Neural network model of gene expression

JIŘÍ VOHRADSKÝ
Institute of Microbiology, CAS,142 20 Prague, Czech Republic

ABSTRACT **Many natural processes consist of networks of interacting elements that, over time, affect each other's state. Their dynamics depend on the pattern of connections and the updating rules for each element. Genomic regulatory networks are networks of this sort. In this paper we use artificial neural networks as a model of the dynamics of gene expression. The significance of the regulatory effect of one gene product on the expression of other genes of the system is defined by a weight matrix. The model considers multigenic regulation including positive and/or negative feedback. The process of gene expression is described by a single network and by two linked networks where transcription and translation are modeled independently. Each of these processes is described by different network controlled by different weight matrices. Methods for computing the parameters of the model from experimental data are discussed. Results computed by means of the model are compared with experimental observations. Generalization to a 'black box' concept, where the molecular processes occurring in the cell are considered as signal processing units forming a global regulatory network, is discussed.—Vohradský, J. Neural network model of gene expression. *FASEB J.* 15, 846–854 (2001)**

Key Words: genetic network · regulation of gene expression · mathematical modeling

The information for constructing and maintaining the molecular components of a living cell lies within its genes. Genes directly encode proteins that make up the cell and directly or indirectly participate in the signaling and control of processes necessary for life. The genes are transcribed to RNAs, which are translated to proteins.

Transcriptional regulation is conferred through the action of gene products that bind to each gene's transcriptional start site. These proteins bind directly to DNA and influence gene expression by altering the activity of transcriptional machinery. The activity of transcription factors is controlled by other gene products. Thus, the transcription of a particular gene can be viewed as a combinatorial action of products of other genes or even of its own gene product. Once the expressed gene is translated into a functional gene product, it affects the state of the cell and may directly or indirectly influence the expression of other genes. In this way the expression state of the cell is controlled by a set of genes expressed at a given moment. These genes regulate the transcription of cell genes leading to a new state. The new state is defined by the concentration of currently expressed gene products, and so on. This means that the state of transcription and translation in the cell is continuously updated from one state to another in a feed-forward-like circuit, where the current state depends on the previous one, which depends on the preceding one, etc.

This approach can be formalized as a concept of 'genetic network' where all genes and their products participating in one regulatory event are members of a network. The transcriptional state of a gene is defined, in principle, by the states of all other genes in the network (**Fig. 1**). This model can be simplified by defining the expression of genes as either on or off {0,1} (1–3). The response of a gene in the model is then defined by Boolean rules that combine the states of expression of genes controlling the particular gene (e.g., gene A is expressed if the formula C AND (B OR D) for genes B, C, D is true). When the Boolean rules for all members of the model are known, the terminal state of the network can be computed. The control of gene expression can thus be simplified to wiring between genes and a set of Boolean rules associated with the synapses (using neural network terminology) connecting the nodes. The synapses and the rules define the next state of the network. From the theory of random Boolean networks, it follows that the final state of the network (the state after a finite number of state updates) is defined by either a point attractor or a basin attraction (1, 4). A point attractor defines a stable state that does not change with the update of the network. The basin of attraction represents a set comprising a limited number of regularly changing states (limited cycle). Several methods to identify the network from experimental data have been suggested (3, 5).

This Boolean simplification ignores the genes that have different regulatory effects depending on their level of expression. Furthermore, these networks cannot address those genes that influence the transcription of various genes to different degrees. The model has therefore been generalized to incorporate different levels of expression of regulatory genes and different degrees of the regulatory effect on a given gene. D'Haesleer (6) suggested a linear model of gene control where the expression level of a gene is defined by

[1] Correspondence: Institute of Microbiology, CAS, Videňská 1083, CZ-14220 Prague, Czech Republic. E-mail: vohr@biomed.cas.cz

0892-6638/01/0015-0846 © FASEB

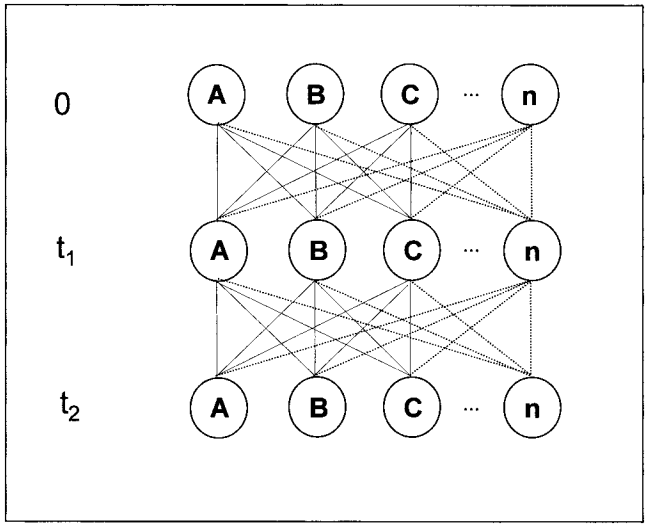


Figure 1. Formal description of transcriptional regulation by a neural network. Each node of the network represents one gene; the connection between nodes represents the regulatory actions of one gene on another. In principle, all genes can control all others (fully connected network); in reality, only a few genes control the expression of one particular gene. The state of the network is updated in a stepwise manner; the state of gene expression at time t_i is determined by the state at time t_{i-1}.

the linear combination of expression levels of other genes weighted by a value reflecting the level of their regulatory effect. The weight matrix then summarizes the interactions among all members of the set. The methods for computing the weight matrix from experimentally measured expression time series are presented. The authors also discuss limits of the approach based on a linear model that oversimplifies the expression kinetics. Weaver et al. (7) extended this approach by introducing a nonlinear 'squashing function' modifying the output from the linear combination of weights and expression levels of the regulatory genes. The additive principle for updating the state of the network was kept. The method of reconstruction of the weight matrix from experimental data was also analyzed. Genetic networks have also been modeled by a set of differential equations (8).

Each such model should be able not only to fit the experimentally measured expression time series 'profiles', but also to correctly describe the underlying process. As the weight matrix has to be derived from the experimental data according to a certain model, an invalid model will produce an invalid matrix and hence an invalid reconstruction of the regulatory connections. Unfortunately, little (if any) experimental data exist that would allow the model to be tested. Nevertheless, rapid developments in DNA chip technology and quantitative proteomics that can provide such data promise to fill this gap. In the current situation, the model can be tested only on simplified artificial examples where the output can be estimated.

MODEL

The new model is based on the assumption that the regulatory effect on the expression of a particular gene can be expressed as a neural network (Fig. 1), where each node represents a particular gene and the wiring between the nodes define regulatory interactions. Each layer of the network represents the level of expression of genes at time t. Output of a node at time $t+\Delta t$ can be derived from the expression levels at the time t (y_j) and connection weights (w_{ij}) of all genes connected to the given gene. Thus g_i, a regulatory effect to gene *i*, is

$$g_i \approx \sum_j w_{ij} y_j. \qquad (1)$$

We propose that this regulatory effect is transformed by a sigmoidal transfer function to the interval <0,1>. The regulatory effect is then defined as

$$g_i = \{1 + \exp[-(\sum_j w_{ij} y_j + b_i)]\}^{-1} \qquad (2)$$

where b_i represents an external input that can be interpreted as a reaction delay parameter. High negative and/or positive values of *b* result in a low influence of factors given in the weight matrix. The actual rate of expression is then modulated by a multiplicative constant k_{1i}, which represents the maximal expression rate of a given gene.

Let the rate of expression of gene *i* be given by the regulatory effects of other genes ρ_i and the effect of degradation

$$dz_i/dt = \rho_i - x_i \qquad (3)$$

The degradation effect x_i is modeled by the kinetic equation of a first order chemical reaction $x_i = k_{2i} \cdot z_i$, and ρ_i represents the regulatory effect reflected in a variable g_i ($\rho_i = k_{1i} g_i$). The constant k_2 represents the rate constant of degradation of the gene product *i* and k_1 is its maximal rate of expression. The whole model has the form:

$$\frac{dz_i}{dt} = k_{1i} \frac{1}{1 + \exp[-(\sum_j w_{ij} y_j + b_i)]} - k_{2i} z_i \qquad (4)$$

Equation 4 represents a special case of a class of recurrent neural networks that can be described by the general formula (9)

$$\tau_i \frac{dy_i}{dt} = f_i(\sum_j w_{ij} y_j - \theta_i) - y_i \qquad (5)$$

where f_i is a nonlinear transfer function, *w* is the weight matrix, and θ is an external input to the node *i*. These types of networks have been used as associative memories and as models of brain activity, and their dynamics have been thoroughly studied. The output of the network at time *t* depends on the weight matrix *w*. If the matrix is symmetric ($w_{ij} = w_{ji}$) the network reaches a stationary state in finite time. If the connection weight

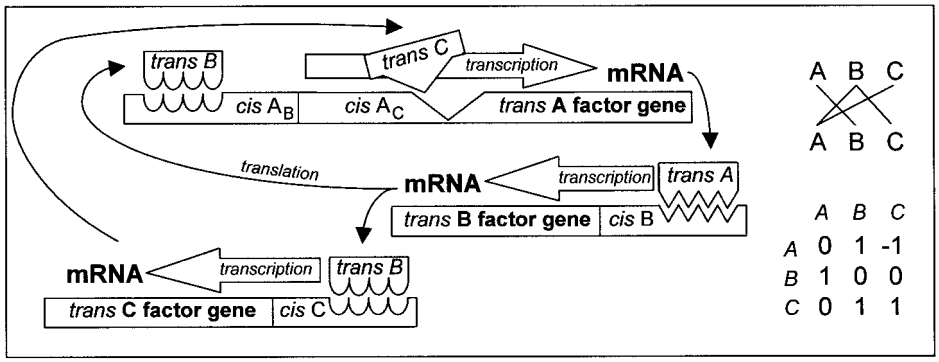


Figure 2. Gene regulation with positive and negative feedback and its representation by a wiring diagram and Boolean weight matrix. Genes code for proteins, which in turn control the expression of other genes. A controls expression of B, which controls C and A. C negatively controls A. The cybernetic formalization is represented by the wiring diagram, shown on the right. The computational rules are given by the Boolean weight matrix. The Boolean rules for this circuit can be written as A=B not C, B=A, C=B.

matrix is far from being symmetric, the state transition of the network can lead to a point attractor (stationary state) or can become oscillatory (limit cycle) or even chaotic, depending on the complexity of the network.

RESULTS

Gene regulation with positive and negative feedback

Let us consider a simple model situation published in ref 1 (**Fig. 2**). Here the product of gene A controls the expression of gene B, which initiates the expression of gene C. Gene B induces the expression of gene A forming positive feedback. Gene C in turn negatively controls the expression of gene A, forming negative feedback. This is a simple feedback circuit that can be characterized by a network and a weight matrix (Fig. 2). Considering only one isolated circuit, the state of the network can reach two attractors, depending on the initial conditions: a point attractor, where the action of the network is stopped; or a two-point attractor perpetually oscillating between two states.

In the case of multiple copies of the circuit, the reaction of the system depends not only on the state of the genes (on/off), but also on the concentration of each compound controlling the process. It is evident that with increasing concentration of the product of A, expression of B will increase with a delay proportional to the reaction kinetics of the transcriptional/translational machinery of gene B. This is also valid for the B/C pair. Therefore, an increased concentration of the A product will be followed by a delayed increase of B and, with another delay, by an increase of C. This scheme will be maintained until the concentration of C product is high enough to block the synthesis of A, reflected in a decrease of the concentration of A, since all products are continuously being degraded. In turn, a decreasing A directly causes a decrease of B and consequently of C. If the concentration of A or B is close to 0, the whole process is terminated. If not, after the decrease of C, A is synthesized again; the whole cycle starts up again and is repeated *ad infinitum*.

This is a qualitative analysis of the behavior of the model of expression shown in Fig. 2. We have reconstructed the response of the network from Fig. 2. using the following weight matrix

$$w_{ij} = \begin{bmatrix} 0 & 10 & -10 \\ 10 & 0 & 0 \\ 0 & 10 & 0 \end{bmatrix} \qquad (6)$$

qualitatively equivalent to the Boolean weight matrix from Fig. 2, and solved the system of equations numerically. The results are shown in **Fig. 3**.

The parameter b (Eq. 2) represents the delay with which the transcription/translation process of one gene reacts to the regulatory impulse from other gene products. High negative values mean a long delay. If all

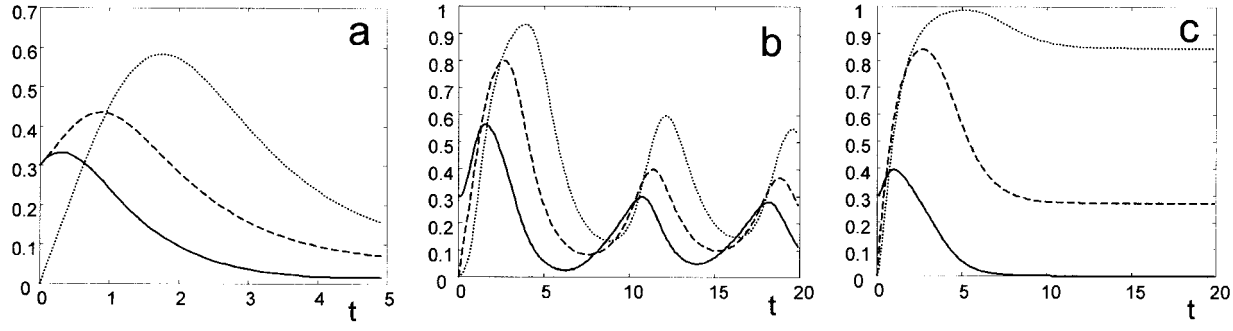
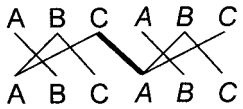


Figure 3. Expression patterns of genes A (solid line), B (dashed), C (dotted) for different values of the reaction delay parameter *b*. *a, c)* The reaction delay parameter *b* is the same for all members of the system; expression stops after some time at constant levels. *b)* If expression of gene A is faster, the system terminates in an oscillatory pattern. Example shown in Fig. 2 was used to calculate the kinetic profiles.



0	10	-10	0	0	0
10	0	0	0	0	0
0	10	0	0	0	0
0	0	10	0	10	-10
0	0	0	10	0	0
0	0	0	0	10	0

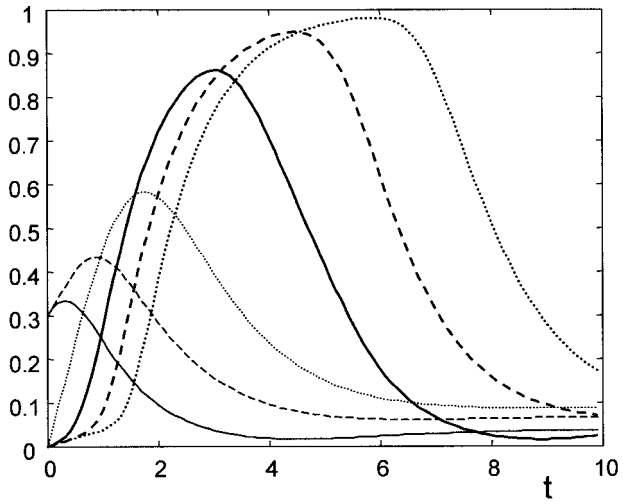


Figure 4. Wiring diagram, weight matrix, and expression profile for two networks from Fig. 2 connected by insertion of interaction between A of the second network and C of the first network. As expected, the second network reacts in the same way as the first one but with a delay, waiting for the signal from the first network. A (solid line), B (dashed), C (dotted). Curves corresponding to the second network are represented by heavier lines.

parameters are set to the same values (Fig. 3*a, c*), the system terminates at a point attractor. If the reaction of one component is faster, e.g., gene A (Fig. 3*b*), then after the concentration of the C product (which blocks transcription of gene A) is low enough, transcription from gene A is quickly renewed and the cycle starts up again. The system terminates in a limit cycle: the expression levels of all components oscillate with time. All other parameters controlling the rate of degradation and expression were kept at 1.

Connected networks

Let us combine two networks by introducing a connection between C of the first network and A of the second one (**Fig. 4**). We would expect that the output of network 2 would be delayed, waiting for a sufficient concentration of the C product. If the weight matrix for the second network is the same, the response of the second network would be delayed but otherwise qualitatively identical to that of network 1. This is exactly what we observe in the kinetic profiles of Fig. 4.

Linked networks: transcriptional and translational control of gene expression

The expression of a gene in the previous examples is influenced by the regulatory effect of other proteins involved in the regulatory event, and the rate of protein formation was replaced by a nonspecific sigmoid curve representing both translation and transcription. If we separate the two processes (**Fig. 5**), transcription and translation can be considered as controlled in two separated compartments. This concept was originally developed by Hoel (10) and Yagil (11) and extended by Hargrowe and Schmidt (12) (readers interested in the validity of the assumptions used in the model should examine these references and their bibliographies). The process can be described by two different networks connected by one synapse. The connection is maintained by the mRNA of the specific gene. If we consider that synthesis of mRNA corresponding to gene i (r_i) is controlled by the regulatory proteins A. . . n (Fig. 5) and the rate of its synthesis again follows a sigmoid, the rate of mRNA accumulation will be described by the equation:

$$\frac{dr_i}{dt} = k_{1ri} \cdot f(p_{j=1..n},\ w_r) - k_{2ri} \cdot r_i \qquad (7)$$

where r_i is the concentration of $mRNA_i$, k_{1ri} and k_{2ri} are the maximal rates of mRNA synthesis and degradation, respectively. Function f is the sigmoid transfer function (as in Eq. 2) 'squashing' the influence of proteins p using the weight matrix w_r.

The rate of protein formation in the process of translation is controlled by the mRNA concentration r_i and the action of other molecules (α_j) involved in the translational process. The following equation describes the rate of accumulation of protein p_i:

$$\frac{dp_i}{dt} = k_{1pi} \cdot r_i \cdot f(r_i,\ \alpha_{j=1..m},\ w_p) - k_{2pi} \cdot p_i \qquad (8)$$

k_{1pi} and k_{2pi} are the maximal rates of translation, and degradation of protein $p_i \cdot f$ is again a sigmoid transfer function modifying the influence of $mRNA_i$ and translational factors α using new weight matrix w_p. Protein p_i then can, in principle, contribute to the regulation of

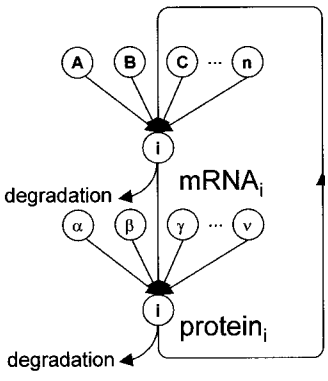


Figure 5. Two-compartment model of gene expression. Regulatory proteins A, B, C. . . n control expression of gene *i*, which is transcribed to mRNA in the first 'compartment'. Translation of $mRNA_i$ to $protein_i$ is controlled by the set of factors $\alpha,\beta,\gamma\ldots\nu$. $Protein_i$ can control expression of itself or other genes. Such a scheme can be drawn for any of the members A. . . n. The formal description of the process is shown in equations 7 and 8.

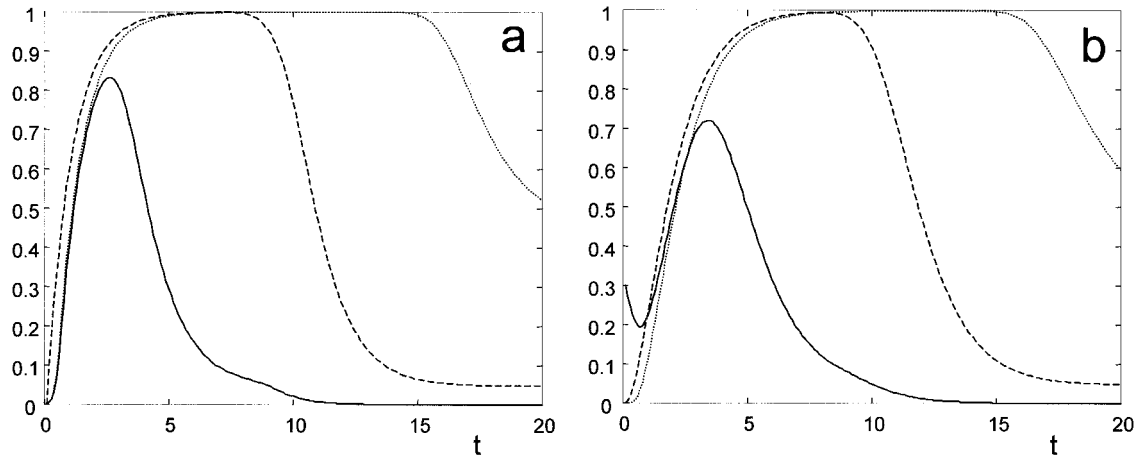


Figure 6. Two-compartment model with no translational control. *a)* mRNA levels; *b)* protein levels. When the translational control is driven by mRNA concentration only (Eq. 9), the protein levels closely follow those of mRNA. The model situation of Fig. 2 was used to compute the profiles (solid line, A; dashed, B; dotted, C).

the expression of other genes of the system (Fig. 5) forming a feedback circuit.

Equations 7 and 8 form a 'two-compartment' model of gene expression. The two networks are connected by one synapse formed by $mRNA_i$. Both networks can be evaluated from experimentally measured mRNA and protein concentrations. Equations 7 and 8 can be written for all n members of the regulatory process forming a system of 2n differential equations.

The translation of an mRNA to the corresponding protein is a several-step process. It has been considered that the components of the system are available in excess. Therefore, the rate of protein accumulation is determined only by the amount of mRNA. The weight matrix thus becomes zero and equation 8 simplifies to:

$$\frac{dp_i}{dt} = k_{1pi} \cdot r_i - k_{2pi} \cdot p_i \qquad (9)$$

The time series of the amounts of protein will then follow that of the mRNA (**Fig. 6**).

We can consider that the rate-limiting step is not only constrained by the mRNA concentration, but by one or more of the processes of the translational machinery. In this case, the elements of the weight matrix w_p gain non-zero values. In principle, the time series of mRNA and protein can differ. This has been observed previously and recently shown experimentally (13, 14). The simulation of such situation is shown in **Fig. 7**.

Reconstruction of the network architecture from experimental data

The network architecture is uniquely determined by the number of nodes and the wiring between the nodes. The wiring is defined by the weight matrix. The

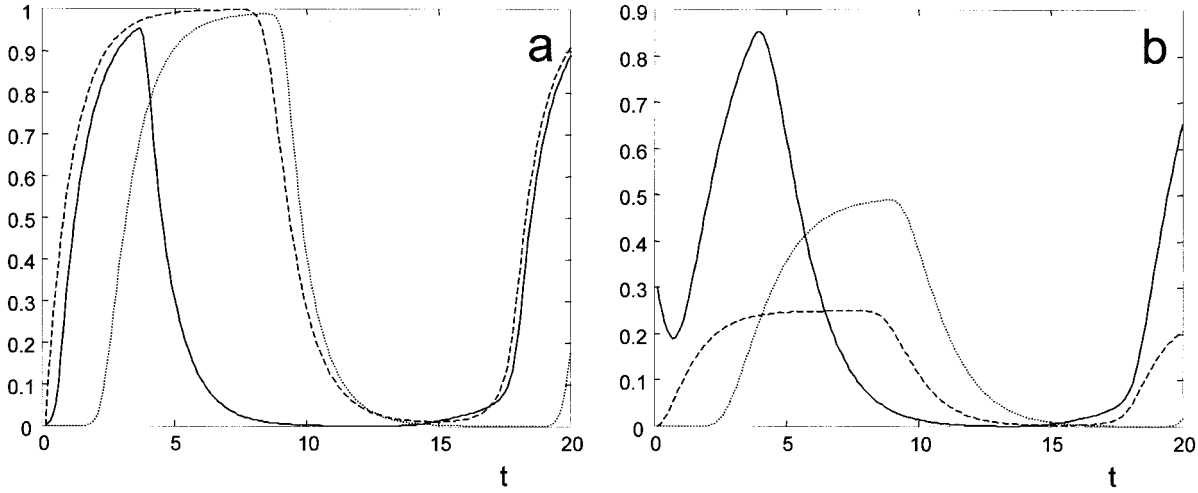


Figure 7. Two-compartment model with translational control. *a)* mRNA levels; *b)* protein levels. Synthesis of protein A (solid line) is enhanced by assigning a positive value to the weight matrix member w_{p11}. The degradation of protein B was increased by increasing the value of the decay parameter of protein B. All other parameters were kept unchanged. As a result, protein and mRNA levels differ. A periodic pattern was reached under current combination of parameters. The model situation of Fig. 2 was used to compute the profiles (solid line, A; dashed, B; dotted, C).

number of nodes is defined as the number of genes or other factors involved in the regulatory event. Ultimate reconstruction of the network therefore means computation of the weight matrix from the experimental data. The only data available are the time series of expression levels of the genes involved in a particular regulated system. We can measure or at least estimate the decay rate constant (k_2) and the maximal rate of expression (k_1) (the decay rate constant can be expressed in terms of a proteins half-life $t_{1/2}$, which is related to the k_2 constant by the formula $t_{1/2}=\ln 2/k_2$).

For every recurrent network, there exists a feed-forward network equivalent to the recurrent network over a finite time period (15). The basis of the strategy is the multiplication of the units and connections of the recurrent network over limited time. In fact, this unfolding is equivalent to the situation shown in Fig. 1. The state of the network at an arbitrary time $t+\Delta t$ is given by the state at the time t and the transition is described by a set of simultaneous differential equations such as those described by equation 4.

Provided that the network relaxes to a steady state, recurrent back-propagation (16) can be used to reconstruct the weight matrix from known initial and terminal conditions. For a general type of response, stochastic optimization like simulated annealing (17), which is capable of finding the global minimum in the parameter space representing the unique solution, has to be adopted.

Difficulties arise when the system is underdetermined (fewer experimental data points then equations). Then an infinite number of solutions of equation 4 exist. Unfortunately, this is the predominant case; the number of experiments is limited by the physiology of an organism and by technical and scale constraints. In this case, a nonlinear interpolation scheme has to be applied to the expression time series that generates a sufficient number of linearly independent equations. A danger of this approach lies in the fact that some rapidly occurring changes in the expression profiles can be hidden between two measuring points within an interval too long to capture these changes. In unequally spaced data, an interpolation yielding equal spacing gives higher weights to points with a longer interval between them. The general rule is that the sampling intervals should be short enough to capture at least the principal trends.

Another problem is introduced by linear dependencies between different kinetic profiles of different gene products. This introduces singularities in to the model and makes the solution impossible. Linearly dependent profiles are results of control by the same set of genes but to a different extent (they are coregulated). As the goal is reconstruction of the network structure, these different profiles represent identical wiring patterns and can thus be replaced by one synapse only. Therefore, preprocessing (using, for example, cluster analysis methods to group profiles similar in shape) has to be applied.

DISCUSSION

A biologically plausible model?

Four principle approaches to the modeling of cell regulatory networks can be found in the literature: *1)* Boolean, which treats gene expression as a network of 'switches' where each member of the network is a gene that is in the state of 0 or 1, i.e., transcribed or not (1, 2, 2, 18, 19, 20); 2*)* models based on kinetic equations and binding equilibria (21, 22, 23) or an artificial system of differential equations (8); *3)* McAdams introduced an approach that interprets a genetic control system as a circuit formally similar to an electronic one (24, 25, 26); *4)* Neural network-based models that replace gene interactions with a weight matrix and linear (6) or sigmoidal (7) transfer functions and a linear combination of the descriptive equations.

In principle, cell regulatory processes can be described by a set of binding equilibria and kinetic differential equations. Such an approach was used, for instance, in the modeling of a biochemical pathway for chemotaxis (21, 22). The constants of the model were obtained from measurements or estimated by fitting the model to the experimental data. A limited genetic algorithm was used to search for sets of binding reactions and associated binding constants expected to give mutant phenotypes in accord with experimental data (23). This approach considers a process in the cell as a physicochemical system analogous to the *in vitro* situation. In such cases, it is assumed that all components of the system and the interactions between them are known and all necessary constants can be measured or estimated. A model that fully describes the behavior of the system can thus be formulated.

In the case of gene expression, such a situation is difficult to achieve as the mechanisms of the processes of gene control are not completely understood. Replacing unknown mechanisms by a kinetic equation with necessarily virtual constants derived from fitting the model to experimental data means the same level of abstraction as replacement of the process by an artificial model like the one presented here, but still with the rigid form of a kinetic equation. For poorly identified systems, where one kinetic equation can comprise several processes, the description by kinetic equation is invalid and does not correctly describe the dynamics of the system. Such a situation can occur in a neural network model as well, but our model does not strictly follow the chemical reaction pathway; it is designed to be flexible enough to follow the observed dynamic behavior. For any system that is not completely understood, a dominant case in most biological systems, the weight matrix (derived from the fitting of the model to the data) can at least give a qualitative indication of the principles of regulatory interaction among the elements of the system. The physicochemical type of model needs to know this information *a priori*. Known experimental errors can provide estimates of the boundaries for the weight matrix elements, i.e.,

whether the mode of action of one component on another is significantly positive, negative or zero. This minimalist qualitative result can draw the attention of experimental biologist to a limited number of significant components of the system. New experiments that bring new information can then be made and the model can be improved. Such an iterative approach leads to an accurate model that can be tested and can finally provide valid predictions.

The Boolean network model interprets gene interactions as connections between genes and the states of gene expression that can be either on or off. The next state of the network is calculated from the known previous state from the rules for interactions between elements of the network and from the known regulatory connections between the genes. The network scheme interpreting the gene control system as a network is identical to the one presented here. The difference is in the interpretation of the regulatory function. In our case, the model can compute expression levels of mRNAs or proteins at any time, whereas the Boolean network model can suggest whether the gene is on or off during the given cycle. The Boolean model does not consider that even when the gene is off the concentration of the gene products can still be high enough to act in the regulatory process. It maintains gene activity even when the gene is already not transcribed. If this shift occurs early in the process, the error is propagated in time and finally can lead to a different state of the network than the one predicted by the Boolean network model. Simulation of behavior of the simple networks presented here showed that between the two limit cases defined by the Boolean network model, a number of different kinetic and terminal states can exist. These states are determined by the combination of parameters comprising the basic physicochemical and kinetic features of the members of the system, and not only by the initial state of expression and the weight matrix as suggested by the Boolean model. For the same weight matrix (same interactions), different kinetic behaviors and terminal states can be reached depending on the other parameters and initial conditions. It is very likely that the behavior of Boolean networks modeling a regulatory event represent a limiting case that can be reached by the system, but not inevitably. Other solutions leading to different terminal states of the system starting from the same initial conditions are also possible. Particular initial conditions can lead to attractors different from those suggested by the Boolean networks. Therefore, the neural network model is more realistic and also more accurate, providing information about instantaneous concentrations of the components of the system, not only about the state of genes transcription. Reconstruction of the Boolean network from the experimental data that are expected to be available, which means mostly DNA chip and proteomics data, would also be difficult.

Circuit simulation of genetic control has a great advantage in the apparent analogy between genetic and electronic circuits and the existence of a broad theoretical background coming from electronic engineering. The circuit-based models may lead to a robust interpretation of complex regulatory networks, but the question remains, how valid is the replacement of gene regulation processes by discrete electronic-like elements? An approach using fuzzy elements (27) seems to improve the validity of this type of model. Reconstruction of the circuit from the experimental data will probably be another challenge.

The advantage of the model presented in this paper is that it is continuous in time, does not use artificial elements, and uses a transfer function that transforms the input to a shape close to the one observed in natural processes. The model allows us not only to fit the experimental data, but can also provide information about the principles of control and mutual interactions of the elements of the modeled system even when only the elements are known in advance, and nothing about their interactions. Furthermore, the model can, from known principles of interaction of the elements of the system, compute the dynamic behavior of the system and the possible terminal states the system can reach. Predictions made by such a model outside the measured interval are likely to be more accurate than those made by totally artificial models.

The model is open for modification. It works with concentrations (amounts) and calculates the instantaneous concentration of the elements of the system by means of the set of equations described above. If there exists a known mechanism of control of the concentration of one of the elements of the system, the equations controlling these elements in the model can be replaced by the new ones. Such a hybrid model has a higher chance to be the correct description of the system than a purely artificial one.

In our model, we consider gene expression as a two-stage process of transcription and translation, as in the real cell. All the previously published models we were able to find consider the process as one stage, ignoring differences in the control of translation and transcription.

The drawback of the proposed model is that it still comprises a large number of parameters that have to be computed from experimental data. Many parameters require large number of data points to correctly compute the parameters. The number of data points obtained from experimental time series is limited by technical constraints that can hardly be overcome. Therefore, interpolation schemes and minimization procedures that bring inaccuracy to the results have to be applied.

Solution of the differential equations representing the model to obtain the weight matrix has to be accomplished numerically. The schemes mentioned above, which in principle search for a minimum in a parametric space, never guarantee that in any particular case the global minimum representing the correct solution will be reached. Therefore, for the best possible results it is necessary to introduce constraints com-

ing from experimental observations and known features of the modeled system. It implies that building up a successful model will be an iterative process involving both mathematicians and experimental biologists. This kind of collaboration is inevitable for any kind of modeling approach that has been or will be invented.

Another simplification in our approach is that the model presumes that all necessary molecules are available as in an *in vitro* solution. Therefore, terms representing the diffusion of molecules to the place where the reaction takes place or compartmentalization (28) of the cell where all necessary agents are pooled should be considered. Other environmental variables can be included to describe a particular problem. On the other hand, compared to the previously published ones, this model already comprises a nonlinear response to the regulatory action of the other genes of the system, the effect of degradation of the gene product, different rates of expression of individual genes, and their regulatory effect. How much these terms are sensitive to random perturbation or experimental error could be tested with data obtained from a known system with a known experimental error. A series of similar models can be constructed by distributing the input data within the known standard deviation for each measurement. Comparing these models would tell us how sensitive the models are with respect to small amounts of noise in the input data.

Once a sufficiently comprehensive model is established, its dynamic properties can be tested by mutageneses *in vivo* and 'in silico' and results compared. This is a general way of refining the functionality of the model. The advantage of a computer-based model is that it can predict situations that can hardly be reached experimentally and can explain underlying mechanisms that are, in principle, inaccessible. Even analysis of the very simple model shown in Fig. 2 demonstrates very complex behavior, depending on the combination of the parameters of the model. Randomly generated expression profiles for the same weight matrix range from a complete stop of expression through oscillatory behavior to saturation. It shows how narrow the range of biologically plausible parameters is and how tight control must be in the living cell.

Results show that the model presented here reproduces well the predicted behavior of these simple artificial examples. By playing around with the values of the parameters, various model situations can be tested. I ran simulated time-course experiments for large networks (up to 100 nodes) with randomly generated weight matrices and parameters. These networks were allowed to iterate until they reached a steady state or a cycle. The dynamics of the network output depended on all parameters of the model and the initial conditions. All networks were generated with randomly selected average connectivity (number of connections per node). Not surprisingly, large networks required a longer period of time to reach a terminal state.

These observations emphasize how finely tuned and tight the regulatory system of the cell must be in order to be flexible enough to adapt to changing environmental conditions and yet maintain global stability. It is encouraging that the neural network approach can reflect this complexity, indicating that it is a good base for modeling of such complex regulated systems.

Although these models include idealization of known features of gene control, they are open to the addition of other different modes of interaction and responses and are already capable of describing simple regulatory events occurring in the cell. To what degree these models are real pictures of reality remains to be proved by comparison with experimental data of the known systems.

The cell as a parallel processing network with a limited number of states

Bray (29) discussed an interpretation of the enzymatic reaction as a memory unit capable of transforming an input (concentration of substrate) to an output (type and concentration of product). The combination of such nodes can form a network that can serve as a decision block adapting its output to its input. The output can be both extremely complex and simple [for example, the decision about the direction of rotation of a bacterial flagellum (22, 28)]. McAdams and Shapiro (24) used the analogy of an electronic circuit to explain the process of decision for the lysogenic or lytic path after phage lambda infection. Both examples consider biochemical reactions as logical (or, more generally, fuzzy logical) units capable of storing and processing information. In our case, basic processes of transcription and translation are modeled by a set of equations principally identical to the artificial recurrent neural networks. Recurrent neural networks have often been used as associative memories. Networks can be trained to reach, after a certain time, some finite number of states depending on the input to the network. When the input is changed, the output state of the network relaxes to another of the possible states: the network can 'remember' the input. This approach leads to the concept of 'black box', where the only required function of the unit is transformation of an input to the output in the observed way. Neural networks are excellent tools for the construction of this architecture. In such a case the weight matrix does not necessarily have to have a direct meaning in terms of influence of certain known parameters, but can be artificial, serving only for the purpose of correct transformation of the input value.

To get the correct concentration of necessary products at a certain time, gene expression is controlled in a complex way, reacting to changes in the developmental program or environmental conditions. The cell is allowed to reach only a limited number of states that maintain its stability. Therefore, a large number of possible input conditions can lead to only a limited number of terminal states over time. The existence of a restricted number of allowed states maintains the regulatory stability of the system on one hand and, on the

other, its variability maintains the necessary flexibility. The situation resembles the attractors of an artificial neural network and suggests that the neural network could be an acceptable model of control of cell processes in general. From this point of view cell control can be considered to be a system of mutually connected networks processing in parallel. Whether this model is plausible and to what level this generalization is valid will be the subject of the future studies. FJ

The author wishes to thank Tom Bickle and Jeremy Ramsden for comments on the manuscript. This work has been supported by grants 204/00/1253 and 203/98/0422 of GACR.

REFERENCES

1. Somogyi, R., and Sniegoski, C. A. (1996) Modeling the complexity of genetic networks: understanding multigenic and pleiotropic regulation. *Complexity* **1**, 45–63
2. Wuensche, A. (1998) Genomic regulation modeled as a network with basins of attraction. *Pacific Symp. Biocomput.* **3,** 89–102
3. Somogyi, R., Fuhrman, S., Askenazi, M., and Wuensche, A. (1997) The Gene expression matrix: towards the extraction of genetic networks architectures. *Nonlinear Analysis, Proc. of Second World Cong. of Nonlinear Analysis (WCNA 96)* **30,** 1815–1824
4. Wuensche, A. (1993) The ghost in the machine: Basins of attractions in random boolean networks. In *SFI Studies in the Science of Complexity,* pp. 465–501, C. G. Langton
5. Samsonova, M. G., and Serov, V. N. (1999) NetWork: an interactive interface to the tools for analysis of genetic network structure and dynamics. *Pacific Symp. Biocomput.* **4,** 102–111
6. d'Haeseleer, P. (1999) Linear modeling of mRNA expression levels during CNS development and injury. *Pacific Symp. Biocomput.* **4,** 41–52
7. Weaver, D. C., Workman, C. T., and Stormo, G. D. (1999) Modeling regulatory networks with weight matrices. *Pacific Symp. Biocomput.* **4,** 112–123
8. Chen, T., He, H. L., and Church, G. M. (1999) Modeling gene expression with differential equations. *Pacific Symp. Biocomput.* **4,** 29–40
9. Bose, N. K., and Liang, P. (1996) *Neural Network Fundamentals with Graphs, Algorithms, and Applications,* p. 312, McGraw-Hill, New York
10. Hoel, D. (1970) A simple, two compartmental model applicable to enzyme regulation. *J. Biol. Chem.* **245,** 5811–5812
11. Yagil, G. (1975) Quantitative aspects of protein induction. *Curr. Top. Cell Regul.* **9,** 183–286
12. Hargrove, J. L., and Schmidt, F. H. (1989) The role of mRNA and protein stability in gene expression. *FASEB J.* **3,** 2360–2370
13. VanBogelen, R. A., Greis, K. D., Blumenthal, R. M., Tani, T. H., and Matthews, R. G. (1999) Mapping regulatory networks in microbial cells. *Trends Microbiol.* **7,** 320–328
14. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19,** 1720–1730
15. Bose, N. K., and Liang, P. (1996) *Neural Network Fundamentals with Graphs, Algorithms, and Applications*, p. 335, McGraw-Hill, New York
16. Pineda, F. J. (1987) Generalization of back-propagation to recurrent neural networks. *Phys. Rev. Lett.* **59,** 2229–2232
17. Bose, N. K., and Liang, P. (1996) Neural Network Fundamentals with Graphs, Algorithms, and Applications, p. 323, McGraw-Hill, New York
18. Akutsu, T., and Miyano, S. (1999) Identification of genetic networks from a small number of gene expression patterns under the boolean network model. *Pacific Symp. Biocomput.* **4,** 17–28
19. Thomas, R. (1973) Boolean formalization of genetic control circuits. *J. Theoret. Biol.* **42,** 563–585
20. Thomas, R. (1978) Logical analysis of systems comprising feedback loops. *J. Theoret. Biol.* **73,** 631–656
21. Hauri, D. C., and Ross, J. (1995) A model of excitation and adaptation in bacterial chemotaxis. *Biophys. J.* **68,** 708–722
22. Bray, D., Bourret, R. B., and Simon, M. I. (1993) Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Mol. Biol. Cell* **4,** 469–482
23. Bray, D., and Bourret, R. B. (1995) Computer analysis of the binding reactions leading to a transmembrane receptor-linked multiprotein complex involved in bacterial chemotaxis. *Mol. Biol. Cell* **6,** 1367–1380
24. McAdams, H. H., and Shapiro, L. (1995) Circuit simulation of genetic networks. *Science* **269,** 650–656
25. McAdams, H. H., and Arkin, A. (1998) Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* **27,** 199–224
26. McAdams, H. H., and Arkin, A. (2000) Towards a circuit engineering discipline. *Curr. Biol.* **10,** R318–R320
27. Arkin, A., and Ross, J. (1994) Computational functions in biochemical reaction networks. *Biophys. J.* **67,** 560–578
28. Bray, D. (1999) Computer-aided resolution of an experimental paradox in bacterial chemotaxis. In *Bacterial Neural Networks*, p. 48, Dourdan, France
29. Bray, D. (1995) Protein molecules as computational elements in living cells. *Nature (London)* **376,** 307–312

Received for publication June 26, 2000.
Revised for publication September 15, 2000.