Hindawi Publishing Corporation
EURASIP Journal on Bioinformatics and Systems Biology
Volume 2006, Article ID 59526, Pages 1–12
DOI 10.1155/BSB/2006/59526

Stochastic Oscillations in Genetic Regulatory Networks: Application to Microarray Experiments

Simon Rosenfeld

Division of Cancer Prevention, Biometry Research Group, National Cancer Institute, Bethesda, MD 20892, USA

Received 19 January 2006; Revised 26 June 2006; Accepted 27 June 2006

Recommended for Publication by Yue Wang

We analyze the stochastic dynamics of genetic regulatory networks using a system of nonlinear differential equations. The system of *S*-functions is applied to capture the role of RNA polymerase in the transcription-translation mechanism. Using probabilistic properties of chemical rate equations, we derive a system of stochastic differential equations which are analytically tractable despite the high dimension of the regulatory network. Using stationary solutions of these equations, we explain the apparently paradoxical results of some recent time-course microarray experiments where mRNA transcription levels are found to only weakly correlate with the corresponding transcription rates. Combining analytical and simulation approaches, we determine the set of relationships between the size of the regulatory network, its structural complexity, chemical variability, and spectrum of oscillations. In particular, we show that temporal variability of chemical constituents may decrease while complexity of the network is increasing. This finding provides an insight into the nature of "functional determinism" of such an inherently stochastic system as genetic regulatory network.

1. INTRODUCTION

According to the "central dogma" in molecular biology, the genetic regulatory process involves two key steps, namely, "transcription," that is, deciphering the genetic code and creation of the messenger RNA (mRNA), and "translation," that is, synthesis of the proteins by ribosomes using the mRNAs as templates. These processes run concurrently for all the genes comprising the genome. Importantly, each molecular assembly responsible for deciphering the genetic code is itself built from the proteins produced through transcription and translation of other genes, thus introducing nonlinear interactions into the regulatory process (Lewin [1]). In the human genome, for example, from 30 to 100 regulatory proteins are usually involved in each transcription event in each of about 30,000 genes. This means that the regulatory network is simultaneously of a very high dimensionality and very high connectivity. Mathematical description of such a network is a challenging task, both conceptually and computationally. Quite paradoxically, however, this seemingly unfavorable combination of two "highs" opens a new avenue for approximate solutions and understanding the global behavior of regulatory systems through the application of asymptotic methods. The novelty introduced by our model is that it does not simplify the processes through decreasing the dimensionality. On the contrary, the model takes advantage of the system being asymptotically large.

In this paper, we pay special attention to quantitative relations between the transcription levels (TLs), that is, the numbers of mRNA molecules of a certain type per cell, and transcription rates (TRs), that is, the numbers of mRNA molecules produced in the cell per unit of time. TLs are the quantities directly derived from microarray experiments, whereas TRs are usually unobservable. Although both of these quantities seem to be legitimate indicators for characterizing gene activity, generally they are different and capture different facets of the regulatory mechanism. The fundamentally nonlinear nature of the gene-to-gene interactions precludes any direct relations between gene-specific TRs and TLs. Also, due to the inherent instability of high-dimensional regulatory systems, nothing like time-independent "gene activity" may be attributed to a living cell. In our view, these conclusions may have serious consequences for the interpretation of microarray experiments where the fluctuating

nature of the mRNA levels is frequently ignored, mRNA abundance is often seen as a direct indicator of the corresponding gene's activity, and the differential expression (i.e., difference in TLs) is taken as evidence of differences in the cells themselves.

2. ASSUMPTIONS AND EQUATIONS

The system of nonlinear ordinary differential equations for the description of proteome-transcriptome dynamics first appeared in [2]

$$\frac{\mathbf{dr}}{\mathbf{dt}} = \mathbf{F}(\mathbf{p}) - \boldsymbol{\beta}\mathbf{r}, \qquad \frac{\mathbf{dp}}{\mathbf{dt}} = \boldsymbol{\gamma}\mathbf{r} - \boldsymbol{\delta}\mathbf{p}, \tag{1}$$

where $\mathbf{r}$ and $\mathbf{p}$ are n-dimensional column vectors of mRNA and protein concentrations measured in numbers of copies per cell; n is the number of genes in genome; $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$ are nondegenerate diagonal matrices corresponding to the rates of production and degradation in transcription and translation. The n-dimensional vector-function, $\mathbf{F}(\mathbf{p})$, is a strongly nonlinear function representing the mechanism of transcription. Chen et al. [2] linearized the system (1) in the vicinity of a certain hypothesized initial point and formulated general requirements of stability. In what follows, we augment the system (1) by an explicitly specified model for $\mathbf{F}(\mathbf{p})$ and attempt to extract the consequences from the essentially nonlinear nature of the problem. Note that according to commonly accepted terminology of chemical kinetics (Zumdahl [3]), production rate is defined as the number of molecules *produced* in the system per unit of time. It may or may not be balanced by an opposite process of degradation. Because transcription is the process of *production* of mRNA, we refer to the quantity $\mathbf{F}(\mathbf{p})$ as *transcription rate*.

As is known from the biology of gene expression, generation of each copy of messenger RNA is preceded by a complex sequence of events in which a large number of proteins bind to the gene's regulatory sites and assemble a reading mechanism known as RNA polymerase (RNAP) (Kim et al. [4]). Each binding represents a separate biochemical reaction involving DNA and proteins and is supported by a number of enzymes and smaller molecules. According to the principles of chemical kinetics, the production term, $\mathbf{F}(\mathbf{p})$, should have the following general form (De Jong [5])

$$F_i(p_1, \ldots, p_n) = \sum_{k=1}^{L_i} \omega_{ik} \prod_{m=1}^{n} p_m^{r_{ikm}}, \tag{2}$$

where L_i is the number of concurrent biochemical reactions for decoding the ith gene; ω_{ik} are the rate constants; and r_{ikm} are the kinetic orders showing how many protein molecules of type m participate in kth biochemical reaction for the transcription of ith gene. A detailed account of the assumptions underlying (2) may be found in [6]. Although these assumptions are not free from inevitable simplifications, they constitute a reasonably solid basis for studying the dynamics of genetic regulatory networks because they recognize the central role of RNAPs in the nonlinear mechanism of gene-to-gene interactions. However, it should be unequivocally stated that many secondary mechanisms of regulation remain beyond the scope of this model. For example, system (1) depicts an important process of mRNA degradation as the first-order chemical reaction. This suggests the idea that the proteins controlling the ribosomes do not return back to the genetic regulatory network and do not become the parts of the deciphering assemblies again. Of course, this is a comparatively crude representation of a more complex process which takes place in reality (Maquat [7]). However, inclusion of this and similar processes into the model does not amount to a new mathematical problem because the system (1)-(2) may be easily augmented by additional terms expressed through the S-function in the same manner as in (2).

Although the biochemical nature of gene expression cannot be doubted, applicability of the standard concepts and descriptors of chemical kinetics to these processes is not out of question. For example, the process that is commonly compartmentalized as "binding" of a protein to the regulatory site is, in fact, a sequence of events of enormous complexity involving a large number of transcriptional coactivators. In a sense, each such binding is a unique adventure which cannot be directly characterized in terms of constant gene-specific chemical rates and stoichiometric coefficients (Lemon and Tjian [8]). The processes of synthesis of the RNAPs may be schematically subdivided into a sequence of steps and rearrangements which may be thought, again with a certain degree of abstraction, as separate biochemical reactions. That is why it is admissible to say that there are many chemical reactions between the proteins and the DNA molecule which run concurrently within the same regulatory site. However, one needs to be careful with excessively straightforward application of standard biochemical terminology and quantitative parameterization to such processes, only in principle similar to simple biochemical reactions.

3. STOCHASTICITY IN GENETIC REGULATORY NETWORKS

There is a large body of theoretical and experimental works devoted to various aspects of randomness and stochasticity in coupled biochemical systems. We briefly summarize some of the key facts here.

As indicated by Gillespie [9], "the temporal behavior of a chemically reacting system of classical molecules is a deterministic process in the $2N$ position-momentum phase space, but it is not a deterministic process in the N-dimensional subspace of the species population numbers. Therefore, both reactive and non-reactive molecular collisions are intrinsically random processes characterized by the collision probability per unit of time. That is why these collisions constitute a stochastic Markov process, rather than a deterministic rate process."

Elf and Ehrenberg [10] observe that "the copy numbers of the individual messenger RNAs can often be very small, and this frequently leads to highly significant relative fluctuations in messenger RNA copy numbers and also to large fluctuations in protein concentrations." In addition, there are inevitable statistical variations in the random partitioning of small numbers of regulatory molecules between daughter cells when cells divide (McAdams and Arkin [11]).

McAdams and Arkin [12] indicate that "time delays required for protein concentration growth depend on environmental factors and availability of a number of other proteins, enzymes and supporting molecules. As a result, the switching delays for genetically coupled links may widely vary across isogenic cells in the population. One consequence of these differing times between cell divisions is progressive desynchronization of initially synchronized cell populations. Within a single cell, random variations in duration of events in each cell-cycle controlling path will lead to uncoordinated variations in relative timing of equivalent cellular events."

Multiple closely spaced ribosomes may process the same strand of mRNA simultaneously. Because the spacings between ribosomes are random, the number of proteins translated from the same transcript may also fluctuate randomly (McAdams and Arkin [11]).

Recent experiments (Cai et al. [13]) demonstrated that even in an individual cell, the production of a protein and supporting enzymes is a stochastic process following a complex pattern of bursting with random distribution of intensities and durations. Similarly, Rosenfeld et al. [14] found that quantitative relations between transcription factor concentrations and the rate of protein production fluctuate dramatically in the individual living cells, thereby limiting the accuracy with which genetic transcription circuits can transfer signals. The processes mentioned above represent various facets of the natural stochasticity of intracellular regulatory systems. In addition, stochastic concepts are engaged as the way of describing extremely intricate quasi-chaotic behavior, even if the system is fully deterministic in principle. As demonstrated by famous examples of the Lorenz attractor (Lorenz [15]), Belousov-Zhabotinsky autocatalytic reactions (Zhang et al. [16]), Lotka-Volterra population dynamics (Lotka [17]), and many other examples (Bower and Bolouri [18]), chaotic behavior may appear even in low-dimensional systems with rather simple structure of nonlinearity. On the contrary, the intracellular biochemical networks are high-dimensional systems with a very complex structure of nonlinearity. These properties make it difficult to overcome mathematical problems without substantial simplifications. In statistical mechanics, a traditional way of formulating a complex multidimensional problem is to introduce the concept of statistical ensemble (Gardiner [19]). In high-dimensional biochemical network there are many ways to introduce a statistical ensemble, but those are preferable that provide tangible mathematical advantages combined with intuitive clarity and ease of interpretation, as discussed below.

The rate constants, ω_{ik}, and kinetic orders, r_{ikm}, are assumed to be time-independent positive real and integer numbers, respectively. For computational purposes, we specify them as random numbers drawn from the gamma and Poisson populations, respectively:

$$\begin{aligned} \Pr\left(\omega_{ik} = x\right) &= \frac{x^{\alpha-1}\exp(-x/\theta)}{\Gamma(\alpha)\theta^{\alpha}}, \\ \Pr\left(r_{ikm} = n\right) &= \frac{\lambda^{n}\exp(-\lambda)}{n!}. \end{aligned} \tag{3}$$

This choice of probabilistic characterization is a matter of mathematical convenience and may be easily replaced by other assumptions compatible with the nature of the problem. Similar to random Boolean networks (Kauffman et al. [20]), the network introduced in (1)–(3) is a collection of identical regulatory units with random assignment of functional properties controlled through the parameters ω_{ik} and r_{ikm}. To avoid a possible misconception, it should be noted that the statistical ensemble introduced through (3) is not intended to mimic a group of isogenic cells. Even less so, the ensemble (3) may be interpreted as a group of neighboring cells in the same tissue because there is always a certain degree of cooperativity and synchronization between the cells under the control of higher loops of regulation (Ptashne [21]). Therefore, these cells will not represent statistically independent members of ensemble. Rather, the ensemble (3) represents the collection of all possible networks of similar types sharing the same probabilistic structure. Simulation experiments show that both summary statistics and global time-independent parameters of such networks generated in independent runs are identical for practical purposes for those networks with size above several hundred regulatory units. Such a notion of statistical ensemble is analogous to that in statistical physics. The states of different members of the ensemble (say, the volumes of ideal gas enclosed in the thermostats with the same temperature) are not supposed to be similar to each other at any fixed moment in time because the trajectories in their respective phase spaces may be entirely different. However, what the members of ensemble do have in common are the integral time-independent statistical characteristics of these trajectories.

The usage of parameterization (3) in this work is twofold. First, it serves as a concise method for generating the network structure in simulation experiments. In the context of this research, we are not interested in peculiarities of the network behavior associated with any specific selection of the coefficients. Rather, we are interested in exploration of global behavior of the whole class of the networks sharing the same probabilistic structure. The second usage of (3) in this work is of purely technical nature. It often happens that the results of mathematical calculations are expressed in terms of summary statistics of the parameters characterizing the system. If the system is asymptotically large, then these summary statistics can be directly related to their expected values, thus allowing for representation of the results in a concise, easily comprehensible form.

4. OUTLINE OF THE SOLUTION

We seek a stationary solution of the system (1). To envision a general structure of this solution, we invoke considerations of the theory of stability of differential equations (Carr [22]). Following standard methodology, we first seek the equilibrium (fixed) point of (1)-(2) and try to determine whether the solution in its vicinity is stable or unstable. Let $\mathbf{P_0}$ be the n-vector of equilibrium protein concentrations, and $\mathbf{X}(t)$ be the vector of relative concentrations normalized by these equilibrium values. After some transformation, system (1)-(2) may be rewritten as

$$\ddot{x}_i + (\beta_i + \delta_i)\dot{x}_i + \beta_i\delta_i x_i = \beta_i\delta_i \sum_{k=1}^{L_i} \Omega_{ik} Y_{ik}, \tag{4}$$

where $\mathbf{Y}$ are the S-functions (Savageau and Voit [23]) defined as

$$\log Y_{ik}(x_1, \ldots, x_n) = \sum_{k=1}^{n} r_{ikm} \log x_m, \tag{5}$$

$$\Omega_{ik} = \frac{\omega_{ik} Y_{ik}(\mathbf{P}_0)}{\sum_{k=1}^{L_i} \omega_{ik} Y_{ik}(\mathbf{P}_0)} \tag{6}$$

(note that by definition $\sum_{k=1}^{L_i} \Omega_{ik} = 1$). System (4) is strongly nonlinear, and there are no reasons to hope that its solution may be obtained in some closed form. However, some important elements of the solution may be understood with the help of the center manifold theory (Perko [24]). A detailed discussion of the application of this theory to biochemically motivated S-systems may be found in Lewis [25]. An informal statement of this theory is that in close vicinity of the equilibrium, the trajectories residing in stable and unstable manifolds (i.e., those associated with eigenvalues of the Jacobian matrix residing in the left and right halves of the complex plane, resp.) are topologically homeomorphic to the corresponding trajectories of the linear system. The solutions associated with the purely imaginary eigenvalues (which would be quasi-periodic in the linear theory) become the sources of extremely intricate chaotic behavior, but importantly these solutions are bounded, thus representing a sort of stationary random-like process. Note that in practical applications it is not usually required that the real parts of the roots in the center manifold are to be exactly zero, they only need to be small enough to justify ignoring nonstationarity during the life time of the process under consideration (Bressan [26]). There are numerous attempts in the literature to describe the oscillatory behavior of genetic regulatory networks in a linear fashion using the concept of feedback loops and other methods widely applied in the control theory (Chen et al. [2]; Wang et al. [27]). Unfortunately, the issue of stability of such oscillatory regimes is extremely difficult to explore within the linear theory; therefore, the *requirements* of stability are to be *imposed* on the matrix of coefficients of the linear system. These requirements lead to a set of very complex relationships between coefficients, and it is far beyond the capabilities of existing theories to elucidate a natural mechanism, biochemical, or other, which would surely maintain these relationships throughout the regulatory process. In light of the above described inherent stochastisity of gene expression, the very existence of such a mechanism seems unlikely. However, postulating a fundamentally nonlinear nature of the problem is out of the question. This is seen from the very fact that the "hardware" of the processes underlying gene expression is predominantly the system of biochemical reactions, and, as such, they are adequately described by the nonlinear equations of chemical kinetics. We therefore make the point that the oscillatory behavior of genetic regulatory networks is possible not in spite of but rather owing to the nonlinearity of the system. This means that the nonlinear effects are able to self-organize themselves in such a manner as to automatically keep the system somewhere in close vicinity of the linear oscillatory regime. In what follows, we show that such a scenario is conceivable.

Qualitatively, the approach to the solution of (4) is based on the following two heuristic considerations. First, we draw attention to the "mixing property" of S-functions which may be explained as follows. Suppose that each of $x_1(t), \ldots, x_n(t)$ is represented by linear superpositions of simple periodic processes with a certain set of frequencies. The "forcing" functions in the right-hand side of (4) are the multivariate polynomials of those quasi-periodic processes containing numerous combinatory frequencies along with the original ones; as such these form essentially continuous spectra of the forcing terms. We can reasonably consider functions with such a complex behavior as stochastic processes. Obviously, functions (2) become even more chaotic if the arguments $x_1(t), \ldots, x_n(t)$ are themselves the random processes. On the other hand, in a system having high dimension and a high degree of nonlinearity, deterministic solutions of (4), even if available, would be completely useless. That is why at the very outset we abandon the idea of obtaining the deterministic solutions and assume that $x_1(t), \ldots, x_n(t)$ are stationary stochastic processes. To this end, the goal of the solution of system (4) is reduced to determination of the statistical characteristics of these processes. To obtain these characteristics, we notice that the right-hand side in (5) is the sum of random variables satisfying Lindeberg's conditions (essentially, boundness of the moments: e.g., Loeve [28]). We also allow the random processes $x_1(t), \ldots, x_n(t)$ to be weakly dependent and satisfy the so-called strong mixing conditions (Bradley [29]). The latter assumption is difficult to substantiate theoretically but easy to demonstrate by simulation under the assumptions of our model. Based on these assumptions, we may conclude that the sums in (5) are asymptotically normal, and therefore the random processes $\eta_{ik}(t) = \log Y_{ik}[\mathbf{X}(t)]$ are approximately Gaussian. The second heuristic consideration we engage is that the random forces corresponding to different genes are basically nonlinear combinations of the same set of variables and therefore, generally speaking, are correlated with each other. Figure 1 illustrates this premise (see Appendix A for more details). In this figure, (a) shows 100 separate quasi-periodic oscillations covering a wide spectrum of frequencies formed from the center manifold

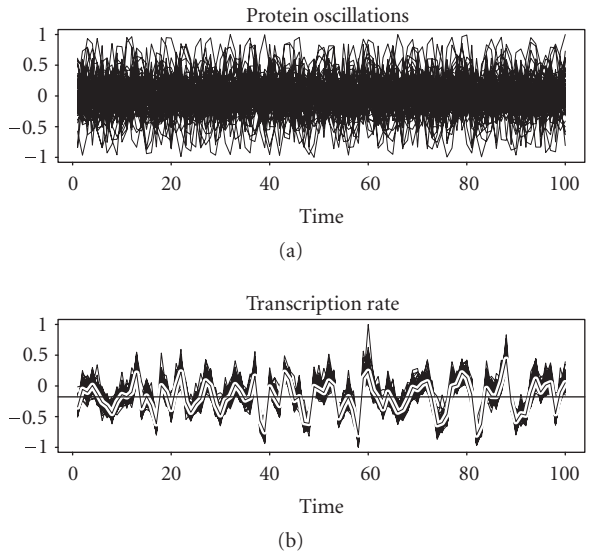

Figure 1: Nonlinear transformation of linear combination of periodic oscillations.

eigenvalues. As shown in (b), corresponding functions $F_i(\mathbf{P}(t))$ in (2) tend to concentrate around a certain stochastic process which is identical for all the genes. This kind of "coherence," that is, the tendency to tightly concentrate around a common limiting process increases as the complexity of the network increases. Statistical analysis shows that the limiting process may be adequately represented as a Gaussian random process. Based on this observation, we assume that all the processes, $\eta_{ik}(t)$, corresponding to different indexes i and k may be replaced by a single Ornstein-Uhlenbeck process (Gardiner [19]), that is, by the process described by the Ito stochastic differential equation (SDE)

$$d\eta_t = -\eta\frac{dt}{\tau_0} + \sqrt{\frac{2}{\tau_0}}\sigma dW_t, \tag{7}$$

where W_t is the unit Wiener process. Considering the asymptotic normality and computing the time averages of both sides in (5), we find that the autocovariance of this process is

$$R_\eta(\tau) = (\lambda^2+\lambda)\sum_{k=1}^{n}\sigma_m^2\exp\left(-|\tau|/\tau_0\right), \tag{8}$$

where $\sigma_m^2 = \mathrm{var}[\ln(x_m)]$ (see Appendix B for details). The correlation radius, τ_0, can be easily estimated computationally through fitting η_t by the first order (i.e., Markov) process.

System (4) is now decoupled on the set of independent equations containing the same "random force," $\exp[\eta(t)]$,

$$\ddot{x}_i + (\beta_i+\delta_i)\dot{x}_i + \beta_i\delta_i x_i = \beta_i\delta_i\exp\left[\eta(t)\right]. \tag{9}$$

Because the process $\eta(t)$ is presumed to be Gaussian, the process $\xi(t) = \exp[\eta(t)]$ is lognormally distributed with the expectation $\exp[\sigma^2/2]$ and variance $\exp(\sigma^2)[\exp(\sigma^2)-1]$. To determine the temporal structure of its autocovariance, we first derive SDE for $\xi(t)$ from (7) and, after some unessential simplifications, find

$$R_e(\tau) = \exp(\sigma^2)\left[\exp(\sigma^2)-1\right]\exp\left(-\frac{\tau}{\tau_0}\frac{\sigma^2}{1-\exp(-\sigma^2)}\right), \tag{10}$$

where

$$\sigma^2 = (\lambda^2+\lambda)\sum_{m=1}^{n}\mathrm{var}\left[\log(x_m)\right]. \tag{11}$$

Comparing (10) and (8), we notice that the correlation radius of the process $\xi(t)$ is always smaller than that of $\eta(t)$, which means that $\xi(t)$ is always closer to white noise than $\eta(t)$. Applying a Fourier transform, (9) can now be easily solved, and the solutions are the stochastic processes with expectations

$$E(x_i) = \beta_i\delta_i\exp(\sigma^2/2), \tag{12}$$

variances

$$\mathrm{var}(x_i) = \frac{\beta_i\delta_i}{\beta_i+\delta_i}\tau_0\frac{\left[\exp(\sigma^2)-1\right]^2}{\sigma^2}, \tag{13}$$

and autocorrelation function

$$\begin{aligned} R_i(\tau) = {} & A_i\exp\left(-|\tau|/\tau_0\right) + B_i\exp\left(-\beta_i|\tau|\right) \\ & + \Delta_i\exp\left(-\delta_i|\tau|\right) \end{aligned} \tag{14}$$

(see Appendix C for details).

The variance, σ^2, should satisfy the conditions of self-consistency derived from the combination of (11) and (13). Simple algebra leads to the transcendental algebraic equation

$$\sigma^2 = (\lambda^2+\lambda)\sum_{i=1}^{n}\ln\left\{1+2\tau_0\frac{\beta_i\delta_i}{\beta_i+\delta_i}\frac{\cosh\sigma^2-1}{\sigma^2}\right\}. \tag{15}$$

In a sense, the solution of the original strongly nonlinear problem is now reduced to solving this equation. Substitution of σ^2 into (12) concludes the procedure of solving the system (4).

5. INTERRELATIONS BETWEEN NONLINEARITY, STABILITY, AND COMPLEXITY

Parameter λ in the Poisson distribution (3) is a natural measure of the complexity of the system. This is because the quantity λn can be interpreted as the average (per gene) number of the proteins participating in the act of transcription. We now formally introduce the "index of complexity," $I_c = (\lambda^2+\lambda)n$. If this index were small, then the vast majority of characteristic roots of the Jacobian matrix would be stable, that is, have negative real parts (see Appendix D for some details regarding characteristic roots). Obviously, this is not the case in reality with I_c usually somewhere between 30 and 100 (Lewin [1]). In the system of such great complexity, a substantial number of the characteristic roots will reside in the right half of the complex plane, thus signifying

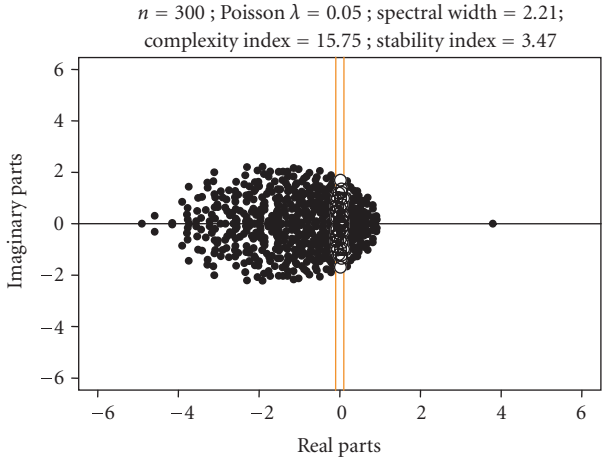

Figure 2: Positions of characteristic roots in case of low complexity.

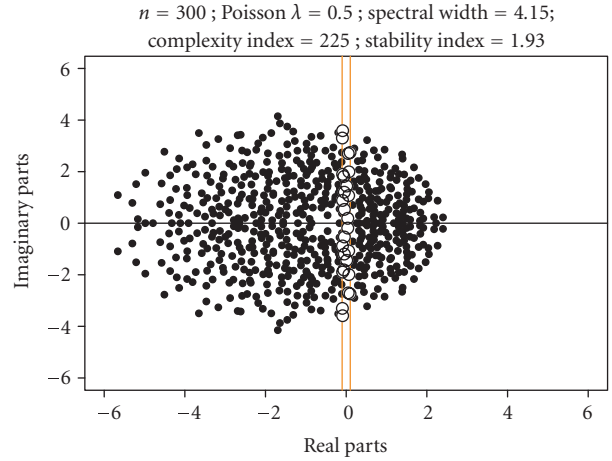

Figure 3: Positions of characteristic roots in case of high complexity.

greater instability of linear oscillatory regime. For this reason, we also define the "index of stability," I_s, assuming that it is the ratio of the number of roots with negative real parts to those with positive ones. Intuitively, it is quite obvious that a certain relationship should exist between the stability, complexity, and spectral width of center manifold. This kind of relationship is not easy to derive theoretically but is fairly easy to demonstrate by simulation (Appendix E). Two examples of the distribution of the characteristic roots over the complex plane for small and large I_c are shown in Figures 2 and 3, respectively. With complexity increasing, the stability decreases, the spectral width of the central manifold increases, thus making the correlation radius, τ_0, smaller and the spectrum of collective "random force," $\xi(t)$, "whiter." Effectively, this means that the more complex the system is, the more favorable the conditions are for applying the proposed approach. Figure 4(a) demonstrates that stability decreases when complexity increases. Figure 4(b) illustrates the fact that the correlation radius of $\xi(t)$ (open circles) is always substantially smaller than that of $\eta(t)$ (solid circles) and both drastically decrease with increasing I_c.

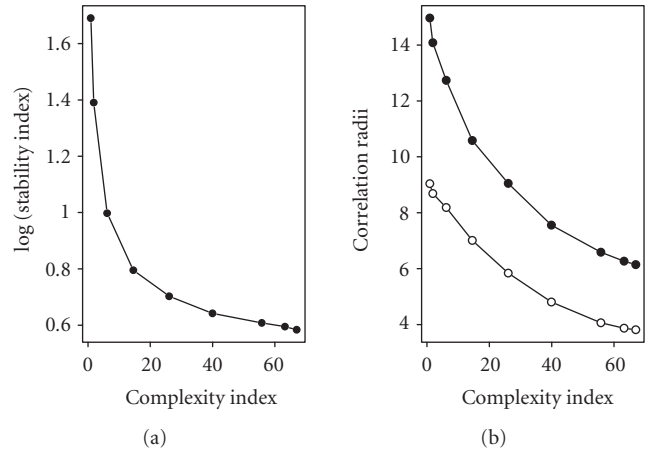

Figure 4: Stability and correlation radii versus complexity of network.

6. INTERRELATIONS BETWEEN TRANSCRIPTION LEVELS AND TRANSCRIPTION RATES

In the model adopted here, the entire gene expression mechanism is seen as being driven by a collective random force which in turn is generated by all the individual transcription-translation events. This kind of "self-consistent" or "average field" approach is widely employed in physics, with such notable examples as Thomas-Fermi equation in atomic physics (Parr and Yang [30]) and Landau-Vlasov equations in the physics of plasma (Chen [31]), to name just a few. Transcription levels (TLs) and transcription rates (TRs) are represented by the quantities r_i and F_i in (1), respectively. In general, since F_i are the stochastic processes generated by the entire network, there are no noticeable correlations between them and any of r_i. Therefore, one cannot expect any substantial similarity between the temporal behavior of TRs and TLs. This conclusion is important for the interpretation of microarray experiments. Also, despite the fact that in our model each mRNA molecule entering the ribosome translates into exactly one protein, there is no similarity between the temporal behaviors of protein and mRNA concentrations. The dissimilarities increase as the network complexity increases because of the longer chain of intermediate events involved in each act of gene expression. To illustrate this fact, Figure 5 depicts the median correlation coefficient (across all the genes) as a function of complexity. As seen from this figure, in the case of high complexity, about a half of all the protein-mRNA pairs is correlated at the level below 0.5. This level of correlation is close to that observed by García-Martínez et al. [32], in their breakthrough experiment where TLs and TRs have been measured simultaneously in budding yeast. It was found the about half of the total 5,500 TLs-TRs pairs turned out not to be correlated with each other. Based on this comparison, we may conclude that the index of complexity of the yeast genetic regulatory network is

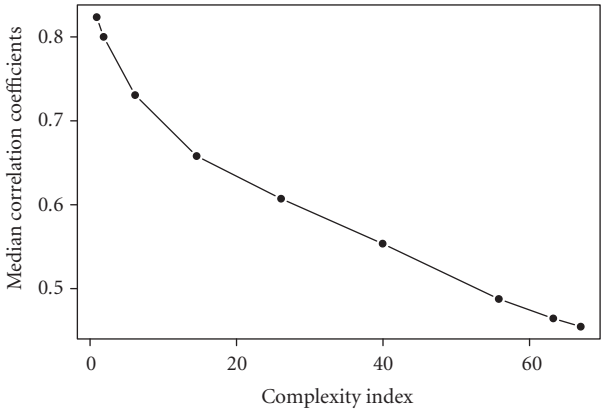

Figure 5: Median correlation coefficients versus complexity.

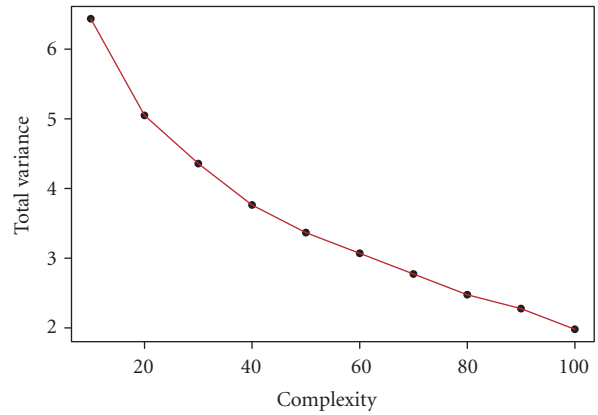

Figure 6: Total variance versus complexity.

about 45–60. Figure 5 shows that in a complex multidimensional system, there are always subsystems which work fast enough to maintain the state of internal synchronization thus displaying apparent steady-state equilibrium. However, this "island" of equilibrium resides amidst the ocean of instability because, due to strong nonlinearity, the system *as a whole* cannot reside in a time-independent steady state. Even an infinitesimally small deviation will cause this state to collapse, and the system will move into the regime of nonlinear stationary stochastic oscillations.

7. INTERRELATIONS BETWEEN COMPLEXITY AND VARIABILITY

It is a fundamental property of living regulatory systems to have precise, highly predictable behavior despite the fact that literally all the components of such systems are intrinsically random and prone to all kinds of failure (McAdams and Arkin [11]). Equation (15) provides an important insight into the nature of this kind of "functional determinism." Simple analysis shows that the solution to this equation exists and is unique if $T_n > I_c\tau_0$, where

$$T_0 = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\beta_i\delta_i}{\beta_i+\delta_i}\right]^{-1}. \tag{16}$$

Parameter T_0^{-1} has a meaning of average, over the entire network, degradation rate of proteins and mRNAs (on this ground we will further refer to T_0 as the "global time of renovation"). If (16) does not hold, then it is not possible to assign any specific variances to the random processes, $x_m(t)$, what essentially amounts to the fact that the system described by (9) may not reside in any stationary oscillatory state. The inequality above, rewritten as $I_c < T_n/\tau_0$, tells us that in a regulatory network with n units there exists an upper limit of complexity determined by two global parameters, that is, by the global time of renovation, T_0, and spectral radius of the collective random force, τ_0. If these parameters reside within the limits required by (16), then (13) may be easily solved numerically. It is quite remarkable that this solution, considered as a function of I_c, is a monotonically decreasing function. Figure 6 shows an example of such dependence $\sigma^2(I_c)$ for the case of the regulatory network with $n = 1000$. According to (13), individual variances, $\mathrm{var}(x_i)$, decrease as well when σ^2 is decreasing. This result suggests the idea that in a large network of fixed size, the precision of regulation increases with the complexity due to an increased number of regulatory loops, despite the presence of numerous pathways of instability.

8. CAUTIONARY NOTES REGARDING MICROARRAY DATA INTERPRETATION

There exist two sets of legitimate quantitative indicators which characterize "gene activity," that is, transcription levels and transcription rates. Microarray experiments provide us with mRNA abundances, that is, transcription levels. What we would rather like to know are the mRNA transcription rates, or the numbers of mRNA copies produced per unit of time. This quantity, if available, would be a more direct measure of gene activity. The difference between TLs and TRs has been repeatedly highlighted in the literature (e.g., Wang et al. [33]); however, it seems to remain largely ignored by the microarray community. As shown above, in a complex regulatory network, transcription level is generally a poor predictor for transcription rates. It is often tacitly assumed in the interpretation of microarray data that there exists some kind of equilibrium between production and degradation of mRNA for each gene separately, in which case a direct proportionality would exist between TLs and TRs. As already mentioned, that may be true with respect to a subset of genes but definitely cannot be true with respect to the entire network. In order to judge which TRs and TLs are in equilibrium and which are not, detailed information about timing of the corresponding biochemical reactions would be required. In principle, in order to cover the entire spectrum of possible chemical oscillations, the sampling rate (number of measurements per unit of time) should be higher than the largest chemical rate among all of the biochemical reactions in the system. Typically, the transcription rate is about five base

pairs per second; therefore, one molecule of mRNA typically requires tens of minutes to be produced (Lewin [1]). The sampling rate capable of capturing the dynamics of these reactions is hardly possible with existing microarray protocols. There are, however, new technologies emerging that combine hybridization with microfluidics which will allow for much higher sampling rates in the foreseeable future (e.g., Peytavi et al. [34]).

Another important implication of the nonlinearity and complexity of a regulatory network is that a living cell cannot reside in a global state of equilibrium, simply because such state cannot be stable. Stochastic oscillatory behavior is in the very nature of the regulatory process. Figuratively speaking, the cell should continuously depart from the point of equilibrium in order to activate the mechanism of returning.

A usual way of thinking in microarray data interpretation is to attribute the differences in mRNA abundances to the cells themselves. However, depending on the frequency of sampling and duration of the sample isolation, the cell can be arrested in different phases of its oscillatory cycle, thus mimicking the differential expression. This means that covariances of expression profiles may be quite different in different time scales. These covariances, usually obtained through cluster analysis or classification, are often used as a basis for the pathway analysis. However, if the temporal dynamics of the regulatory processes is ignored, this analysis may produce misleading results. Many statistical procedures in microarray data analysis, especially in the context of disease biomarker discovery, include the notion that only small subsets of all the genes participate in the disease process and, due to this reason, are actually differentially expressed, while a vast majority of genes are not involved in this process and "do business as usual." Contrary to this notion, it is quite possible that rapidly fluctuating components of the regulatory network are the integral parts of the process as a whole, and their high-frequency variations manifest the preparatory work of supplying the mRNAs for slower processes with bigger amplitudes of variation.

9. DISCUSSION

The model formalized by (1)–(3) possesses a rich variety of features capable of simulating the properties of living cells. We briefly discuss some of them here. Formally speaking, (1)–(3) are written for the entire genome, and therefore, as shown in [25], there is only one global fixed point (i.e., equilibrium). However, if random sets of r_{ikm} and ω_{ik} are clustered into a number of comparatively independent subsets through assigning the gene-specific λ_i, then the entire system (1) is also decomposed into comparatively independent subsystems possessing their own fixed points. In this case, it would be reasonable to expect that the system may switch between different equilibria and produce different oscillatory repertoires. The concept of differentiation, that is, the ability of living cells to perform different functions despite the fact that they have basically identical molecular structures, has been extensively discussed within a number of previously proposed regulatory models (De Jong [5]). The model proposed here has the capability of mimicking the cell differentiation as well. Results of extensive simulations of "tunneling" between different oscillatory repertoires will be published elsewhere.

Regulatory mechanisms in living systems are highly redundant and able to maintain their functionality even when a number of regulatory elements are "knocked out." In the model proposed herein, all the individual transcription-translation subunits are driven by the "collective" random force whose stochastic structure is basically determined by the spectrum of center manifold. Because this spectrum is generated by a large number of individual processes, it follows that if a certain number of genes is "knocked out," then the majority of the remaining genes will not generally change their behavior. For the same reason, the model suggested here has wide basins of attractions (Wuensche [35]), that is, low sensitivity to initial conditions. This property is considered desirable for any formal scheme in models of living systems.

In this work, the *S*-system has been selected to represent nonlinear interactions within genetic regulatory networks for two reasons. First, the *S*-system originates from and adequately represents the dynamics of biochemical reactions, a material basis of all the intracellular processes. Second, the *S*-system is known to be the "universal approximator," that is, to have the capability of representing a wide range of nonlinear functions under mild restrictions on their regularity and differentiability (Voit [36]). However, the *S*-approximation is in no way unique in this sense. Sometimes it would be desirable to maintain a more general view on the nonlinear structure, such as provided by the artificial neural networks (ANN), for example. Our numerical experiments show that a properly constructed ANN retains many of the same features as the *S*-functions. In fact, the only requirement necessary when selecting a nonlinear model is that it must have the "mixing" capability, that is, provide a strong interaction between normal oscillatory modes resulting in stochastic-like behavior of $\mathbf{F}(\mathbf{p})$.

In this work an attempt has been made to directly link the stochastic properties of random fluctuations in the nonlinear regulatory system to the spectrum of quasi-periodic oscillations near the point of equilibrium. Currently, we are able to offer only heuristic considerations and numerical simulation in support of this viewpoint. Attempts to create a rigorous theoretical basis for extension of center manifold theory to stochastic systems are still very rare, highly involved mathematically, and do not seem to be readily digestible in practical applications (Boxler [37]). Intuitively, however, the link between the center manifold theory and stochastic dynamics seems to be quite natural. As shown above, under certain conditions, variance of fluctuations around the equilibrium point may decrease with increase in the network size, which means that, despite strong nonlinearity, the system may nevertheless mostly reside in close vicinity of the equilibrium. Therefore, it seems reasonable to think that the spectrum of nonlinear oscillations is somewhat similar to the spectrum of linear oscillations but with distortions of amplitudes and phases introduced by nonlinear interactions between linear

oscillatory modes. Figuratively speaking, a strong nonlinear "pressure" of a very big network is what forces the system to be nearly linear. This intriguing hypothesis is currently among the priorities of the author's future research.

In the natural sciences, it is always desirable to have a way of experimental verification of theoretical results. However, it would be risky to claim that any of the existing models are already mature enough to generate a verifiable prediction regarding biological behavior of the genetic regulatory networks. So far it is not even quite clear what kind of features or criteria should be selected to compare theory and experiment. It is our personal opinion that among the most important questions to elucidate are the ones pertaining to the global structure of the network connectivity, that is, whether the network under consideration is "scale-free," "exponential," or intermediate (Newman [38]). Equally significant are the questions pertaining to the spectrum of temporal variations of the chemical constituents. In general, whatever the criteria are selected for comparison, attention should be primarily focused on the characteristics of global behavior, rather than on the intricacies of the behavior of individual genes.

APPENDICES

A. MIXING PROPERTY AND COHERENCE

Let us assume that $x_i(t) = a_i \cos[\nu_i t + \varphi_i(t)]$, where frequencies ν_i are randomly selected from the center manifold spectrum and a_i are some positive numbers. Also, let us assume that the phases, $\varphi_i(t)$, are independent stationary Gaussian delta-correlated random processes with identical variances σ_φ^2. In this simulation, we assume that the random fluctuations of phases are weak, that is, $\sigma_\varphi \ll 2\pi$; therefore, the oscillations $x_i(t)$ are very close to being purely periodic. For the fixed set of coefficients ω_{ik}, r_{ikm}, and a_i, we compute the set of response functions

$$F_i(t) = \sum_{k=1}^{L_i} \omega_{ik} \exp\left[\sum_{m=1}^{n} r_{ikm} x_m(t)\right]. \tag{A.1}$$

The goal of this computation is to demonstrate the following.

(1) Although the trajectories, $x_i(t)$, are independent random processes, nevertheless the random "forces," $F_i(t)$, are highly correlated, that is, coherent.

(2) Although the trajectories, $x_i(t)$, are almost deterministic, that is, have large correlation radii, nevertheless random "forces," $F_i(t)$, are chaotic, that is, have small correlation radii.

(3) Although random processes, $x_i(t)$, are very far from being Gaussian, nevertheless the logarithms of random "forces," $\log[F_i(t)]$, are very close to Gaussian. Graphical representations of the functions $x_i(t)$ and $\log[F_i(t)]$ are shown in Figure 1. Usually n is in thousands, but to make the curves visually distinguishable we have selected $n = 100$, $\lambda = 0.5$, and $\sigma_\varphi = \pi/16$. Parameters associated with this figure are given in Table 1.

The following definitions have been used in these calculations.

Table 1

	Cross-correlation	Correlation radius	Kurtosis
$x_i(t)$	< 0.001	18.9	-1.41
$\log[F_i(t)]$	0.706	1.23	0.18

(1) Correlation radius, $\tau_0 = \int_0^\infty |r(\tau)| d\tau$, where $r(\tau)$ is the autocorrelation function defined as

$$r(\tau) = E[x^*(t)x^*(t+\tau)]/E[x^*(t)x^*(t)], \qquad x^*(t) = x(t) - E[x(t)]. \tag{A.2}$$

(2) Cross-correlation, $R_{ij} = E[x_i^*(t)x_j^*(t)]/\sqrt{E[(x_i^*)^2]E[(x_j^*)^2]}$.

Under the condition of stationarity, $r(\tau)$ and R_{ij} are independent on t. Assuming ergodicity, the expectations may be computed as time averages: $E[g(t)] = \lim_{T\to\infty}[T^{-1}\int_0^T g(t)dt]$.

Note that (a) both $x_i(t)$ and $\log[F_i(t)]$ have symmetric density distributions; (b) distribution of periodic functions with infinitesimally small fluctuations of phase is the arcsine distribution with kurtosis equal to $-\sqrt{2}$; (c) closeness of the distribution of $\log[F_i(t)]$ to normal is signified by the closeness of its kurtosis to zero.

B. DERIVATION OF (8)

The goal here is to find statistical characteristics of the random processes

$$Y_{ik}[x_1(t), \ldots, x_n(t)] = \exp(S_{ik}), \quad S_{ik} = \sum_{m=1}^{n} r_{ikm} \log[x_m(t)]. \tag{B.1}$$

Under the assumptions that $y_m(t) = \log[x_m(t)]$ have finite moments (Lindeberg's condition), the sums S_{ik} are asymptotically normal with expectations

$$e_{ik} = E_y(S_{ik} \mid r_{ikm}) = \sum_{k=1}^{n} r_{ikm} E(\log[x_m(t)]) = \sum_{k=1}^{n} r_{ikm}\mu_m \tag{B.2}$$

and variances, θ_{ik}^2,

$$\theta_{ik}^2 = \mathrm{var}_y(S_{ik} \mid r_{ikm}) = \sum_{p,q}^{n,n} r_{ikp} r_{ikq} \mathrm{cov}[y_p(t) y_q(t)]. \tag{B.3}$$

Therefore,

$$S_{ik}(t) = e_{ik} + \sqrt{\theta_{ik}^2}\eta_{ik}(t), \tag{B.4}$$

where $\eta_{ik}(t)$ are standard normal Gaussian processes with yet unknown autocorrelation structures. Note that y_m are not required to be statistically independent; weak dependence satisfying the "strong mixing conditions" is sufficient for asymptotic normality (Bradley [29]). Since $S_{ik}(t)$ asymptotically normal, the $\exp[S_{ik}(t)]$ are asymptotically lognormal

with expectations and variances equal to

$$E(Y_{ik} \mid r_{ikm}) = \exp(e_{ik} + 0.5\theta_{ik}^2),$$
$$\operatorname{var}(Y_{ik} \mid r_{ikm}) = \exp(\theta_{ik}^2)[\exp(\theta_{ik}^2) - 1]. \tag{B.5}$$

We now need to evaluate the sums in (B.2), (B.3), and for this purpose we use again the central limit theorem. We notice that when n is sufficiently large

$$e_{ik} \approx E_r(e_{ik}) + \sqrt{\operatorname{var}_r(e_{ik})}\,\zeta_{ik},$$
$$\theta_{ik}^2 \approx E_r(\theta_{ik}^2) + \sqrt{\operatorname{var}_r(\theta_{ik}^2)}\,\xi_{ik}, \tag{B.6}$$

where ζ_{ik} and ξ_{ik} are standard normal iid, and subscript r indicates averaging with respect to distribution of r_{ikm}. Simple algebra provides the following results:

$$E_r(e_{ik}) = \lambda \sum_{m=1}^{n} \mu_m; \qquad \operatorname{var}_r(e_{ik}) = \lambda \sum_{m=1}^{n} \mu_m^2, \tag{B.7}$$

$$E_r(\theta_{ik}^2) = \lambda \sum_{p=1}^{n} \sigma_p^2 + \lambda^2 \sum_{p,q}^{n,n} \operatorname{cov}(y_p y_q), \tag{B.8}$$

$$\operatorname{var}_r(\theta_{ik}^2) = 4\lambda^3 \sum_{p,q,\nu}^{n,n,n} \operatorname{cov}(y_p y_q)\operatorname{cov}(y_p y_\nu) + \lambda^2 \sum_{p,q}^{n,n} [5\sigma_p^2 + \operatorname{cov}(y_p y_p)]\operatorname{cov}(y_p y_q) + \lambda \sum_{p=1}^{n} \sigma_p^4. \tag{B.9}$$

Due to asymptotic normality, the terms containing variances in (B.6) have order $O(n^{1/2})$ and may be neglected when compared with the expectation terms having the order $O(n)$. If, in addition to that, we also neglect the cross-covariances (not required in numerical computations!), that is, assume that $\operatorname{cov}(y_p y_q) = \sigma_p \sigma_q \delta_{pq}$, then we come out with (8) in the main text,

$$S_{ik}(t) = \lambda \sum_{m=1}^{n} \mu_m + \left[(\lambda^2 + \lambda) \sum_{m=1}^{n} \sigma_m^2 \right]^{1/2} \eta_{ik}(t). \tag{B.10}$$

C. DERIVATION OF (11)–(13)

We calculate statistical characteristics of the processes $x_i(t)$ satisfying differential equations (9), where $\eta(t)$ is the OUP satisfying the SDE (7). Spectral density of the latter process is (Gardiner [19])

$$\Phi(\omega) = \frac{\sigma^2 \tau_0}{\pi} \frac{1}{1 + \omega^2 \tau_0^2}. \tag{C.1}$$

We introduce new processes, $\xi_i(t) = \beta_i \delta_i \{\exp[\eta(t)] - \exp(\sigma_\eta^2/2)\}$. These processes satisfy SDEs

$$d\xi_i(t) = -\frac{1}{\tau_0} \frac{\sigma_\eta^2}{1 - \exp(\sigma_\eta^2)} \xi_i(t)dt + \sqrt{\frac{2}{\tau_0}} \beta_i \delta_i \sigma_\eta \exp(\sigma_\eta^2)\, dW_t. \tag{C.2}$$

Applying Fourier transform to (9) (index i is temporarily omitted) we find

$$R_x(\tau) = \frac{D}{2}\left[\frac{1}{\delta} \frac{\exp(-\delta|\tau|)}{(\beta^2 - \delta^2)(\chi^2 - \delta^2)} + \cdots \right], \tag{C.3}$$

where ellipsis stands for the terms obtained by cyclic permutations of β, δ, and χ with

$$D = \frac{2}{\tau_0} \beta_i^2 \delta_i^2 \sigma_\eta^2 \exp(2\sigma_\eta^2),$$
$$\chi = \frac{1}{\tau_0} \frac{\sigma_\eta^2}{1 - \exp(\sigma_\eta^2)}. \tag{C.4}$$

Since $\beta_i \tau_0 \ll 1$ and $\delta_i \tau_0 \ll 1$ for the majority of genes, we find that

$$\operatorname{var}(x_i) = R_i(0) = \frac{D}{2} \frac{1}{\chi^2 \beta_i \delta_i (\beta_i + \delta_i)} = \frac{\beta_i \delta_i}{\beta_i + \delta_i} \tau_0 \frac{[\exp(\sigma^2) - 1]^2}{\sigma^2}. \tag{C.5}$$

D. JACOBIAN MATRIX AND EIGENVALUES

In (1), let $\{p_i^0, r_i^0\}$ be the equilibrium (fixed) point in the $2n$-dimensional phase space of the system (1). At this point $\mathbf{F}(\mathbf{p^0}) = \boldsymbol{\beta}\mathbf{r^0}$, $\boldsymbol{\delta}\mathbf{p^0} = \boldsymbol{\gamma}\mathbf{r^0}$. Let $\{p_i', r_i'\}$ be the deviations from this point, then the quantities $\xi_i = p_i'/p_i^0$ and $\rho_i = r_i'/r_i^0$ satisfy the equations

$$\frac{d\xi_i}{dt} = \delta_i(\rho_i - \xi_i), \qquad \frac{d\rho_i}{dt} = \beta_i\left(\sum_{k=1}^{n} \Omega_{ik}\xi_k - \rho_k \right), \tag{D.1}$$

where $\boldsymbol{\Omega} = \|\partial\mathbf{F}/\partial\mathbf{p}\|$ is the Jacobian matrix. Compound matrix of the system (D.1) (not shown to save space) is the basis for the calculation of eigenvalues. Because $\boldsymbol{\Omega}$ is a nonsymmetric matrix with positive elements, its eigenvalues are complex numbers having, generally speaking, both positive and negative real parts.

Existence of a fixed point is the necessary condition for existence of a stationary solution. Provided all the coefficients in (1) are known, the search for the fixed point $\mathbf{F}(\mathbf{p^0}) = (\boldsymbol{\beta\delta}/\boldsymbol{\gamma})\mathbf{p^0}$ may be a difficult task by itself. In order to avoid this problem, which is not central in our consideration, we postulate that a unique equilibrium point for protein concentration $\mathbf{p^0}$ does exist and is the part of the model parameterization. With this reparameterization, vectors $\mathbf{r^0}$ and $\boldsymbol{\gamma}$ are expressed through $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and $\mathbf{p^0}$, as seen in (4), (6), and (D.1).

Table 2

Size of network, n	Complexity index, I_c		
—	10	40	100
100	1.30	1.57	2.73
500	1.52	2.43	3.12
1000	1.47	2.51	3.56

E. RELATION BETWEEN COMPLEXITY AND SPECTRAL WIDTH OF CENTER MANIFOLD

Table 2 shows that for a fixed size of the network, the half-width of the center manifold spectrum grows approximately as the logarithm of the complexity index.

REFERENCES

[1] B. Lewin, *Genes VIII*, Prentice-Hall, Upper Saddle River, NJ, USA, 2004.

[2] T. Chen, H. L. He, and G. M. Church, "Modeling gene expression with differential equations," in *Pacific Symposium on Biocomputing (PSB '99)*, pp. 29–40, Mauna Lani, Hawaii, USA, January 1999.

[3] S. Zumdahl, *Chemical Principles*, Houghton Mifflin, New York, NY, USA, 2005.

[4] J. T. Kim, T. Martinetz, and D. Polani, "Bioinformatic principles underlying the information content of transcription factor binding sites," *Journal of Theoretical Biology*, vol. 220, no. 4, pp. 529–544, 2003.

[5] H. De Jong, "Modeling and simulation of genetic regulatory systems: a literature review," *Journal of Computational Biology*, vol. 9, no. 1, pp. 67–103, 2002.

[6] A. Sorribas and M. A. Savageau, "Strategies for representing metabolic pathways within biochemical systems theory: reversible pathways," *Mathematical Biosciences*, vol. 94, no. 2, pp. 239–269, 1989.

[7] L. E. Maquat, "Nonsense-mediated mRNA decay in mammals," *Journal of Cell Science*, vol. 118, no. 9, pp. 1773–1776, 2005.

[8] B. Lemon and R. Tjian, "Orchestrated response: a symphony of transcription factors for gene control," *Genes & Development*, vol. 14, no. 20, pp. 2551–2569, 2000.

[9] D. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.

[10] J. Elf and M. Ehrenberg, "Fast evaluation of fluctuations in biochemical networks with the linear noise approximation," *Genome Research*, vol. 13, no. 11, pp. 2475–2484, 2003.

[11] H. H. McAdams and A. Arkin, "It's a noisy business! Genetic regulation at the nanomolar scale," *Trends in Genetics*, vol. 15, no. 2, pp. 65–69, 1999.

[12] H. H. McAdams and A. Arkin, "Stochastic mechanisms in gene expression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 3, pp. 814–819, 1997.

[13] L. Cai, N. Friedman, and X. S. Xie, "Stochastic protein expression in individual cells at the single molecule level," *Nature*, vol. 440, no. 7082, pp. 358–362, 2006.

[14] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, "Gene regulation at the single-cell level," *Science*, vol. 307, no. 5717, pp. 1962–1965, 2005.

[15] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of the Atmospheric Sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[16] D. Zhang, L. Gyorgyi, and W. R. Peltier, "Deterministic chaos in the Belousov-Zhabotinsky reaction: experiments and simulations," *Chaos*, vol. 3, no. 4, pp. 723–745, 1993.

[17] A. J. Lotka, *Elements of Physical Biology*, Williams and Wilkins, Baltimore, Md, USA, 1925.

[18] J. M. Bower and H. Bolouri, Eds., *Computational Modeling of Genetic and Biochemical Networks*, MIT Press, Cambridge, Mass, USA, 2001.

[19] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Springer, New York, NY, USA, 1983.

[20] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Random Boolean network models and the yeast transcriptional network," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 25, pp. 14796–14799, 2003.

[21] M. Ptashne, "Regulated recruitment and cooperativity in the design of biological regulatory systems," *Philosophical Transactions of the Royal Society A*, vol. 361, no. 1807, pp. 1223–1234, 2003.

[22] J. Carr, *Applications of Center Manifold Theory*, Springer, New York, NY, USA, 1981.

[23] M. Savageau and E. Voit, "Recasting nonlinear differential equations as S-systems: a canonical nonlinear form," *Mathematical Biosciences*, vol. 87, pp. 83–115, 1987.

[24] L. Perko, *Differential Equations and Dynamical Systems*, Springer, New York, NY, USA, 3rd edition, 2001.

[25] D. Lewis, "A qualitative analysis of S-systems: Hopf bifurcation," in *Canonical Nonlinear Modeling. S-System Approach to Understanding Complexity*, E. Voit, Ed., pp. 304–344, Van Nostrand Reinhold, New York, NY, USA, 1991.

[26] A. Bressan, "Tutorial on the Center Manifold Theory," 2003, SISSA, Trieste, Italy, http://www.math.psu.edu/bressan/PSPDF.

[27] R. Wang, Z. Jing, and L. Chen, "Modelling periodic oscillation in gene regulatory networks by cyclic feedback systems," *Bulletin of Mathematical Biology*, vol. 67, no. 2, pp. 339–367, 2005.

[28] M. Loeve, *Probability Theory*, The University Series in Higher Mathematics, Van Nostrand, New York, NY, USA, 1963.

[29] R. Bradley, "Basic properties of strong mixing conditions. A survey and some open questions," *Probability Surveys*, vol. 2, pp. 107–144, 2005.

[30] R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules*, Oxford University Press, New York, NY, USA, 1989.

[31] F. Chen, *Introduction to Plasma Physics and Controlled Fusion*, Plenum Press, New York, NY, USA, 1984.

[32] J. García-Martínez, A. Aranda, and J. E. Pérez-Ortín, "Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms," *Molecular Cell*, vol. 15, no. 2, pp. 303–313, 2004.

[33] W. Wang, J. M. Cherry, D. Botstein, and H. Li, "A systematic approach to reconstructing transcription networks in Saccharomyces cerevisiae," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 26, pp. 16893–16898, 2002.

[34] R. Peytavi, F. R. Raymond, D. Gagné, et al., "Microfluidic device for rapid (< 15 min) automated microarray hybridization," *Clinical Chemistry*, vol. 51, no. 10, pp. 1836–1844, 2005.

[35] A. Wuensche, "Genomic regulation modeled as a network with basins of attraction," in *Pacific Symposium on*

Biocomputing (PSB '98), vol. 3, pp. 89–102, Maui, Hawaii, USA, January 1998.

[36] E. Voit, Ed., *Canonical Nonlinear Modeling. S-System Approach to Understanding Complexity*, Van Norstand Reinhold, New York, NY, USA, 1991.

[37] P. Boxler, "A stochastic version of center manifold theory," *Probability Theory and Related Fields*, vol. 83, no. 4, pp. 509–545, 1989.

[38] M. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

Simon Rosenfeld has his M.S. degree in physics (major in molecular physics) from Lomonosov State University, Moscow, Russia and Ph.D. degree in physics (major in upper atmosphere and space physics from the National Institute of Applied Geophysics, Moscow, Russia. He published numerous works in aerodynamics of ultrarefied supersonic flows, satellite remote sensing, GPS radio-occultation tomography, applied statistical methods, time-series analysis, physical acoustics, and chemical thermodynamics. Substantial part of Dr. Rosenfeld's research experience is in nonlinear wave dynamics with applications to stochastic phenomena in atmospheric and oceanic waves. During the last seven years he has been employed with the National Cancer Institute, Bethesda, MD, USA. His current research activity is in cancer-related functional genomics and systems biology with focus on microarray data analysis, computational proteomics, bioinformatics, and dynamics of genetic regulatory networks.