CHAPTER 1

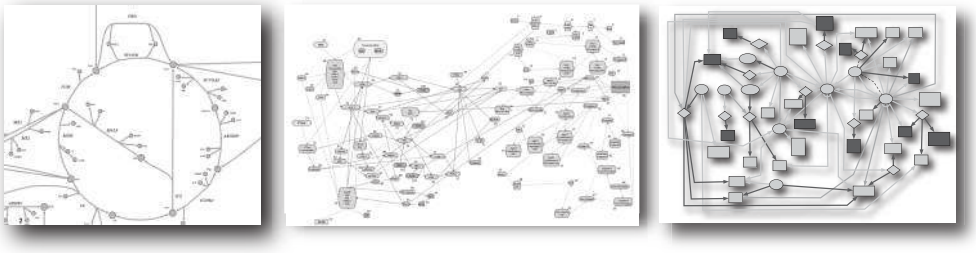
Introduction

Systems biology has been brought to the forefront of life-science-based research and development. The need for systems analysis is made apparent by the inability of focused studies to explain whole network, cell, or organism behavior, and the availability of component data is what is fueling and enabling the effort. This massive amount of experimental information is a reflection of the complex molecular networks that underlie cellular functions. Reconstructed networks represent a common denominator in systems biology. They are used for data interpretation, comparing organism capabilities, and as the basis for computing their functional states. The companion book [89] details the topological features and assessment of functional states of biochemical reaction networks and how these features are represented by the stoichiometric matrix. In this book, we turn our attention to the kinetic properties of the reactions that make up a network. We will focus on the formulation of dynamic simulators and how they are used to generate and study the dynamic states of biological networks.

1.1 Biological networks

Cells are made up of many chemical constituents that interact to form networks. Networks are fundamentally comprised of *nodes* (the compounds) and the *links* (chemical transformations) between them. The networks take on functional states that we wish to compute, and it is these physiological states that we observe. This text is focused on dynamic states of networks.

There are many different kinds of biological network of interest, and they can be defined in different ways. One common way of defining networks is based on a preconceived notion of what they do. Examples include metabolic, signaling, and regulatory networks; see Figure 1.1. This



(a) Metabolic (b) Signaling (c) Regulatory

Figure 1.1 Three examples of networks that are defined by major function. (a) Metabolism. (b) Signaling. From Arisi *et al. BMC Neuroscience* 2006 7(Suppl 1):S6 DOI: 10.1186/1471-2202-7-S1-S6. (c) Transcriptional regulatory networks. Image courtesy of Christopher Workman, Center for Biological Sequence Analysis, Technical University of Denmark.

approach is driven by a large body of literature that has grown around a particular cellular function.

Metabolic networks Metabolism is ubiquitous in living cells and is involved in essentially all cellular functions. It has a long history – glycolysis was the first pathway elucidated in the 1930s – and is thus well known in biochemical terms. Many of the enzymes and the corresponding genes have been discovered and characterized. Consequently, the development of dynamic models for metabolism is the most advanced at the present time.

A few large-scale kinetic models of metabolic pathways and networks now exist. Genome-scale reconstructions of metabolic networks in many organisms are now available. With the current developments in metabolomics and fluxomics, there is a growing number of large-scale data sets becoming available. However, there are no genome-scale dynamic models yet available for metabolism.

Signaling networks Living cells have a large number of sensing mechanisms to measure and evaluate their environment. Bacteria have a surprising number of two-component sensing systems that inform the organism about its nutritional, physical, and biological environment. Human cells in tissues have a large number of receptor systems in their membranes to which specific ligands bind, such as growth factors or chemokines. Such signaling influences the cellular fate processes: differentiation, replication, apoptosis, and migration.

The functions of many of the signaling pathways that is initiated by a sensing event are presently known, and this knowledge is becoming more detailed. Only a handful of signaling networks are well known,

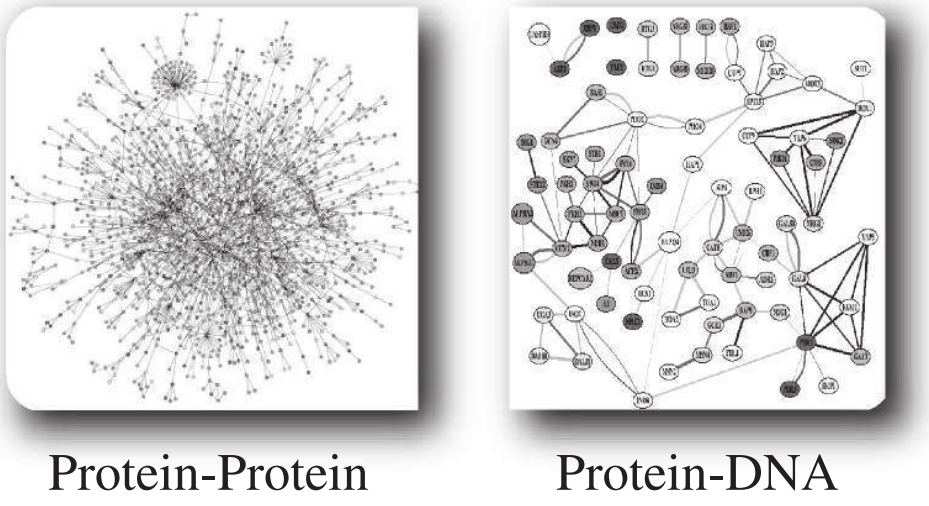


Figure 1.2 Two examples of networks that are defined by high-throughput chemical assays. Images courtesy of Markus Herrgard.

such as the JAK-STAT signaling network in lymphocytes and the Toll-like receptor system in macrophages. A growing number of dynamic models for individual signaling pathways are becoming available.

Regulatory networks There is a complex network of interactions that determine the DNA binding state of most proteins, which in turn determine whether genes are being expressed. The RNA polymerase must bind to DNA, as do transcription factors and various other proteins. The details of these chemical interactions are being worked out, but in the absence of such information, most of the network models that have been built are discrete, stochastic, and logistical in nature.

With the rapid development of experimental methods that measure expression states, the binding sites, and their occupancy, we may soon see large-scale reconstructions of transcriptional regulatory networks. Once these are available, we can begin to plan the process to build models that will describe their dynamic states.

Unbiased network definitions An alternative way to define networks is based on chemical assays. Measuring all protein–protein interactions regardless of function provides one such example; see Figure 1.2. Another example is a genome-wide measurement of the binding sites of a DNA-binding protein. This approach is driven by data-generating capabilities. It does not have an *a priori* bias about the function of molecules being examined.

Table 1.1 Web resources that contain information about biological networks (prepared by Jan Schellenberger)

		Metabolic	Protein–protein	Regulatory/signaling	Organisms	Curation[a]
KEGG	http://www.genome.jp/kegg/	x			many	C
BiGG	http://bigg.ucsd.edu/	x			many	M
BioCyc[b]	http://biocyc.org/	x		x	many	C/M
MetaCyc	http://metacyc.org/	x			many	C/M
Reactome	http://reactome.org/	x	x	x	many	M
BIND	http://www.bindingdb.org/		x		many	E/M
DIP	http://dip.doe-mbi.ucla.edu/		x		many	M
HPRD	http://www.hprd.org/		x		human	M
MINT	http://mint.bio.uniroma2.it/		x		many	M
Biogrid	http://www.thebiogrid.org/		x		many	E
UniHI	http://theoderich.fb3.mdc-berlin.de:8080/unihi/		x		human	E/M
Yeastract	http://www.yeastract.com/			x	yeast	M
TRANSFAC	http://www.gene-regulation.com			x	many	M
TRANSPATH	http://www.gene-regulation.com			x	many	M
RegulonDB	http://regulondb.ccg.unam.mx/			x	many	C/E
NetPath	http://www.netpath.org/			x	human	M

[a] M = manual/literature, C = computational, E = experimental.
[b] Links to other *Cyc databases.

Network reconstruction Metabolic networks are currently the best-characterized biological networks for which the most detailed reconstructions are available. The conceptual basis for their reconstruction has been reviewed [100], the workflows used detailed [35], and a detailed standard operating procedure (SOP) is available [117]. Some of the fundamental issues associated with the generation of dynamic models describing their functions have been articulated [52].

There is much interest in reconstructing signaling and regulatory networks in a similar way. The prospects for reconstruction of large-scale signaling networks have been discussed [49]. Given the development of new omics data types and other information, it seems likely that we will be able to obtain reliable reconstructions of these networks in the not too distant future.

Public information about pathways and networks There is a growing number of networks that underlie cellular functions that are being unraveled and reconstructed. Many publicly available sources contain this information; see Table 1.1. We wish to study the dynamic states of such networks. To do so, we need to describe them in chemical detail and

incorporate thermodynamic information and formulate a mathematical model.

1.2 Why build and study models?

Mathematical modeling is practiced in various branches of science and engineering. The construction of models is a laborious and detailed task. It also involves the use of numerical and mathematical analysis, both of which are intellectually intensive and unforgiving undertakings. So why bother?

Bailey's five reasons The purpose and utility of model building has been succinctly summarized and discussed [15]:

1. *"To organize disparate information into a coherent whole."* The information that goes into building models is often found in many different sources and the model builder has to look for these, evaluate them, and put them in context. In our case, this comes down to building data matrices (see Table 1.3) and determining conditions of interest.
2. *"To think (and calculate) logically about what components and interactions are important in a complex system."* Once the information has been gathered it can be mathematically represented in a self-consistent format. Once equations have been formulated using the information gathered and according to the laws of nature, the information can be mathematically interrogated. The interactions among the different components are evaluated and the behavior of the model is compared with experimental data.
3. *"To discover new strategies."* Once a model has been assembled and studied, it often reveals relationships among its different components that were not previously known. Such observations often lead to new experiments, or form the basis for new designs. Further, when a model fails to reproduce the functions of the process being described, it means there is either something critical missing in the model or the data that led to its formulation is inconsistent. Such an occurrence then leads to a re-examination of the information that led to the model formulation. If no logical flaw is found, the analysis of the discrepancy may lead to new experiments to try to discover the missing information.
4. *"To make important corrections to the conventional wisdom."* The properties of a model may differ from the governing thinking about

process phenomena that is inferred based on qualitative reasoning. Good models may thus lead to important new conceptual developments.

5. *"To understand the essential qualitative features."* Since a model accounts for all interactions described among its parts, it often leads to a better understanding of the whole. In the present case, such qualitative features relate to multi-scale analysis in time and an understanding of how multiple chemical events culminate in coherent physiological features.

1.3 Characterizing dynamic states

The dynamic analysis of complex reaction networks involves the tracing of time-dependent changes of concentrations and reaction fluxes over time. The concentrations typically considered are those of metabolites, proteins, or other cellular constituents. There are three key characteristics of dynamic states that we mention here right at the outset, and they are described in more detail in Section 2.1.

Time constants Dynamic states are characterized by change in time; thus, the rate of change becomes the key consideration. The rate of change of a variable is characterized by a *time constant*. Typically, there is a broad spectrum of time constants found in biochemical reaction networks. This leads to time-scale separation, where events may be happening on the order of milliseconds all the way to hours, if not days. The determination of the spectrum of time constants is thus central to the analysis of network dynamics.

Aggregate variables An associated issue is the identification of the biochemical, and ultimately physiological, events that are unfolding on every time scale. Once identified, one begins to form *aggregate concentration variables*, or *pooled variables*. These variables will be combinations of the original concentration variables. For example, two concentration variables may interconvert rapidly, on the order of milliseconds and thus on every time scale longer than milliseconds these two concentrations will be "connected." They can, therefore, be "pooled" together to form an aggregate variable. An example is given in Figure 1.3.

The determination of such aggregate variables becomes an intricate mathematical problem. Once solved, it allows us to determine the *dynamic structure of a network*. In other words, we move hierarchically away from the original concentration variables to increasingly interlinked aggregate

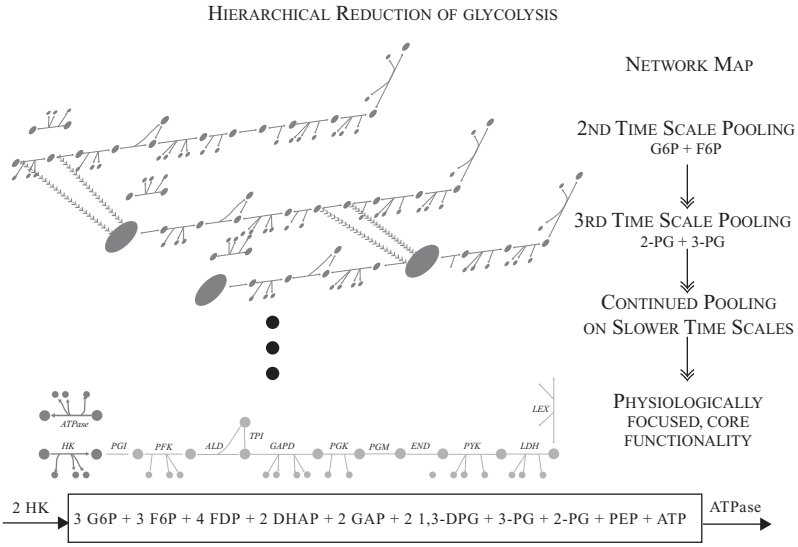


Figure 1.3 Time-scale hierarchy and the formation of aggregate variables in glycolysis. The "pooling" process culminates in the formation of one pool (shown in a box at the bottom) that is filled by hexokinase (HK) and drained by ATPase. This pool represents the inventory of high-energy phosphate bonds. From [52].

variables that ultimately culminate in the overall dynamic features of a network on slower time scales. Temporal decomposition, therefore, involves finding the time-scale spectrum of a network and determining what moves on each one of these time scales. A network can then be studied on any one of these time scales.

Transitions Complex networks can transition from one steady state (i.e., homeostatic state) to another. There are distinct types of transition that characterize the dynamic states of a network. Transitions are analyzed by *bifurcation theory*. The most common bifurcations involve the emergence of *multiple steady states*, *sustained oscillations*, and *chaotic* behavior. Such dynamic features call for a yet more sophisticated mathematical treatment. Such changes in dynamic states have been called *creative functions*, which in turn represent willful physiological changes in organism behavior. In this book, we will only encounter relatively simple types of such transition.

1.4 Formulating dynamic network models

Approach Mechanistic kinetic models based on differential equations represent a *bottom-up approach*. This means that we identify all the

Table 1.2 Assumptions used in the formulation of biological network models

Assumption	Description
(1) Continuum assumption	Do not deal with individual molecules, but treat medium as a continuum
(2) Finer spatial structure ignored	Medium is homogeneous
(3) Constant-volume assumption	V is time-invariant, $dV/dt = 0$
(4) Constant temperature	Isothermal systems Kinetic properties a constant
(5) Ignore physico-chemical factors	Electroneutrality and osmotic pressure can be important factors, but are ignored

detailed events in a network and systematically build it up in complexity by adding more and more new information about the components of a network and how they interact. A complementary approach to the analysis of a biochemical reaction network is a *top-down approach*, where one collects data and information about the state of the whole network at one time. This approach is not covered in this text but typically requires a Bayesian or Boolean analysis that represents causal or statistically determined relationships between network components. The bottom-up approach requires a mechanistic understanding of component interactions. Both the top-down and bottom-up approaches are useful and complementary in studying the dynamic states of networks.

Simplifying assumptions Kinetic models are typically formulated as a set of deterministic ordinary differential equations (ODEs). There are a number of important assumptions made in such formulations that often are not fully described and delineated. Five assumptions will be discussed here (Table 1.2).

1. Using deterministic equations to model biochemistry essentially implies a "clockwork" of functionality. However, this modeling assumption needs justification. There are three principal sources of variability in biological dynamics: internal thermal noise, changes in the environment, and cell-to-cell variation. Inside cells, all components experience thermal effects that result in *random molecular motion*. This process is, of course, one of molecular diffusion, called *Brownian motion* with larger observable objects. The ODE assumption involves taking an ensemble of molecules and averaging out the stochastic effects. In cases where there are very few molecules of a particular species inside a cell or a cellular compartment, this assumption may turn out to be erroneous.

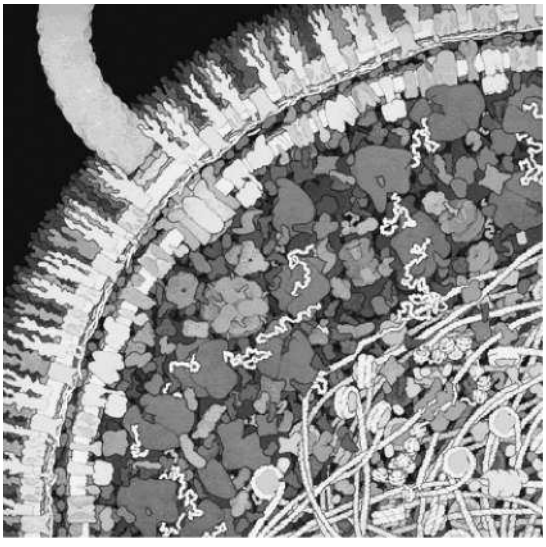


Figure 1.4 The crowded state of the intracellular environment. Some of the physical characteristics are viscosity ($>100 \times \mu_{H_2O}$), osmotic pressure (<150 atm), electrical gradients (≈300 000 V/cm), and a near-crystalline state. ©David S. Goodsell 1999.

2. The finer architecture of cells is also typically not considered in kinetic models. Cells are highly structured down to the 100 nm length scale and are thus not homogeneous (see Figure 1.4). Rapidly diffusing compounds, such as metabolites and ions, will distribute quickly throughout the compartment and one can justifiably consider the concentration to be relatively uniform. However, with larger molecules whose diffusion is hindered and confined, one may have to consider their spatial location. Studying and describing cellular functions of the 100 nm length scale is likely to represent an interesting topic in systems biology as it unfolds.
3. Another major assumption in most kinetic models is that of *constant volume*. Cells and cellular compartments typically have fluctuations in their volume. Treating variable volume turns out to be mathematically difficult. It is, therefore, often ignored. However, minor fluctuations in the volume of a cellular compartment may change all the concentrations in that compartment and, therefore, all kinetic and regulatory effects.
4. Temperature is typically considered to be a constant. Larger organisms have the capability to control their temperature. Small organisms have a high surface-to-volume ratio, making it hard to control heat flux at their periphery. Further, small cellular dimensions lead to rapid thermal diffusivity and a strong dependency on the thermal characteristics of the environment. Rate constants are normally a strong function of temperature, often described by the Arrhenius law. Thus, treating cells as isothermal systems is a simplification under which the kinetic properties are described by kinetic *constants.*

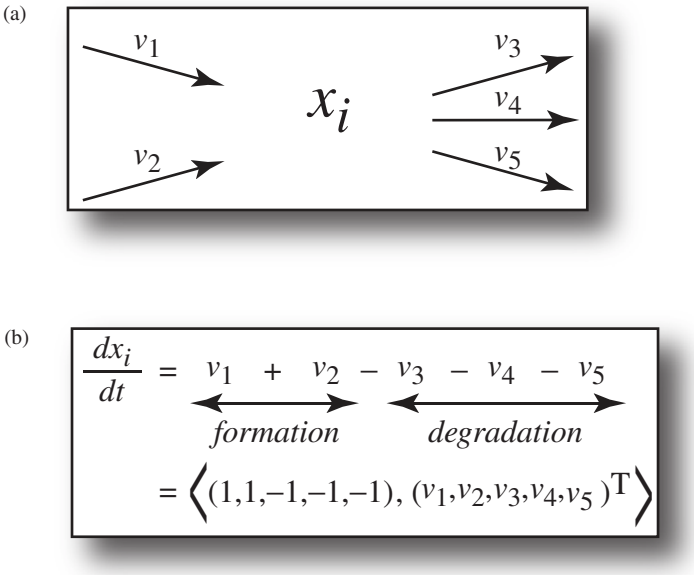


Figure 1.5 The dynamic mass balance on a single compound. (a) All the rates of formation and degradation of a compound x_i (a graphical representation called a node map). (b) The corresponding dynamic mass balance equation that simply states that the rate of change of the concentrations x_i is equal to the sum of the rates of formation minus the sum of the rates of degradation. This summation can be represented as an inner product between a row vector and the flux vector. This row vector becomes a row in the stoichiometric matrix in Eq. (1.1).

5. All cells and cellular compartments must maintain electroneutrality; therefore, the exchange of any species in and out of a compartment or a cell must also obey electroneutrality. Considering the charge of molecules and their pH dependence is yet another complicated mathematical subject and, thus, often ignored. Similarly, significant internal osmotic pressure must be balanced with that of the environment. Cells in tissues are in an isotonic environment, while single-cellular organisms and cells in plants build rigid walls to maintain their integrity.

The dynamic mass balance equations Applying these simplifying assumptions, we arrive at the dynamic mass balance equations as the starting point for modeling the dynamic states of biochemical reaction networks. The basic notion of a dynamic mass balance on a single compound, x_i, is shown in Figure 1.5.

The combination of all the dynamic mass balances for all concentrations $\mathbf{x}$ in a biochemical reaction network are given by a matrix equation:

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v}(\mathbf{x}), \tag{1.1}$$