

14 COMPUTATIONAL CELL BIOLOGY AND COMPLEXITY

Charles Coward, MD*
Banu Onaral, PhD*

Chapter Contents

14.1 Computational Biology

- 14.1.1 Computational Modeling of Cellular Processes
- 14.1.2 Modeling Control Mechanisms within the Cell
- 14.1.3 Modeling Metabolic Pathways within the Cell

14.2 The Modeling Process

- 14.2.1 Methods of Modeling
- 14.2.2 Equations of Modeling

14.3 Bionetworks

14.4 Introduction to Complexity Theory

- 14.4.1 Complexity in Metabolic and Control Networks

Exercises

Suggested Reading

At the end of this chapter, the reader will be able to:

- Describe how computational modeling is used to model cellular processes.
- Develop simple models of cellular control.
- Describe the basic concepts defining complexity theory.
- Describe how complexity theory applies to cellular biology.

*With contributions by Susan Blanchard and Anne-Marie Stomp.

14.1 COMPUTATIONAL BIOLOGY

Computational biology involves modeling, measuring, or classifying the processes within a cell, such as metabolic or control pathways, or the proteins that are active within those pathways. Specifically, computational biology includes the use of computer systems to search the genome/proteome through the use of the huge genomic and proteomic databases or computer systems and can be used to analyze cellular messenger RNA (mRNA) levels within a tissue through microarray testing. Computational biology tools can be used to model pathways to understand how pathways interact or to find drug targets to treat specific diseases. These systems can bridge the purely theoretical through modeling or utilize or generate experimental data. Ultimately, these different tools and approaches are used to generate information about cellular behavior and to integrate the information to develop an understanding about how cells function within an organism.

14.1.1 Computational Modeling of Cellular Processes

Modeling of cellular processes can be divided into two areas, modeling the internal control of the cell or modeling the metabolic pathways within the cell. Both types of processes will, in general, interact with the environment around the cell. Control of the cell can be affected by stimuli from other cells nearby, from tissues distant from the cell of interest, and from the extracellular matrix to which the cell is attached. Metabolite concentrations may be affected by glucose concentrations in the environment around the cell. Both of the areas (control and metabolic modeling) may therefore consider the exterior of the cell. Cellular modeling may also require simulating the effects of different compartments (see Chapter 12) that may restrict the interactions between different control paths or interactions or metabolic concentrations. Examples of different compartments within a eukaryotic cell are cytoplasm, mitochondria, and nucleus. Other compartments may also be considered such as Golgi apparatus or sarcoplasmic reticulum.

14.1.2 Modeling Control Mechanisms within the Cell

Control within a cell is distributed throughout the entire cell. Receptors interact with the exterior of the cell and initiate effects within the cell. Some receptors interact with DNA to stimulate or inhibit production of mRNA that leads to production of proteins. These receptors may function on the cell membrane or within the cytoplasm and exert their influence on the mRNA/protein production through cascades of protein messengers or interact with second messengers, ultimately influencing protein concentrations within the cell. Some of these proteins may stimulate the initial conversion to cancer. These types of receptors are called proto-oncogenes and may be tyrosine kinases. Alternatively, receptors exist within the nucleus for steroid hormones which can travel through cell membranes that enclose the cytoplasm and separate the cytoplasm from the interior of the nucleus. Receptors may also facilitate

absorption of substrate for metabolic pathways, such as GLUT receptors which allow a cell to absorb glucose from the bloodstream.

14.1.3 Modeling Metabolic Pathways within the Cell

Metabolic pathways within the cell provide the ability to convert energy within various molecules (such as glucose, ketoacids, and certain amino acids) into energy needed by the cell to perform various maintenance functions, cell division, or the functions required by the tissue to maintain the organism as a whole. The energy within these molecules is stored in chemical bonds between specific atoms within the molecules of interest. These metabolic pathways may involve a process called *respiration* that utilizes the characteristics of oxygen molecules to assist in the chemical conversion and liberation of energy from the fuel molecules. Energy can also be extracted from glucose without the need of oxygen, although this is significantly less efficient. Energy generated in this manner produces lactic acid, a molecule that can change the pH and adversely affect the activity of enzymes within the cell.

The process of energy extraction or energy conversion is a pathway requiring many enzymes, with the corresponding complexity due to many control points within the pathway, and will cross into several compartments within the cell. The citric acid cycle (Kreb's cycle) functions in part in the cytoplasm and in part within the mitochondria. If the glycolysis pathway and the Kreb's cycle are to be modeled, the cytoplasm and mitochondria must be modeled. Provisions must be made to allow metabolites within the Kreb's cycle to cross into the mitochondria, and adenosine triphosphate (ATP) must be permitted to exit the mitochondria.

14.2 THE MODELING PROCESS

In developing a model of cellular processes, whether control or metabolic, the purpose of the model must be established first. Is the purpose of the model to simply understand the processes, to identify the key control points within a process (for a drug discovery process), or to attempt to develop information or data for future experimentation? This question is important because it can influence how the model is developed and which assumptions are made in constructing the model. The types of assumptions may influence decisions on which computational platforms or software packages are used to execute the model. If a preexisting software package is used for the model (or simulation), then the assumptions made during its design and implementation should be evaluated to determine if they are consistent with the goals of the model.

The goals of the model may also determine the computational time required for the model to perform its simulation with simpler models requiring several minutes on a personal computer or workstation and larger models requiring hours to days of processing on supercomputers or Linux clusters.

14.2.1 Methods of Modeling

Computational models can be built using existing higher-level software packages such as MATLAB/SIMULINK[®], V-Cell (<http://www.nrcam.uchc.edu/>, Fig. 14.1), E-Cell (<http://www.e-cell.org/>), Jig Cell (<http://jigcell.biol.vt.edu/>), LabVIEW[®], and Microsoft[®] Excel, or through programming languages such as C/C++, Java, or Basic. The higher-level software packages provide an environment that allows the user to quickly create a model or simulation with reduced debugging time and with predefined graphics capabilities. Some of the packages may allow the user to simply click and drag various icons onto the workspace to simulate processes. The disadvantage of these packages is that it may not be possible to simulate desired features, and the simulations may have considerable overhead that will lead to extended processing times for complicated simulations. The advantages of using programming languages include higher processing speeds and additional flexibility to create a detailed simulation that meets the requirements. The disadvantages of using programming languages are the additional time needed to debug the programs and the difficulty in developing

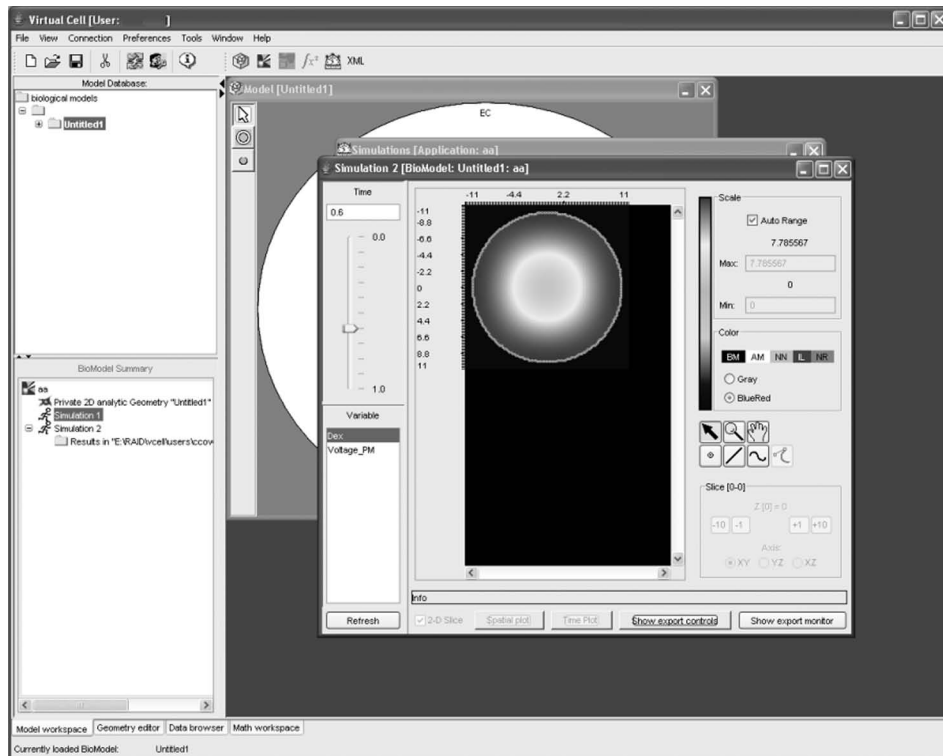


Figure 14.1 V-Cell (<http://www.nrcam.uchc.edu/>), which allows a user to simulate cellular processes. The colored circle in the center of the screen indicates diffusion into a cell at a specific time. The time can be changed with controls.

output that demonstrates the model's results. Usually, a greater understanding of a programming language is required to create a model than to create a simulation using a software package such as V-Cell or E-Cell. The higher-level software packages can be combined with the models developed in one of the programming languages to allow a simulation to have the speed and flexibility afforded by the language and the graphics capabilities of the software packages.

Computational models are built using equations to represent the various aspects that define the simulation. Usually, models or simulations (the terms are used here interchangeably) are based on modeling changes within a cell or tissue. Modeling the changes within the cell allows the modeler to represent a signal pathway (cellular control) or an ion/metabolite concentration or flux increase within the cytoplasm. Change can be spatial (concentration gradient between different compartments within the cell) or temporal (change in calcium concentration within the cytoplasm in a muscle cell during contraction). Changes are usually modeled with ordinary or partial differential equations, but they can be modeled with other types of methods or equations (perhaps, for example, a Boolean network with delays). The model could consist of the closed form solution to a solvable differential equation or a numerical solution of a more difficult, ordinary, or partial differential equation.

An example of a simple model using a closed form solution to an ordinary differential equation would be a simple exponential solution. The coefficients could be modified to compare results using different characteristics.

14.2.2 Equations of Modeling

As mentioned in Section 14.2.1, the equations of modeling principally involve change. This change will be dependent on many factors, such as enzyme concentrations, substrate concentrations, initiating factors, concentration gradients, electrical gradients, decay of complex molecules (mRNA, proteins, hormones), and spatio-temporal dimensions. First-order or second-order differential equations are used to model the change within the cell. When a simulation is based on a differential equation, a method of numerical solution must be used to integrate the equation within the software. If the solution of differential equations is numerical, a grid or coordinate system is usually defined. This coordinate system is used to define points spatially for arrays that contain the values of the concentrations throughout the cell. By using an array (one-, two-, or three- dimensional) to store the concentrations, spatial change can be modeled and concentration gradients can be established.

In solving first-order differential equations, the prototypical method of solution is the Euler approximation. In this approximation, the first-order derivative is based on the general definition of a derivative, as shown for the following differential equation:

$$\frac{df}{dx} = f(x, y) \quad (14.1)$$

Using the definitions of derivatives:

$$\frac{df(x)}{dx} = \frac{\Delta f(x, y)}{\Delta x} = (f(x_2, y) - f(x_1, y)) / \Delta x \quad (14.2)$$

the equation can be manipulated to generate the Euler method:

$$x_{n+1} = x_n + \Delta x f(x_n, y_n) \quad (14.3)$$

This approximation is relatively inaccurate and, in some equations, may lead to instability of the solution. The Euler method evaluates the function at the current (x_{n+1}) position by using the value of the function at the previous position (x_n), so for a rapidly changing function, the results can lead to the inaccuracy mentioned earlier. Alternatively, a second-order or fourth-order Runge–Kutta method or one of several other methods could be used to model the derivative, and this would lead to a more stable and accurate solution.

The second-order Runge–Kutta approximation is:

$$k_1 = hf(x_n, y_n) \quad (14.4)$$

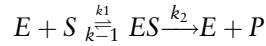
$$k_2 = hf(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1) \quad (14.5)$$

$$x_{n+1} = x_n + k_2 + O(h^3) \quad (14.6)$$

The Runge–Kutta method evaluates the function at the midpoint, so the results more closely reflect the behavior of the function itself. All of these approaches may have problems if the rate of change of the function changes over an interval (i.e., the function is slowly changing and then begins to change quickly) and these differential equations are said to be “stiff.” Stiff ordinary differential equations simulating this type of changing behavior may occur in biological models when signal transduction paths are modeled. A common approach to solving stiff ordinary differential equations is to change the integration step (decrease Δx) when the function begins to change rapidly with respect to x . Many of the higher-level simulation packages allow the user to select which method should be used to solve the first-order differential equations required in the simulation.

The Michaelis–Menten equation provides one of the basic first-order differential equations necessary for analysis of metabolic networks. The Michaelis–Menten equation relates the rate of production of a specific molecule to the product formation velocity (V_{\max}), substrate concentration [S], and substrate concentration when the product formation velocity is one-half of V_{\max} . The equation reflects a binding of the enzyme (E) and the substrate (S). During the binding the enzyme speeds the transition of the substrate to the product. The transition speed is increased by lowering the activation energy required for the conversion to occur. Without the enzyme, the transition from substrate to product would occur at a very low rate, so the enzyme, by lowering the activation energy, increases the probability and speed of conversion. With every enzymatic reaction, there is the possibility that the substrate will be “released” from the enzyme without making a conversion to the product, and this is reflected in the constant k_m . The k_m constant is equal to the ratio of the rates of breakdown of the enzyme-substrate complex to the rate of the creation of the enzyme-

substrate complex and therefore reflects the stability of the enzyme-substrate complex. This equation assumes that enzyme concentration is insignificant when compared to the substrate concentration. The equation mathematically represents the following chemical equation:



The Michaelis–Menten equation is

$$\frac{d[P]}{dt} = V_{\max}[S]/(k_m + [S]) \quad (14.7)$$

where

$$k_m = (k_{-1} + k_2)/k_1 \quad (14.8)$$

and

$$V_{\max} = k_2[E_o] \quad (14.9)$$

This velocity (V_{\max}) (maximal rate of conversion) occurs when all of the enzymes are saturated with substrate and is not a function of the rates of enzyme-substrate complex creation or dissociation. $[E_o]$ is the enzyme concentration, $[S]$ is the substrate concentration, k_m is the substrate concentration at which product formation velocity equals $\frac{1}{2}V_{\max}$, $\frac{d[P]}{dt}$ = rate of increase of the product, k_1 = rate for synthesis of E and S to ES , k_{-1} = rate for catabolism of ES to E and S , k_2 = rate for conversion of ES to E and P (production of P), k_2 is also called k_{cat} , the turnover number, and $\frac{d[P]}{dt} = -\frac{d[S]}{dt}$ is the rate of decrease of the substrate.

The Michaelis–Menten equation demonstrates both zero-order and first-order kinetics with respect to the substrate concentration. Initially, as the substrate concentration is low, the equation demonstrates first-order kinetics when the rate of conversion increases as the substrate concentration increases. Eventually, as the substrate concentration increases, the enzymes in the system become saturated and the rate of increase of the product decreases (second time derivative of $[P]$ is negative), leading to the asymptotic approach to V_{\max} . In this region the equation demonstrates zero-order kinetics, in which the velocity is no longer a linear function of the substrate concentration. This relationship exists because the second step in the chemical equation is assumed to be essentially irreversible, meaning that the enzymes in the system do not catalyze the conversion of the product into the substrate. The equation can be modified to reflect multiple reactions prior to the irreversible reaction and to incorporate the rates of those reactions.

To model spatial change, the diffusion equation is used. The equation relates spatial changes in concentrations to temporal changes and can be modified to include one-, two-, or three-dimensional change. The following equation is the one-dimensional diffusion equation in Cartesian coordinates:

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} \quad (14.10)$$

The following equation is the three-dimensional diffusion equation:

$$\frac{\partial u}{\partial t} = D \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) \quad (14.11)$$

This equation can be used to model concentration gradients that exist across cell membranes or to simulate the diffusion that exists due to interstitial pressure within tissues.

Signal Transduction Example

A common control mechanism in both control and metabolic processes in a cell is phosphorylation (Fig. 14.2). In the mechanism, an enzyme receives a phosphate (PO_4^{3-}) group from a member of a class of proteins called kinases. When the enzyme is phosphorylated, the behavior of the enzyme is changed. In many cases, the enzyme is activated. A differential equation can be created that models the process of phosphorylation. The equation will reflect the rates of change from an unphosphorylated form to a phosphorylated form and back. Each direction will have a different rate constant and will be dependent upon the stimulus.

The quantity of R_p , which may be either concentration or amount depending on the model, will provide the output used to interface to another process.

The general form of the equation is:

$$\frac{dR_p}{dt} = k_1 S(R - R_p) - k_2 R_p \quad (14.12)$$

k_1 = rate of phosphorylation of enzyme R by a kinase

k_2 = rate of loss of the phosphorylation of enzyme R by a phosphatase

S = stimulus or signal

R = amount or concentration of enzyme R in the unphosphorylated form

R_p = amount or concentration of enzyme R in the phosphorylated form

The solution to this equation is a hyperbolic (Fig. 14.3). Equation 14.12 reflects the rates of the phosphorylation and loss of phosphorylation of the enzymes and the quantities of each form of the enzyme.

A more accurate representation of the reaction uses the following equation:

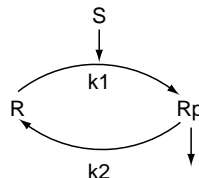


Figure 14.2 A diagram of the phosphorylation of enzyme R_p caused by stimulus S . The phosphorylation is completed by a kinase and the phosphorus is then transferred from R_p to another enzyme.

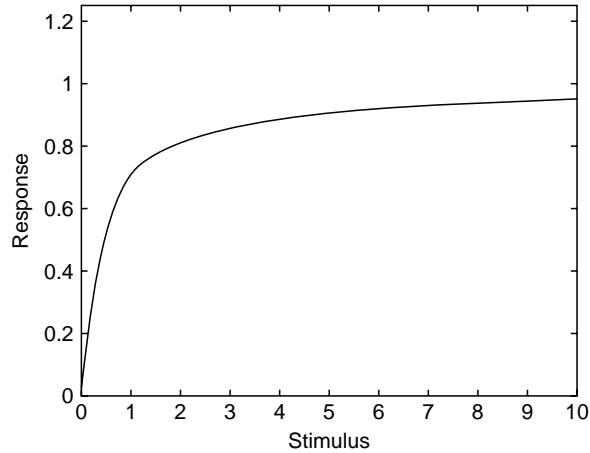


Figure 14.3 Hyperbolic curve defined by Equation 14.12.

$$\frac{dR_p}{dt} = [V_1 S(R - R_p)]/[k_1 + R - R_p] - [V_2 R_p]/[k_2 + R_p] \quad (14.13)$$

k_1 = Michaelis constant of phosphorylated enzyme R_1

k_2 = Michaelis constant of phosphorylated enzyme R_2

V_1 = max rate of phosphorylation

V_2 = max rate of loss of phosphorylation

S = stimulus or signal

R = amount or concentration of enzyme R in the unphosphorylated form

R_p = amount or concentration of enzyme R in the phosphorylated form

This approach is more realistic because these enzymes follow Michaelis–Menten kinetics (Fig. 14.4).

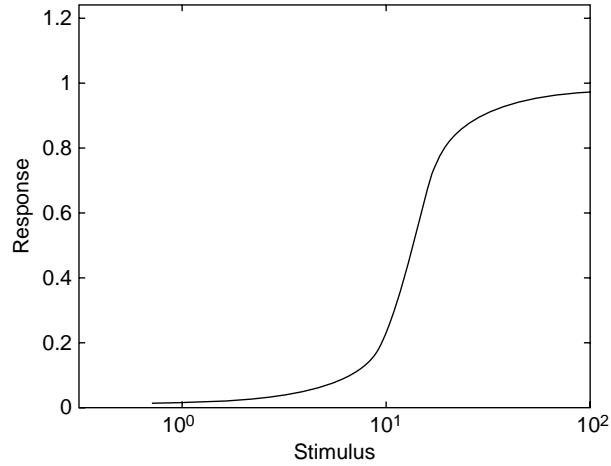
If phosphorylation of one enzyme leads to phosphorylation of another enzyme, then the signal is transmitted through a cascade. In this example (Figs. 14.5 and 14.6), enzyme R is a kinase.

$$\frac{dR_p}{dt} = [V_1 S(R - R_p)]/[k_1 + R - R_p] - [V_2 R_p]/[k_2 + R_p] \quad (14.14)$$

$$\frac{dR_1 p}{dt} = [V_3 R_p(R_1 - R_1 p)]/[k_3 + R_1 - R_1 p] - [V_4 R_1 p]/[k_4 + R_1 p] \quad (14.15)$$

In the cascade shown in Figure 14.7, three phosphorylations lead to control of a system. The characteristics of the curves are determined by the rate constants (Fig. 14.8). A third phosphorylation equation is added to model the last step in the cascade:

$$\frac{dR_2 p}{dt} = [V_5 R_p(R_2 - R_2 p)]/[k_5 + R_2 - R_2 p] - [V_6 R_2 p]/[k_6 + R_2 p] \quad (14.16)$$



```

function sigmoid
options = odeset('RelTol',1e-4, 'AbsTol',[1e-4 1e-4 1e-5]);
Y = zeros(3,1); % a column vector
[t,Y] = ode23@s_model,[0 100],[0 0 0],options);
figure, semilogx(t,Y(:,1),'k-'); % Plots fig. 14.4
axis([0 100 0 1.25]);
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k- -'); % Plots fig. 14.6
axis([0 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k- -',t,Y(:,3),'k.') % Plots fig. 14.8
axis([1 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')

function dy = s_model(t,y)
dy = zeros(3,1);
dy(1) = (0.05*t*(1 - y(1)))/(0.2 + (1 - y(1))) - (0.50*y(1))/(0.10 + y(1));
dy(2) = (0.87*y(1)*(1 - y(2)))/(0.10 + (1 - y(2))) - (0.10*y(2))/(0.1 + y(2));
dy(3) = (0.85*y(2)*(1 - y(3)))/(0.08 + (1 - y(3))) - (0.08*y(3))/(0.1 + y(3));

```

Figure 14.4 Sigmoidal curve defined by Equation 14.13.

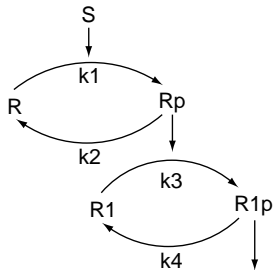
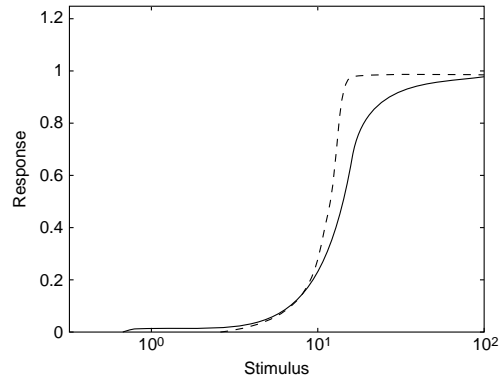


Figure 14.5 A diagram of a cascade of phosphorylation of a series of enzymes.



```
function sigmoid
options = odeset('RelTol',le-4, 'AbsTol',[1e-4 le-4 le-5]);
Y = zeros(3,1); % a column vector
[t,Y] = ode23@s_model,[0 100],[0 0 0],options);
figure, semilogx(t,Y(:,1),'k-'); % Plots fig. 14.4
axis([0 100 0 1.25]);
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k-'); % Plots fig. 14.6
axis([0 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1), 'k-',t,Y(:,2), 'k-',t,Y(:,3),'k.') % Plots fig. 14.8
axis([1 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')
```

```
function dy = s_model(t,y)
dy = zeros(3,1);
dy(1) = (0.05*t*(1 - y(1)))/(0.2 + (1 - y(1))) - (0.50*y(1))/(0.10 + y(1));
dy(2) = (0.87*y(1)*(1 - y(2)))/(0.10 + (1 - y(2))) - (0.10*y(2))/(0.1 + y(2));
dy(3) = (0.85*y(2)*(1 - y(3)))/(0.08 + (1 - y(3))) - (0.08*y(3))/(0.1 + y(3));
```

Figure 14.6 Plot of Equations 14.14 and 14.15. In these sigmoidal curves, the stimulus leads to the first phosphorylation (solid line), which in turn causes the second phosphorylation (dashed line).

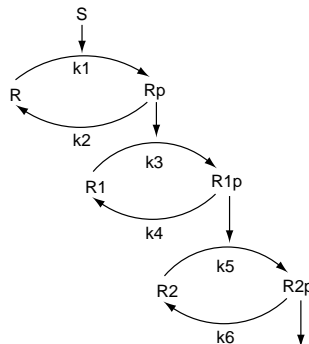
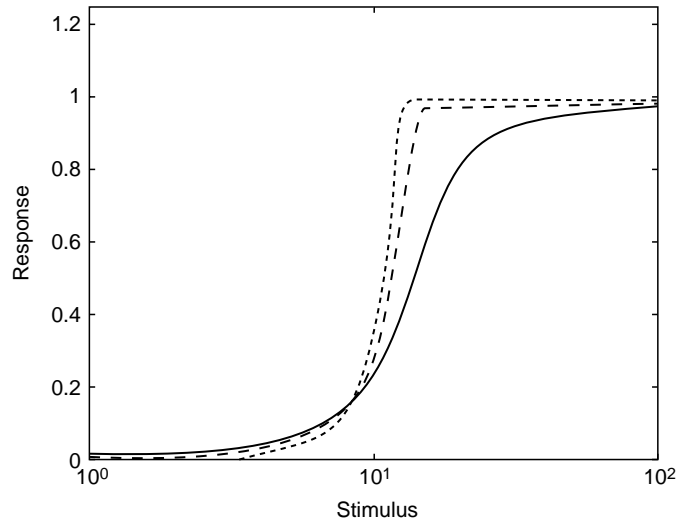


Figure 14.7 A diagram of a cascade of phosphorylation of a series of enzymes.



```

function sigmoid
options = odeset('RelTol',le-4, 'AbsTol',[1e-4 le-4 le-5]);
Y = zeros(3,1); % a column vector
[t,Y] = ode23@s_model,[0 100],[0 0 0],options);
figure, semilogx(t,Y(:,1),'k-'); % Plots fig. 14.4
axis([0 100 0 1.25]);
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k-'); % Plots fig. 14.6
axis([0 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k-',t,Y(:,3),'k.') % Plots fig. 14.8
axis([1 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')

function dy = s_model(t,y)
dy = zeros(3,1);
dy(1) = (0.05*t*(1 - y(1)))/(0.2 + (1 - y(1))) - (0.50*y(1))/(0.10 + y(1));
dy(2) = (0.87*y(1)*(1 - y(2)))/(0.10 + (1 - y(2))) - (0.10*y(2))/(0.1 + y(2));
dy(3) = (0.85*y(2)*(1 - y(3)))/(0.08 + (1 - y(3))) - (0.08*y(3))/(0.1 + y(3));

```

Figure 14.8 Plots of Eqs. 14.14, 14.15, and 14.16. In these sigmoidal curves, the stimulus leads to the first phosphorylation (solid line), which in turn causes the second phosphorylation (dashed line). The second phosphorylation then effects the third phosphorylation (dotted line). The constants in the phosphorylation equations do not reflect the behavior of the actual MAPK cascade but are for illustration only.

This model is a simplified approximation of the mitogen-activated protein kinase (MAPK) pathway (Fig. 14.9), which is involved in many aspects of cellular regulation, including growth, differentiation, inflammation, and apoptosis. The MAPK pathway

is a conserved cascade made up of kinases which receive stimuli from a variety of sources, such as integrins, G protein-coupled receptors (GPCRs), and tyrosine kinase receptors (RTKs). From integrins, the MAPK cascade receives signals from the extracellular matrix about the stress or strain imparted on the cell through the basement membrane. G protein-coupled receptors will, generally, provide signals from extracellular nonsteroid hormones about the status of the tissues farther from the cell to the MAPK cascade and, finally, tyrosine kinase receptors will provide signals from extracellular growth factors (e.g., Epidermal Growth Factor (EGF)) for the cascade. The MAPK cascade integrates these signals (along with others) to determine if gene expression will occur. Through this process, cellular regulation is initiated and controlled. The MAPK cascade family includes several pathways within a cell, such as ERK1/2, ERK5, p38, and JNK. In this pathway, a signal stimulates the initial phosphorylation which leads to a cascade. Through the cascade, the response to the stimulus develops bistable (switch-like) behavior (on/off). This simplified model is based on papers by Bhalla and Iyengar (1999), Huang and Ferrell (1996), Kholodenko (2000), and Shvartsman et al. (2001).

Protein Production Example

In this example, the phosphorylation equation is used to simulate the production of mRNA (transcription), which in turn leads to translation. In this model, the first

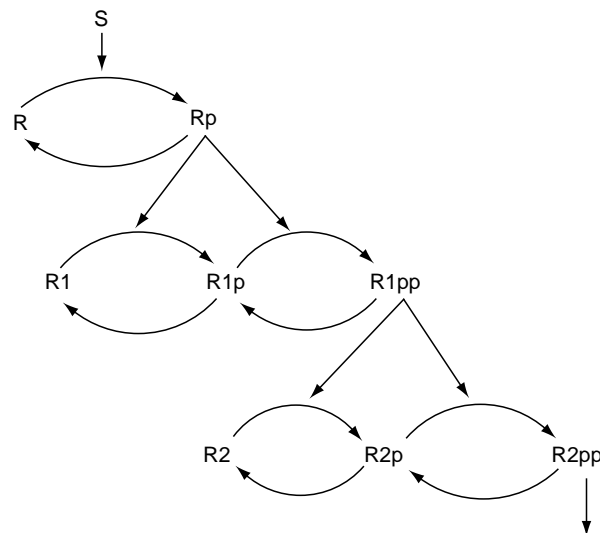


Figure 14.9 The MAPK pathway. This signal pathway is conserved in many species. It maintains tight control over specific cellular processes. The signal transduction cascade involves multiple phosphorylations to transmit the signal through the enzymes. The first phosphorylation is made to an MAPKKK (R) in response to the stimulus. MAPKKK then phosphorylates MAPKK (1) twice. In turn, MAPKK phosphorylates MAPK (R_2).

equation models a promoter that enables transcription to proceed. The promoter is phosphorylated and transcription begins. After transcription begins, mRNA is produced and begins to decay. Existing mRNA enables translation, and protein levels increase and then begin to decay. The rate of decay of mRNA is faster than the rate of decay of protein (Fig. 14.10).

$$\frac{dR_p}{dt} = [V_1 R_p (R - R_p)] / [k_1 + R - R_p] - [V_2 R_p] / [k_2 + R_p] \quad (14.17)$$

$$\frac{d[mRNA]}{dx} = k_3^* [R_p] - k_4^* [mRNA] \quad (14.18)$$

$$\frac{d[protein]}{dx} = k_5^* [mRNA] - k_6^* [protein] \quad (14.19)$$

k_3 = rate of production of mRNA

k_4 = rate of decay of mRNA

k_5 = rate of production of protein

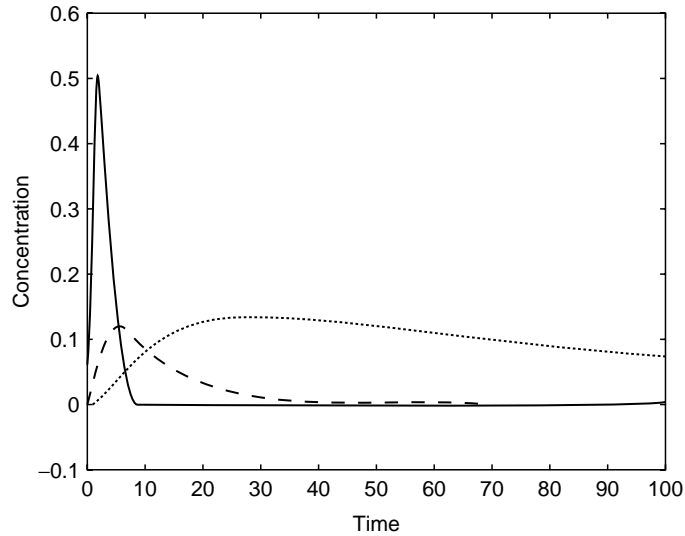
k_6 = rate of decay of protein

The model reflects the initial production of protein and represents a model for the production of a protein that is only needed periodically for cellular processes. To model proteins that are needed at some nominal concentration to maintain the viability of the cell, the stimulus must be received at a periodic rate which is related to the decay rate. Figure 14.11 reflects the model with periodic stimulus to produce necessary proteins.

14.3 BIONETWORKS

The previous examples illustrated a small series of reactions that represented several steps within a metabolic pathway. A bionetwork is a pathway within the cell that performs a specific purpose. This pathway involves many steps that require different proteins and may be connected to other pathways at many points. At any of these contact points, the pathway of interest may receive molecules (source) or export molecules (sink). Examples of the various bionetworks are glycolysis, the citric acid cycle, and β -oxidation. The challenge of this approach is that all of the various bionetwork examples interact either through substrates, products, and/or the various molecules that establish the state of the cell (ATP, NADH, NAD, etc.). This interaction adds high levels of complexity to this type of analysis.

In the network shown in Figure 14.12, each step can be defined by a single chemical equation that relates the substrates necessary for the creation of the product. Each product then becomes a substrate for the next chemical reaction in the pathway. Thus a series of linear equations can be written to define the stoichiometry of the system. This series of linear equations can be combined to create a stoichiometric matrix, and



```

function simulation
options = odeset('RelTol',1e-4, 'AbsTol',[1e-4 1e-4 1e-5]);
Y = zeros(3,1); % a column vector
[t,Y] = ode23(@simmodel,[0 100],[0 0 0],options);
figure, plot(t, Y(:,1),'k-',t,Y(:,2), 'k--',t,Y(:,3),'k.');
```

xlabel('time')

ylabel('Concentration')

```

function dy = simmodel(t,y)
dy = zeros(3,1);% a column vector

k11 = 0.15; % Scaling purposes only
k1 = 0.2;
k2 = 0.1;
k3 = 0.1;
k4 = 0.1;
k5 = 0.1;
k6 = 0.01;
v1 = 0.6;
v2 = 0.15;

if t < 2 % Used to model a limited stimulus
s = 5; % Length of stimuli
else (t > 2)
s = 0;
end
dy(1) = (v1*k11*s*(1 - y(1)))/(k1 + (1 - y(1))) - (v2*y(1))/(k2 + y(1));
dy(2) = k3*y(1) - k4*y(2);
dy(3) = k5*y(2) - k6*y(3);
```

Figure 14.10 A simple model of promoter leading to transcription and then to translation. The solid line is the level of the phosphorylated promoter, the dashed line is the level of mRNA, and the dotted line is the level of the protein produced.

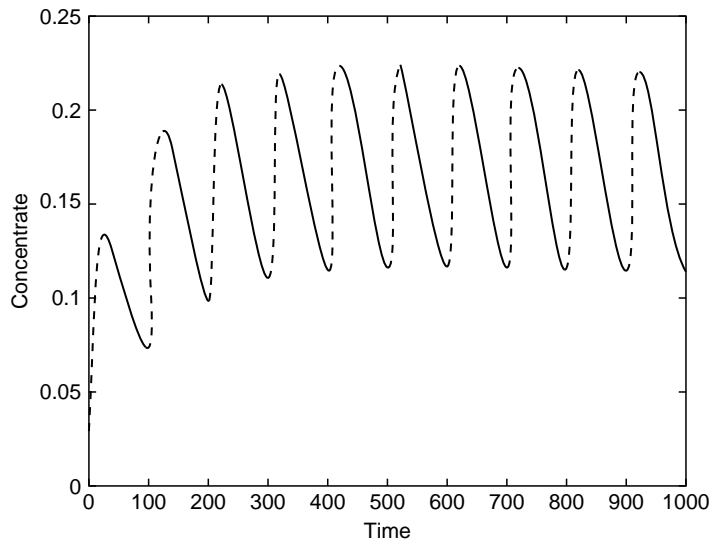


Figure 14.11 Protein levels. With simple modifications to the MATLAB code, leading to periodic stimuli (every 100 time units), protein production can compensate for the protein decay. In this manner, protein levels can remain above a required threshold.

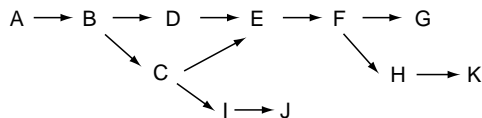


Fig. 14.12 Schematic of a simple network.

the matrix can be evaluated to derive various qualities of the system, including gene networks that are part of the mechanism that exerts control on the cellular machinery. By modeling multiple pathways that interact, the influence of inactivating particular proteins or reducing kinetic activity can be modeled. When the bionetwork reflects the changes that occur due to disease, the outcome of the inactivation of pathways or portions of pathways can reflect drugs that are competitive inhibitors of specific proteins. A library of many metabolic and regulatory pathways can be found in the Kyoto Encyclopedia of Genes and Genomes (<http://www.genome.jp/kegg/>). Bionetworks, both metabolic and control, demonstrate complex behavior due to the redundancy of pathways, the interaction of various pathways, and distribution of the control of cellular behavior over the various pathways.

14.4 INTRODUCTION TO COMPLEXITY THEORY

Complexity theory is the study of how systems interact and evolve. Complexity theory is applied to the study of biological systems to understand or characterize how various biological mechanisms function to maintain life. The study of complexity is based on observing the behavior of a system due to both the system response to specific stimuli and as a result of the interactions and functions of the various component subsystems. Through the study of complex systems, common characteristics have been identified that both characterize the systems and are required to maintain system viability.

Typically, complex systems have the following qualities:

- There can be rich interactions between a large numbers of local subsystems.
- Interactions between subsystems can be inhibitory, stimulatory, competitive, or cooperative and may be dependent upon conditions or other subsystems.
- Due to the resulting subsystem interactions, a large variety of possible behaviors (many degrees of freedom) exist.
- The possible behaviors will change over space and time (temporal and spatial degrees of freedom) (Fig. 14.13).
- Global behavior of the system can be characterized more by the interactions between subsystems and less by the behavior of the subsystems themselves.
- The global observed behavior is nonlinear, both due to the interactions of subsystems and due to the qualities of the subsystems themselves.
- Subsystems are not controlled by a central process but have control distributed over the interactions that exist among the subsystems (self-organization).
- The emergent global behavior is determined by self-organization.

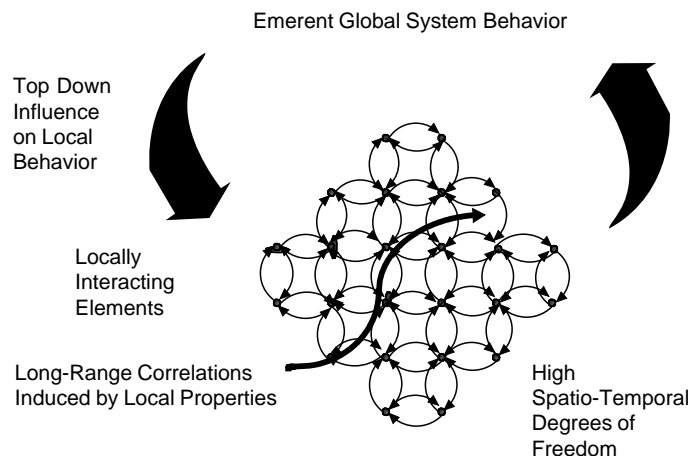


Figure 14.13 Schematic of the interactions between varying scales and emergence of a systemic behavior.

- A complex system is adaptive and stable with relatively small changes in the environment.
- A complex system demonstrates critical phenomena; it can undergo an abrupt change leading to a different state of the system.

Complex systems are systems that are composed of richly interacting subsystems (Fig. 14.14). In essence, they are systems that are composed of subsystems that interact in many ways and affect each other and may effect change in the outcome of other subsystems. These systems have many degrees of freedom and are spatially distributed, which leads to system effects that have both (or either) temporal and spatial degrees of freedom. This suggests that the events “caused” by stimuli will change over time and space. So the observed global behavior is determined more by the interaction of the subsystems than by the behavior of the individual subsystems; it “emerges” from the interactions. The observed global behavior is, in effect, established by the spatially and temporally distributed interactions, leading to distributed “self-organization.” Systemic stability and adaptive behavior are related to the distributed self-organization. The interactions and the behavior of the individual subsystems are nonlinear.

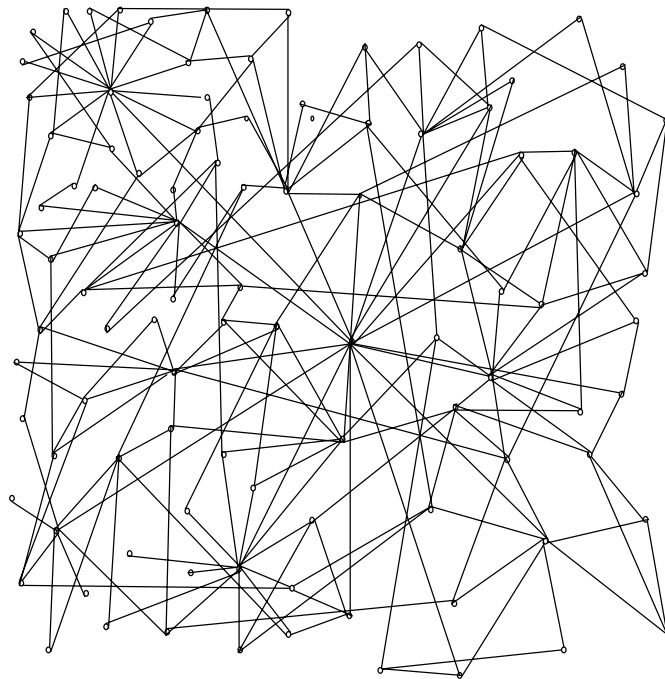


Figure 14.14 This is an example of an interaction map. Each node could indicate a protein, transcription site, etc. The chart indicates that each node interacts with other nodes and that any one node does not interact with all other nodes. The interactions that are distributed over the nodes determine which processes are active and define the system behavior.

14.4.1 Complexity in Metabolic and Control Networks

Although knowledge about biological functionality can be gained from modeling individual metabolic or control pathways by using linear differential equations, this reductionist approach will not demonstrate the full complexity of a biological organism. Both individual metabolic and control pathways interact with other pathways within a cell. These interactions between multiple metabolic pathways may lead to substrate being directed down other paths and used for other purposes. Interactions between control pathways (and metabolic paths in certain cases) communicate the status of the cell to other control processes/pathways. This communication can provide information about the environment surrounding the cell (e.g., ion concentrations and paracrine or autocrine interactions), information about distant conditions of the organism or tissue (e.g., endocrine), or status information about the cell itself. Because the status of the environment (outside and inside of the cell) is communicated to other pathways, this mechanism provides a methodology for distributed “decision making” in terms of which other pathways should be active or passive. The cellular behavior dictated by the active and passive pathways will direct the cell to perform a specific role within the organism, its phenotype, and this system behavior can be defined to be adaptive, another important quality of complexity. Examples of the cellular activities include: secretion of hormones, ions, or proteins; contraction; growth; mitosis; apoptosis; or senescence. This distributed interaction provides a mechanism for the emergence of behavior, one of the hallmarks of complexity theory. *Emergence of behavior* is a quality that can be defined colloquially as “the whole is greater than the sum of the parts” and specifically as “the interaction of multiple pathways or systems that produces system behavior that cannot be predicted from the study of the pathways/systems individually.”

An example of the communication between pathways is the generation of ATP within the cell. Because it is the currency of energy within the cell, its cellular concentration will also indicate whether the cell is in an energetic state or not. ATP, at higher cellular concentrations, which is indicative of an energetic cellular status, will have a greater probability of being used to provide the phosphate (PO_4^{3-}) group that activates existing cellular proteins, either activating or inactivating specific metabolic or control pathways. Similar examples can be found with other intracellular messengers such as GTP and cyclic AMP.

With this complex interaction between pathways leading to distributed “decision making,” the cell also has to demonstrate stability. Stability occurs if a system exposed to small stimuli or perturbations in the environment results in small changes in the system’s state (e.g., with small changes in extracellular sodium or potassium concentration, the cell remains viable). Large changes in the environment may or may not lead to catastrophic changes in the system (e.g., the cell may not be viable). In contrast, an unstable system will exhibit catastrophic and unpredictable results when exposed to a small stimulus or perturbation.

Specifically, an organism has to demonstrate stability in metabolic, control, and reproductive functions. Though stability is important to maintain a particular organism, evolution requires change. So, whereas a cell must be stable in terms of its

metabolism and distributed control, it may be advantageous for nonzero error rates to occur in copying of DNA during the S-phase (synthesis) of the cell cycle. In this case, mutations in the DNA or RNA allow changes in functional components (e.g., proteins) which may increase the fitness of the cell. The specific organism then competes within the environment and is subsequently selected by the environment and reproduces according to its acquired fitness. This type of change probably led to sickle cell anemia.

During modeling, the processes are generally assumed to be linear systems. This assumption is made to simplify the solution and will yield a solution and an approximation of the actual process. Most systems are actually nonlinear, and this is another quality of complexity.

Although this discussion has been directed to the ways that cellular behavior demonstrates complexity, in fact, complexity exists on all levels of life from subcellular to population and ecological systems. Cellular systems demonstrate complex behavior through the interactions of various metabolic and control pathways that lead to an ability to respond to internal and external stimuli and maintain viability (i.e., they are adaptive and stable). Cellular systems demonstrate a behavior that is different from what would be predicted if systems were studied individually; they demonstrate emergence of behavior.

Through the use of computer systems to identify the purpose of particular proteins within the cell, analyze and cluster mRNA expression levels, and model metabolic or control networks, research is continuing to develop new understanding of cellular functionality and to identify drug targets to treat disease. These three methodologies can be combined to provide access to existing experimental information, develop a theoretical understanding through analysis of various pathways, and then analyze experiments to validate those hypotheses. Computer systems are able to emulate some of the cellular complexity through various metabolic and control simulations, and these models will continue to demonstrate greater complexity as computer hardware becomes more powerful, computer simulation designs and architectures develop, and molecular biology provides greater understanding of the interaction between cellular functions and additional quantitative cellular data.

EXERCISES

1. The process described by the Michaelis–Menten equation can be represented by a series of first-order differential equations. These differential equations define the rate of change of each substance to be equal to the rate constant multiplied by the concentration of each molecule in the chemical equation. Develop the four equations and describe the meaning of each term.
2. Develop the four first-order ordinary differential equations that model the Michaelis–Menten equation when the transition step from ES to $E + P$ is reversible (not assumed to be irreversible).

3. The characteristics of a sigmoidal curve are dependent upon the constants and the stimulus. Change the constants and observe the effects. What range of values leads to a curve that appears to be hyperbolic?
4. The MAPK cascade is an important cellular signal transduction pathway. Which diseases are implicated in changes to the MAPK cascade?
5. In the MAPK cascade example, feedback was not modeled. What cellular mechanisms would allow negative or positive feedback to control the cascade?
6. The NF- κ B signal pathway is active in inflammation. Several models of the pathway have been developed (Hoffmann et al., 2002; Lipniacki et al., 2004). What are the important proteins that are active in the pathway? How is the pathway controlled? Is there feedback in the pathway? What is the result of the active NF- κ B pathway? Which diseases are attributed to the NF- κ B pathway?
7. Cell division models have demonstrated cellular protein oscillations. Which proteins are included in the models? What role does phosphorylation play in the models? How do the proteins regulate the cell cycle?
8. The combination of stimuli frequency, production, and decay rates determine protein levels within a cell. Modify the protein production code to increase the time of the model by making the following changes to functions simulation and simmodel.

Simulation:

```
[t,Y] = ode23(@simmodel,[0 1000],[0 0 0],options);
```

Simmodel:

```
if mod(t,100) < 2 % Used to model a limited stimulus
s = 5;
else( mod(t,100) > 2)
s = 0;
end
```

Increase and decrease the stimuli frequency, production, and decay rates and compare the result with Fig. 14.11.

9. Relating to the phosphorylation cascade example, what are the Goldbeter–Koshland function and ultrasensitivity? How would they affect the example?
10. In the phosphorylation cascade example, the cell was assumed to be in a high energy state. How is the energy state exhibited in a cell? If the cell was in a low energy state how would the curves change?
11. The phosphorylation cascade example can also be described by a series of first-order ordinary differential equations, but some of these equations reflect the input stimulus. Develop the ordinary differential equations for each of the molecules in the cascade. (Each phosphorylated molecule is different from a nonphosphorylated molecule.)
12. Emergent behavior is present in normal cellular and pathologic processes. How does emergent behavior lead to disease?

13. One of the metabolic enzymes in a cell is phosphofructokinase (PFK-1). In which pathways is it active? (See the Kyoto Encyclopedia of Genes and Genomes.) Phosphorylation controls PFK-1. How does phosphorylated PFK-1 activate and deactivate metabolic pathways? How does this relate to the energy state of the cell? What other molecules regulate PFK-1 activity and how?
14. Find example pathways for the following pathway components (Tyson et al., 2003):
 - a. Feed-forward loops
 - b. Perfect adaptation
 - c. Positive feedback
 - d. Negative feedback
 - e. Substrate-depletion oscillators
 - f. Activator-inhibitor oscillators
15. Sick cell anemia is thought to increase the fitness (resistance to specific environmental conditions) of individuals with the mutation. What is the environmental pressure that performs the selection? What other disease also increases the fitness of individuals with the mutations against the same environmental conditions?

SUGGESTED READINGS

- Allman, E.S. and Rhodes, J.A. (2004). *Mathematical Models in Biology*. Cambridge, Cambridge, U.K.
- Bhalla, U.S. and Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387.
- Fall, C.P., Marland, E.S., Wagner, J.M. and Tyson, J.J. (Eds.). (2002). *Computational Cell Biology*. Springer-Verlag, New York.
- Hoffmann, A., Levchenko, A., Scott, M.L. and Baltimore, D. (2002). The I κ B-NF- κ B signaling module: Temporal control and selective gene activation. *Science* **298**, 1241–1245.
- Huang, C.Y. and Ferrell, J.E. (1996). Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Nat. Acad. Sci. USA* **93**, 10078–10083.
- Keener, J. and Sneyd, J. (1998). *Mathematical Physiology*. Springer-Verlag, New York.
- Kholodenko, B.N. (2000). Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.* **267**, 1583–1588.
- Nelson, D.L. and Cox, M.M. (2005). *Lehninger Principles of Biochemistry*. W.H. Freeman, New York.
- Lipniacki, T., Paszek, P., Brasier, A.R., Luxon, B. and Kimmel, M. (2004). Mathematical model of NF- κ B module. *J. Theoret. Biol.* **228**, 195–215.
- Lodish, H. et al. (1999). *Molecular Cell Biology*, W.H. Freeman & Co., New York.
- Markevich, N.I., Hoek, J.B. and Kholodenko, B.N. (2004). Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J Cell Biol.* **164**, 353–359 (plus supplemental materials).

- Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (2002). *Numerical Recipes in C++: The Art of Scientific Computing*, 2nd Ed. Cambridge Univ. Press, Cambridge, U.K.
- Sauro, H.M. and Kholodenko, B.N. (2004). Quantitative analysis of signaling networks. *Prog. Biophys. & Molecular Biol.* **86**, 5–43.
- Shvartsman, S.Y. et al. (2001). Autocrine loops with positive feedback enable context-dependent cell signaling. *Am. J. Physiol. Cell Physiol.* **282**, C545–C559.
- Stephanopoulos, G.N., Aristidou, A.A. and Nielsen, J. (1998). *Metabolic Engineering Principles and Methodologies*, Academic, San Diego, CA.
- Tyson, J.J. (1991). Modeling the cell division cycle: cdc2 and cyclin interactions. *Proc. Nat. Acad. Sci. USA* **88**, 7328–7332.
- Tyson, J.J., Chen, K.C. and Novak, B. (2003). Sniffers, buzzers, toggles, and blinkers: Dynamics of regulatory and signaling pathways in the cell. *Curr. Op. Cell Biol.* **15**, 221–231.
- Weng, G., Bhalla, U.S. and Iyengar, R. (1999). Complexity in biological signaling systems. *Science* **284**, 92–96.
- Wolf, D.M. and Arkin, A.P. (2003). Motifs, modules and games in bacteria. *Curr. Op. Microbiol.* **6**, 125–134.