to image that all found genes with unstable regulation are due to noise data. Experimental research in stability analysis of gene regulation is urgently needed. Three categories of gene regulation are useful concepts for characterization of dynamic behavior of gene regulations. To perform analysis for classifying all genes with stable gene regulation in the yeast and other organisms into one of three categories should be planned in the future.

REFERENCES

[1] T. E. Allen, B. O. Palsson. Sequence-based analysis of metabolic demands for protein synthesis in prokaryotes. J Theor Biol 220, 1-18, 2003.

[2] R. J. Cho, M.J. Campbell, E.A. Winzeler, L. Steinmetz, A.Conway, L.Wodicka, T. G Wolfsberg, A.E. Gabrielian, D. Landsman, D. J. Lockhart, R. W. Davis. A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell. 2(1):65-73, 1998.

[3] D. Ferber. Synthetic biology. Time for a synthetic biology Asilomar? Science. 303(5655):159, 2004.

[4] P. Ganju, J. Hall. Potential applications of siRNA for pain therapy. Expert Opin Biol Ther. 4(4):531-42, 2004.

[5] G C. Guet, M. B. Elowitz , W. Hsing, S. Leibler. *Combinatorial synthesis of genetic networks.* Science. 296(5572):1466-7,2002..

[6] K. Ogata. System dynamics. Third Edition. Prentice Hall. Upper Saddle Rive, New Jersey. 1998.

[7] H. Kobayashi,M. Kaern, M. Araki, K. Chung, T. S. Gardner, C. R. Cantor, J. L. Collins. Programmable cells: interfacing natural and engineered gene networks. Proc Natl Acad Sci U S A. 101:8414-9, 2004.

[8] I. H. Tarner, U. Muller-Ladner, C. G Fathman. Targeted gene therapy: frontiers in the development of 'smart drugs'. Trends Biotechnol. 22(6):304-10, 2004.

[9] P. Smolen, D. A.Baxter, J.H. Byrne. Modeling transcriptional control in gene networks--methods, recent results, and future directions. Bull Math Biol. 62:247-92,2000.

[10] S. H. Strogatz .Exploring complex networks Nature. 410:268-276. Review. 2001.

[11] J. Hasty, D. McMillen, J.J.Collins. Engineered gene circuits. Nature. 420: 224-30,2002.

[12] M.E. Wall, W.S. Hlavacek, M.A. Savageau. Design of gene circuits: lessons from bacteria. Nat Rev Genet. 5(1):34-42,2004.

[13] M.J. Wood, B. Trulzsch, A. Abdelgany, D. Beeson. Ribozymes and siRNA for the treatment of diseases of the nervous system. Curr Opin Mol Ther. 5(4):383-8, 2003.