

## Chapter 2

# Perturbation-Response Approach for Biological Network Analysis

A cell contains thousands of genes/proteins/metabolites in a highly inhomogeneous intracellular environment with spatiotemporal effects of molecular crowding and diffusion. The molecules are highly interconnected, leading to complex networks. Hence, there is precedence that all molecular interactions with their detailed reaction kinetics and spatial organization are required to be known for the proper understanding of biological responses. In fact, the goal of the human genome project is to first generate a complete parts list of genetic materials within cells and, second, from it construct the detailed regulatory features that connect them [1, 2]. However, even in such perceived complexity involving myriad interacting components, simple mass-action models using a limited set of crucial molecules, constituting functional biological modules, were successfully used to interpret dynamic responses (see Chap. 1). How can such simplicity be held within a complex system?

For a long time, biological networks have been investigated on a much smaller scale through in vitro experiments (e.g., the 10-step reaction analysis of glycolysis; see Fig. 1.5b). This approach was mainly in the hope to understand and control all reaction kinetics within the entire known connectivity in detail. In most circumstances, the investigations considered a ‘closed’ system, where there was no exchange of material between the internal and external environments. Hence, chemical and thermal equilibrium can be assumed, and as a result, the law of mass action is expected to work within the system. In other words, the approach describes a well-mixed, homogeneous, and isothermal environment where each reaction in the network is connected through first-order, higher-order mass-action or enzyme kinetic equations, depending on the knowledge gained for an individual reaction (Box 2.1).

Most, if not all, studies adopting in vitro experiments determine the parameter values of reaction species for computational modeling from an artificial environment where the species are deliberately purified from its physiological neighbors. This approach is because, until now, in vivo kinetic parameters could be reliably measured using current experimental technologies. Notably, there have been various reports that claim the kinetic parameters determined through in vitro and in vivo experiments can differ by several orders of magnitude [3]. As a result, when

### Box 2.1 Mass-Action Kinetics

A chemical system is perturbed from a stable steady state. When the system subsequently returns to its original state, the system is said to be in equilibrium. For example, heating water at room temperature and subsequently removing the heat source results in the decay of the water temperature to its original reading. This event happens in a 'closed' system where external conditions, such as pressure, energy, or mass, do not enter or leave the system that is in cooling.

Let us consider a simple case involving only the decay process, that is, the point onward when a heat source is removed. The temperature ( $T$ ) changes over time following Newton's law of cooling:

$$\frac{dT}{dt} = -k[T - T_0]$$

where  $T_0$  is the room temperature and  $k$  is the rate of cooling. The minus sign indicates temperature drop in time.

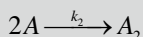
Now consider a closed chemical reaction of molecular species A into B:



The arrow indicates the equilibrium state lies far to the right, that is, the reverse reaction ( $B \rightarrow A$ ) proceeds only at an infinitesimal extent. For every species B formed (concentration units in moles), an A species disappears:

$$\frac{d[B]}{dt} = \frac{-d[A]}{dt} = k_1[A]$$

The constant  $k_1$  is called the rate constant and has the unit of per second. The rate constant provides a direct measure of how fast this reaction is occurring. The higher the  $k_1$  value, the faster the reaction. This type of reaction is called a first-order reaction, as its rate depends on the first power of the reactant concentration. A second-order reaction occurs typically when two species react to form another species. An example:



The rate of such a reaction is proportional to the second power of the concentration of reactant, for the reaction can occur only when two molecules collide:

$$\frac{-d[A]}{dt} = k_2[A]^2$$

where  $k_2$  is the second-order rate constant. It has dimensions of  $(\text{mol/L})^{-1} \text{s}^{-1}$ .

(continued)

**Box 2.1** (continued)

It follows that for any system in chemical equilibrium, the rate of an elementary reaction is proportional to the product of the concentrations of the reacting species.

**Enzyme Kinetics**

For reactions that require the aid of other species, such as enzymes that enhance catalytic reactions, the mechanism to account for such reactions assumes that the species  $A$  combines with the species  $E$  in a reversible manner to give complex  $EA$ , which then dissociates reversibly or reacts irreversibly to produce  $B$  while leaving  $E$  unchanged.



In this case, the rate of  $B$  formation can be shown to be [9]:

$$\frac{dB}{dt} = \frac{V_{\max} [A]}{[A] + K_M}$$

where  $V_{\max} = k_{cat} [E]_t$  and  $K_M = \frac{k_{cat} + k_{-1}}{k_1}$

The reaction rate increases with increasing  $[A]$ , approaching an asymptote at  $V_{\max}$ , when all enzyme is bound to  $A$ .  $[E]_t$  is the total enzyme concentration and  $k_{cat}$  is the maximum number of enzymatic reactions catalyzed per second.

There are various forms of enzyme kinetics, depending on the types of intermediates or cofactors affecting the overall reactions [9].

combining these errors into the model, the final predictions could differ by several orders of magnitude. For example, the steady-state concentration of the glycolytic metabolite 3-phosphoglycerate in *Trypanosoma brucei* was underpredicted by an order of seven [4]. Hence, despite modularizing a small functional network in biology, the computational prediction of steady-state glycolytic metabolite concentrations may not reflect realistic in vivo values because of the lack of accuracy in determining kinetic reaction parameter values using an in vitro system.

Moreover, the kinetic approach requires (i) a priori knowledge of all detailed connectivities of reacting species and (ii) the assumption of steady-state conditions. This requirement highly limits their applicability for analyzing a cellular system where the knowledge of network topology is sparse, for example, in signal transduction and gene regulatory networks, and for interpreting non-steady-state or

complex dynamic response to external perturbation. Therefore, there have been numerous efforts to find alternative theoretical and computational methodologies that would analyze complex dynamics and overcome the difficulty of not having enough *in vivo* biological data.

One popular method that overcomes the issue of parameter reliability is the flux-balance analysis (FBA) [5]. Here, only the reaction topologies or stoichiometry of the network need be known. Constraints are introduced by the stoichiometric coefficients in the system for the optimization of certain biological function, such as growth or production of certain compounds. Although the FBA requires the assumption that metabolite concentrations remain at the steady state for analysis, it has been successfully used to interpret important physiological functions of a living cell. For example, Palsson and colleagues experimentally verified their prediction for the primary carbon source and oxygen uptake rates for maximal cellular growth in *Escherichia coli* [6]. Nevertheless, the FBA requires steady-state response and connectivity of all species in a biological network to be known, which highly limits its applicability for signaling or gene network analysis.

On the other hand, as mentioned in Chap. 1, the use of a simple mass-action model in a highly limited module of tumor necrosis factor (TNF), lipopolysaccharide (LPS), epidermal growth factor (EGF), and nerve growth factor (NGF) signaling has produced very useful insights into the regulation of key transcription factors and their gene expressions. These achievements may seem to contradict studies that aim to understand the details of the entire biological part list before interpreting complex dynamics. The truth, obviously, depends on the type of biological questions that one seeks to find solution. In the case of the signaling studies mentioned, the goal was to understand how the IKK and MAPK module responded to different stimuli. The investigations demonstrated that it was not necessary to know the detailed dynamics of each signaling species activated for each stimulus to understand their core regulatory features. So, why is it that not all details are required to understand IKK and MAPK modular response or to understand the physiological growth rate of *E. coli*?

The answer may lie within the organizing principles observed for biological networks. Chapter 1 highlighted that biological networks are organized in a scale-free manner, where controlling the ‘nodes’ would not overtly change the stable response of the network, while removing the ‘hubs’ is likely to produce an unstable or deleterious response. In a pioneering work on understanding *E. coli* chemotactic behavior, Leibler and colleagues showed that the adaptation precision of bacterial chemotaxis was insensitive to the large variation of its network parameter values [7, 8]. This mechanism, therefore, allows *E. coli* to display robust behavior to a wide range of attractant and repellent concentrations. However, at the same time, other properties, such as adaptation time and steady-state tumbling frequency, were variable with stimulant concentration. Overall, the work concludes that the core network structure does not depend on the precision of its parameter values. This important finding is fundamental to the success of the numerous simple models, such as those for IKK and MAPK modules or for carbon metabolism, where it is the network structure rather than reaction kinetics that is crucial for the robust and repeatable behavior of the system.

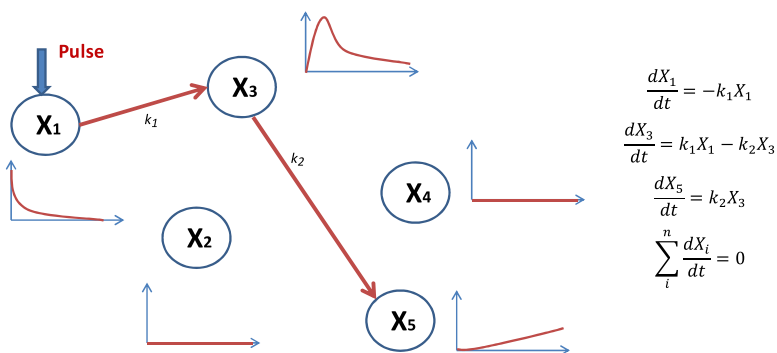
In this chapter, and for most of this book, we focus on the “perturbation-response” approach, where instead of trying to understand the microscopic details of the detailed kinetics of each reaction, a macroscopic or ‘top-down’ view of the biological network behavior is used to understand the system structure and behavior. More technically, instead of measuring each reaction kinetics to investigate biological network response, one can alternatively observe the dynamic patterns of the network species to a given perturbation, and then directly determine the response parameter values for the model.

## The Perturbation-Response Approach

Let a fixed perturbation be given to one of the species in a cellular system: this will result in the propagation of response waves among those species that interact. Obviously, those species that are not connected will remain at baseline or steady-state levels. Figure 2.1 schematically illustrates a hypothetical system with five species. Here, the perturbation of  $X_1$  results in the response of species  $X_3$  and  $X_5$  only, suggesting the existence of connectivity between the three species. Next, by analyzing the time to reach peaks, the actual peak levels, and the decay rates, the order of causal relationships or connectivity between the species can be determined.

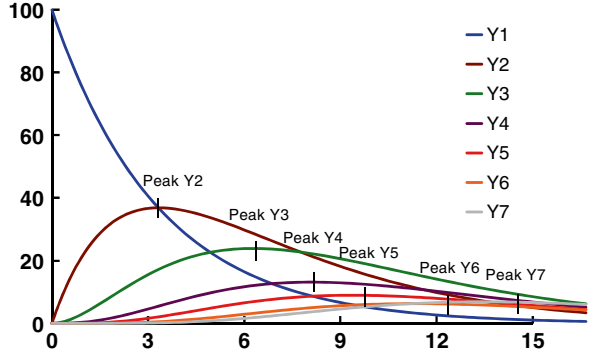
For example, in a closed system, if the concentration of (i)  $X_1$  decreases only, (ii)  $X_3$  increases, go through a maximum, and decreases, and (iii)  $X_5$  increases only, by the law of mass flow conservation it can be shown that the connectivity is a linear chain of reactions:  $X_1 \rightarrow X_3 \rightarrow X_5$  (Fig. 2.1). That is, the temporal order of responses can yield the causal connectivity of the species in the reaction mechanism.

To illustrate this further, let us construct a linear chain of reactions for  $Y$  species ( $Y_1 \rightarrow Y_2 \rightarrow Y_3 \rightarrow \dots \rightarrow Y_{N-1} \rightarrow Y_N$ ) using first-order mass-action response equations. By applying a pulse perturbation to the concentration of the first species,  $Y_1$ , we observe the propagation of response waves along the pathway (Fig. 2.2). The experiments



**Fig. 2.1** Essence of the perturbation-response approach. The summation of all reaction rates will be zero in a mass-conserved closed system

**Fig. 2.2** *Propagation of response waves.* Temporal profiles of species concentration in a linear chain reaction network for pulse perturbation of species  $Y_1$  (with  $N=7$ ; see text). The units are arbitrary, scaled by rate coefficients

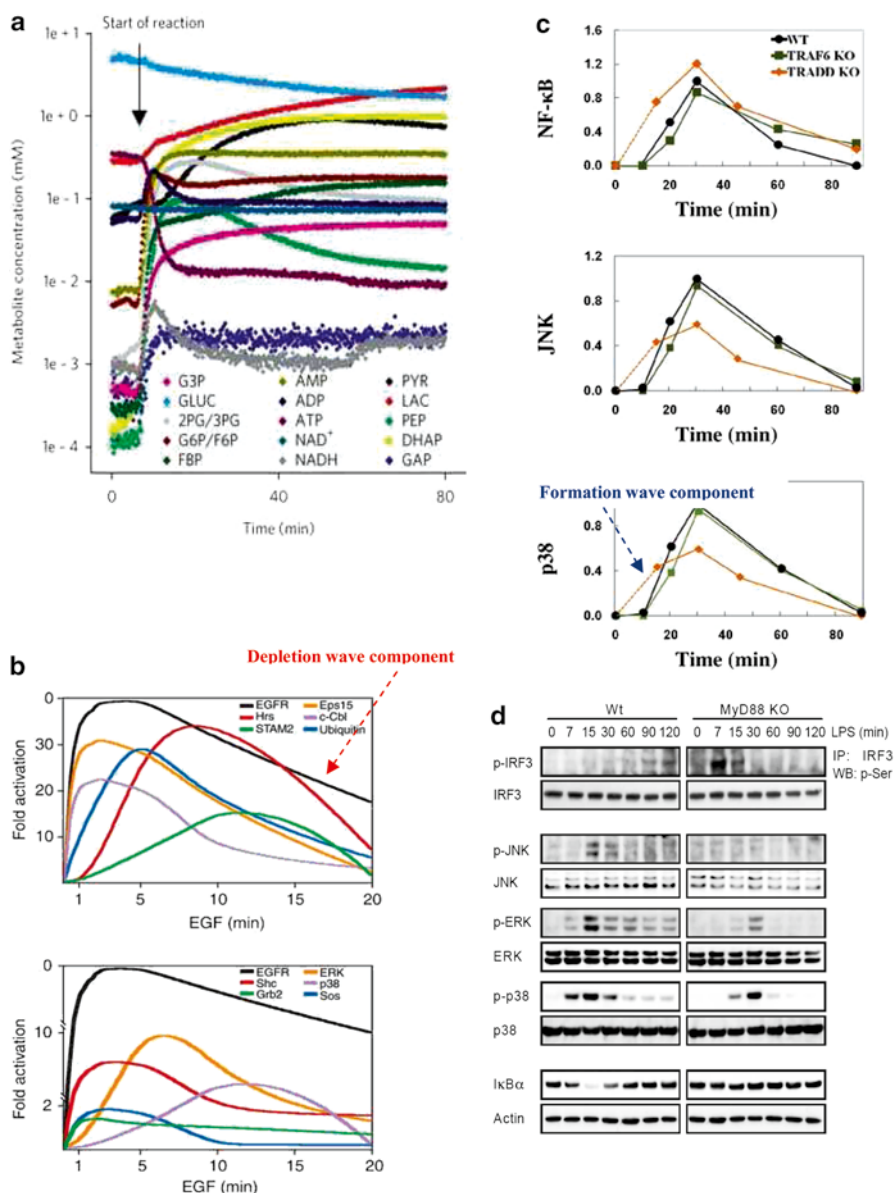


preserve the total input and output fluxes. Under such conditions, a linear superposition of propagation response waves connects the species between input and output fluxes. Thus, simple linear rules can be derived for the system, notably, (1) the time to reach peak values (*Peak Y2*, *Peak Y3*, etc., in Fig. 2.2) increases and its amplitude decreases as one moves down the reaction network, unless there are other features such as feedback reactions; (2) the initial response gradient can be used to determine the location of a reactant species in a network, that is, the steepest gradient is the closest to the perturbed reactant species and the lowest gradient is the farthest; (3) reactant species that are not connected to the system do not show any response profile; and (4) as the law of mass conservation is used for pulse propagation, the sum of all species deviations from the steady state (weighted by stoichiometric coefficients) is constant. Therefore, it can help to determine the correct stoichiometric coefficients [10].

Despite the simplicity of the approach, linear response is visually apparent in the dynamic response profiles of several intracellular molecules activated in glycolysis, EGF- and TLR3/4 signaling to their respective perturbations (Fig. 2.3a–d). In other words, the perturbation-response analysis clearly provides evidence that complex biological networks could be governed by simple mass-action response equations, resulting in the linear superposition of propagation response waves.

To understand why biological network dynamics can follow linear response, let us reconsider pulse perturbation,  $(\alpha, 0)$  at  $t=0$ , given to a simple two-species chain governed by first-order mass-action equations;  $\mathbf{X}=(X_1, X_2)$ :  $X_1 \xrightarrow{k_1} X_2 \xrightarrow{k_2} \rightarrow$ . The perturbation wave  $\delta\mathbf{X}=(\delta X_1, \delta X_2)$ , applied to the system with rate constants  $k_1$  and  $k_2$  for  $X_1$  and  $X_2$ , where  $X_1$  has  $k_1$  as depletion and  $X_2$  has  $k_1$  as formation and  $k_2$  as depletion, can be represented by

$$\frac{d\delta\mathbf{X}}{dt} = \begin{pmatrix} -k_1 & 0 \\ k_1 & -k_2 \end{pmatrix} \delta\mathbf{X} \quad (2.1)$$



With initial conditions,  $\mathbf{X}_0$ , solving Eq. 2.4 yields the sum-of-exponentials:

$$\delta X_1 = \alpha e^{-k_1 t} \quad (2.2)$$

$$\delta X_2 = \frac{k_1}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t}) \quad (2.3)$$

Factorize Eq. 2.3 with respect to  $e^{-k_1 t}$  if  $k_2 > k_1$  (or  $e^{-k_2 t}$  if  $k_1 > k_2$ ) and obtain:

$$\delta X_2 = \frac{k_1}{k_2 - k_1} (1 - e^{-(k_2 - k_1)t}) e^{-k_1 t} \quad (2.4)$$

Equation 2.4 can be rewritten to constitute both the *formation* and *depletion* wave terms:

$$\delta X = \alpha \times \underbrace{(1 - e^{-p_1 t})}_{\text{Formation wave}} \times \underbrace{e^{-p_2 t}}_{\text{Depletion wave}} \quad (2.5)$$

↙
↓
↘  
 Perturbation                      Formation wave                      Depletion wave

where  $\alpha$  represents the amount of perturbation and  $p_1$  and  $p_2$  represent the measure of formation and depletion response propagation waves, respectively. In general,  $p_1$  and  $p_2$  are not equal to  $k_1$  and  $k_2$ , respectively, and are determined by fitting with experimental data. Thus, the formation and depletion terms observed for several biological network responses (Fig. 2.3) are likely to have their origins from first-order or linear response equations.

## The Origins of Linear Response

To investigate the origins of formation and depletion waves in the response of various complex biological networks outlined so far, let a stable network consisting of  $n$  species be perturbed from the reference steady state. In general, the resultant changes in the concentration of species are governed by the kinetic evolution equation:

$$\frac{\partial X_i}{\partial t} = F_i(X_1, X_2, \dots, X_n), i = 1, \dots, n \quad (2.6)$$

where the corresponding vector form of Eq. 2.1 is  $\frac{\partial \mathbf{X}}{\partial t} = \mathbf{F}(\mathbf{X})$ .  $\mathbf{F}$  is a vector of any nonlinear function including diffusion and reaction of the species vector  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ , which represents activated concentration levels of the reaction species (hence the partial derivative used). The response to perturbation can be written by  $\mathbf{X} = \mathbf{X}_0 + \delta \mathbf{X}$ , where  $\mathbf{X}_0$  is the reference steady-state vector and  $\delta \mathbf{X}$  is the relative response from steady states ( $\delta \mathbf{X}_{t=0} = \mathbf{0}$ ).

The generally nonlinear kinetic evolution equation (Eq. 2.6) can be approximated or linearized by using the Taylor series:

$$\frac{\partial \delta \mathbf{X}}{\partial t} = \left. \frac{\partial F(\mathbf{X})}{\partial \mathbf{X}} \right| \delta \mathbf{X} + \left. \frac{\partial F^2(\mathbf{X})}{\partial \mathbf{X}^2} \right| \delta \mathbf{X}^2 + \dots \quad (2.7)$$

As the general volume of perturbing substance is usually very small (order of 1 %) compared to the total volume of cells that are perturbed, now consider a small perturbation around the steady state in Eq. 2.7, in which higher-order terms become negligible, and resulting in the approximation of the first-order term.

In vector form  $\frac{d\delta \mathbf{X}}{dt} \equiv \left. \frac{\partial F(\mathbf{X})}{\partial \mathbf{X}} \right|_{\mathbf{X}=\mathbf{X}_0} \delta \mathbf{X}$  (note the change from partial derivative to total derivative of time), where the zero-order term  $F(\mathbf{X}_0)=0$  at the steady-state  $\mathbf{X}_0$  and the *Jacobian* matrix, or linear stability matrix, is  $\mathbf{J} = \left. \frac{\partial F(\mathbf{X})}{\partial \mathbf{X}} \right|_{\mathbf{X}=\mathbf{X}_0}$ . The elements

of  $\mathbf{J}$ , based on the initial activation topology, are chosen by fitting  $\delta \mathbf{X}$  with corresponding experimental profiles. Hence, the amount of response (flux propagated) along a biological pathway can be approximated using *first-order mass-action response*, i.e.,  $\frac{d\delta \mathbf{X}}{dt} = \mathbf{J} \delta \mathbf{X}$ . That is, the basic principle so far suggests that the response rate of species in a mass-conserved system at an initial steady state can be approximated by the first-order mass-action response equation, given a small perturbation to one or more species.

Note that Jacobian matrix elements (or response coefficients) can represent not only reaction information but also spatial information such as diffusion and transport mechanisms. Thus, each species in the perturbation-response model can represent a molecule, a different modified state of a molecule (e.g., ubiquitinated state), or a molecular process such as diffusion or endocytosis. That is, each species in the biological network does not necessarily represent a specific molecular species. For illustration, in a pathway  $q1 \rightarrow q2 \rightarrow q3 \rightarrow q4 \rightarrow q5$ ,  $q1$  to  $q5$  can each be a different protein or the same protein at different stages in signaling, for example,  $q1$  being internalized ( $q2$ ), transported to a different organelle ( $q3$ ), ubiquitinated ( $q4$ ), and become part of a protein complex ( $q5$ ).

Thus, in contrast to bottom-up kinetic models, which use fixed network topology, the perturbation-response approach considers the network as a sequence of events rather than just molecular species. As molecular networks are largely not fully understood, this difference is crucial as it prevents rigidly fixing the network topology, and allows it to be modified according to experimental data so as to prevent overfitting problems and, as a consequence, identify novel features of biological networks.

In addition, as cellular processes involve a large number (thousands) of intracellular molecular interactions, it is currently not plausible to model the dynamics of all possible reactions with the generally limited data. To overcome such difficulties, this approach permits the lumping of several molecules into a species, and the

resultant averaging nature of the response equations does not require detailed kinetics. This postulation will become clearer in later chapters where the lumping of several molecules in signaling networks does not distort the overall properties of the system response.

## Limitations of Linear Response Approach

As with all modeling approaches, there are certain limitations that require mentioning. First, the linear response approach does not comprehensively model the details of the kinetics of an individual signaling reaction. Second, the small perturbation assumption leading to the deterministic first-order mass-action equations represents an average cell response in a well-mixed homogeneous environment. Hence, the approach cannot be used to study single-cell stochastic behavior or oscillatory dynamics in a heterogeneous environment (see Chap. 9). Third, the model predictions show relative, and not absolute, activation levels. Fourth, the first-order response may be used only for a relatively short primary response period (up to 2 h), where secondary effects such as autocrine or paracrine feedback regulations are usually not significant. Fifth, the first-order response is best suited for a fixed perturbation strategy; that is, the model cannot be tested with the same parameter values across a variable range (volume) of perturbing stimuli. In that situation, a different set of parameters or network features (e.g., feedback loop) could suffice for different levels of perturbation. Although linear response is observed for numerous cellular responses, their application may not necessarily be applicable in several other cellular processes, such as in interpreting cell fate decision or neural signaling, where nonlinear approaches are required [15, 16].

Nevertheless, the linear response approach is not restricted to specific pathways. It can be applied to model any pathways that experimentally display formation and depletion waves (Fig. 2.3). Although the position of a particular species in a reaction network requires all related species be identified and measured, simple models using linear response that lump several molecular species can still provide useful information in revealing overt novel network features, for example, the identification of distinct IKK–I $\kappa$ B–NF- $\kappa$ B network wirings in LPS, TNF, EGF, and NGF stimulation (see Chap. 1). More examples with better details are shown in the following chapters through the investigations of various innate immune signaling networks, where several missing components from the current knowledge of their network species are revealed.

In the following chapter, Chap. 3, we introduce some basic aspects in immunology and Toll-like receptor (TLR) signaling. Chapters 4, 5, 6, and 7 consider the use of the linear response approach to analyze the dynamics of TLR4 (Chaps. 4 and 5), TLR3 (Chap. 6), TNF (Chap. 7), and TRAIL (Chap. 8) signaling. Chapter 9 onwards introduce the issue of nonlinearities observed in biology and set the tone for future systems biology research.

## References

1. Venter JC et al (2001) The sequence of the human genome. *Science* 291:1304–1351
2. Venter JC (2011) Genome-sequencing anniversary. The human genome at 10: successes and challenges. *Science* 331:546–547
3. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* 3:1871–1878
4. Bakker BM, Michels PA, Oppendoes FR, Westerhoff HV (1997) Glycolysis in bloodstream-form *Trypanosoma brucei* can be understood in terms of the kinetics of the glycolytic enzymes. *J Biol Chem* 272:3207–3215
5. Edwards JS, Palsson BO (1998) How will bioinformatics influence metabolic engineering? *Biotechnol Bioeng* 58:162–169
6. Edwards JS, Ibarra RU, Palsson BO (2001) In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 19:125–130
7. Barkai N, Leibler S (1997) Robustness in simple biochemical networks. *Nature* 387:913–917
8. Alon U, Surette MG, Barkai N, Leibler S (1999) Robustness in bacterial chemotaxis. *Nature* 397:168–171
9. Segel IH (1993) *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady state enzyme systems*. Wiley, New York
10. Vance W, Arkin A, Ross J (2002) Determination of causal connectivities of species in reaction networks. *Proc Natl Acad Sci USA* 99:5816–5821
11. Bujara M, Schümperli M, Pellaux R, Heinemann M, Panke S (2011) Optimization of a blueprint for in vitro glycolysis by metabolic real-time analysis. *Nat Chem Biol* 7:271–277
12. Blagoev B, Ong SE, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 22:1139–1145
13. Helmy M, Gohda J, Inoue J, Tomita M, Tsuchiya M, Selvarajoo K (2009) Predicting novel features of toll-like receptor 3 signaling in macrophages. *PLoS One* 4:e4661
14. Selvarajoo K, Takada Y, Gohda J, Helmy M, Akira S, Tomita M, Tsuchiya M, Inoue J, Matsuo K (2008) Signaling flux redistribution at toll-like receptor pathway junctions. *PLoS One* 3:e3430
15. Maamar H, Raj A, Dubnau D (2007) Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* 317:526–529
16. Longtin A (1993) Stochastic resonance in neuron models. *J Stat Phys* 70:309–327

Immuno Systems Biology

A macroscopic approach for immune cell signaling

Selvarajoo, K.

2013, IX, 146 p. 69 illus., 61 illus. in color.,

ISBN: 978-1-4614-7690-0