Part A

METHODOLOGICAL CHAPTERS

1

Two Challenges of Systems Biology

William S. Hlavacek

Theoretical Division, Los Alamos National Laboratory, Los Alamos, USA

1.1 Introduction

Articulating the challenges faced by a field, or rather the problems that seem worthy of pursuit, can be a useful exercise. A famous example is Hilbert's problems (Hilbert 1902), which profoundly influenced mathematics. This entire handbook can be viewed as an attempt to define challenges faced by systems biologists, especially challenges requiring a statistical approach, and to provide tools for addressing these challenges. This particular chapter provides a personal perspective. If the editors had invited someone else to write it, I am sure it would have materialized in a very different form, as there are many challenges worthy of pursuit in this field.

Much has been written about systems biology, and excellent overviews of the field are available (for example, see Kitano 2002). For many years, I believe one of the challenges faced by systems biology has been community building. In recent years, thanks in part to events such as the International Conference on Systems Biology (Kitano 2001) and the q-bio Summer School and Conference (Edwards *et al.* 2007), a community of systems biology researchers has established itself. This community identifies with the term 'systems biology' and broadly agrees upon its loose definition and the general goals of the field. If one is looking for a definition of systems biology, the work presented at the meetings mentioned above serves the purpose well.

A characteristic of systems biology is the systems approach, which is aided by modeling. In systems biology, there has been special interest in the study and modeling of cellular regulatory systems, such as genetic circuits (Alon 2006). This chapter will be focused on a discussion of challenges faced by modelers. Of course, modeling would be aided by technology development, the generation of new quantitative data and in other ways, but I will be concerned mainly with the practice of modeling. Moreover, I will focus on modeling of cell signaling systems, although I believe the discussion is relevant for modeling of other types of cellular regulatory systems.

We need models of cellular regulatory systems to accelerate elucidation of the molecular mechanisms of cellular information processing, to understand design principles of cellular regulatory systems (i.e. the

relationships between system structures and functions), and ultimately, to engineer cells for useful purposes. Cell engineering encompasses a number of application areas, including manipulation of cellular metabolism for production of desirable metabolic products (Keasling 2010), such as biofuels; identification of drug targets for effective treatment or prevention of diseases (Klipp *et al.* 2010), such as cancer; and creation of cells with entirely new functions and properties through synthetic biology approaches (Khalil and Collins 2010). In each of these areas, predictive models can be used to guide the manipulation of cellular phenotypes through interventions at the molecular level (e.g. genetic modifications or pharmacological perturbations). In traditional engineering fields, predictive models play a central role, and I believe predictive modeling will be just as central to cell engineering efforts in the future. Another more immediate reason to seek models of cellular regulatory systems is that model-guided studies of the molecular mechanisms of cellular information processing can accelerate the pace at which these mechanisms and their design principles are elucidated (Wall *et al.* 2004; Alon 2006; Novák and Tyson 2008; Rafelski and Marshall 2008; Mukherji and van Oudenaarden 2009). A model of a cellular regulatory system essentially represents a hypothesis about how the system operates and such a model can be used to interpret experimental observations, to design experiments, and generally to extend the reach of human intuition and reasoning. A model can be constructed in a systematic step-by-step fashion and the logical consequences of a model, no matter how complicated, can usually be determined. Although models are useful because they make predictions, models are also useful for other reasons, which are sometimes overlooked, as recently discussed by Lander (2010).

It is often assumed, presumably because of the long tradition of modeling in science and engineering and the many successes of modeling (e.g. celestial modeling), that modeling of cell signaling systems should be fairly straightforward and that all the necessary tools (e.g. numerical methods and software implementations of these methods) should be more or less readily available. I do not believe that this is the case or even close to the case. I believe that predictive modeling of cell signaling systems will require the solution of unique and difficult problems and that there is much work to be done.

1.2   Cell signaling systems

Cell signaling (Hunter 2000; Scott and Pawson 2009) controls the phenotypes and fates of cells. By conveying information through a signal-initiated cascade of molecular interactions, a cell signaling system enables a cell to sense changes in its environment and respond to them. Cell signaling is mediated in large part by complex networks of interacting proteins, each of which is itself complex. A signaling protein may contain a catalytic subunit that is tightly regulated, such as a tyrosine kinase (Bradshaw 2010; Lemmon and Schlessinger 2010), as well as other functional components, including modular protein interaction domains [e.g. a Src homology 2 (SH2) domain] (Pawson and Nash 2003), short linear motifs (e.g. a proline-rich sequence that can be recognized by an SH3 domain) (Gould *et al.* 2010), and sites of post-translational modification (e.g. substrates of kinases and phosphatases) (Walsh *et al.* 2005). Many of the components of signaling proteins enable and regulate interactions with binding partners, and regulated assembly of protein complexes plays an important role in cell signaling by co-localizing enzymes and substrates, which affects both enzyme activity and specificity (Ptashne and Gann 2001). Although many of the molecular-level properties of specific cell signaling systems have been elucidated and characterized, we are only now beginning to investigate system-level properties, i.e. how the parts work together to process information (Aldridge *et al.* 2006; Kholodenko 2006; Kholodenko *et al.* 2010).

A systems-level understanding of cell signaling is central to a basic understanding of cell biology. This type of understanding is also relevant for finding better treatments for a number of diseases involving dysregulation of cell signaling systems, such as cancer. The malignancy of cancer cells is driven and/or sustained in part by mutations that affect signaling proteins, many of which are the products of oncogenes and tumor

suppressor genes (Hanahan and Weinberg 2000). We can hope to reverse the effects of deleterious mutations by understanding the system-level properties of cell signaling systems (Kreeger and Lauffenburger 2010; Xu and Huang 2010). However, the complexity of cell signaling systems poses a formidable barrier to understanding and rational manipulation of system-level properties. There is a pressing need to find better ways to reason about cell signaling, so that we can relate the molecular interactions in cell signaling systems to the higher level phenomena that emerge from these interactions (Nurse 2008). Addressing this need now seems timely, given recent advances in proteomics that allow quantitative multiplex assays of the molecular states of cell signaling systems (Gstaiger and Aebersold 2009; Choudhary and Mann 2010; Kolch and Pitt 2010) and opportunities to take advantage of molecular therapeutics (Gonzalez-Angulo *et al.* 2010), i.e. drugs that specifically target signaling proteins, such as imatinib (Deininger *et al.* 2005; Hunter 2007).

A fairly standard aid for reasoning about a cell signaling system is a network wiring diagram, which is used to provide a visual representation and summary of available knowledge about a system, in particular the molecules involved and their interactions. Wiring diagrams come in many forms, from simple cartoons to elaborate depictions that make use of formalized iconography (Kohn 1999; Kitano *et al.* 2005). A testament to the importance of wiring diagrams is the Systems Biology Graphical Notation (SBGN) project, an effort to develop standardized conventions for diagrammatic representation of biological knowledge (Le Novère *et al.* 2009). Although diagrams are clearly useful, they lack predictive power.

A more powerful aid for reasoning about a cell signaling system is a mathematical model (Kholodenko *et al.* 2010; Kreeger and Lauffenburger 2010; Xu and Huang 2010). Models of cell signaling systems are less commonly used and less accessible than network wiring diagrams, but modeling provides a means to cast available knowledge in a form that allows predictions to be made via computation. Once a model has been formulated, computational procedures can usually be applied to elucidate the logical consequences of the model. We can expect modeling to be useful for understanding cell signaling systems, especially given emerging capabilities in synthetic biology (Bashor *et al.* 2010; Grünberg and Serrano 2010; Lim 2010), which present opportunities for model-guided cell engineering. However, it is becoming clear that traditional modeling approaches, such as the formalism of ordinary differential equations (ODEs), will not be entirely adequate. Modeling of cell signaling systems presents unique challenges for which new approaches are needed. Two of these challenges are discussed below.

1.3 The challenge of many moving parts

To address many interesting questions, we need large models (Hlavacek 2009). The reason is twofold. First, we need models of entire cell signaling systems and indeed models that encompass sets of connected cell signaling systems and other regulatory systems (e.g. metabolic circuits) to understand how the molecular states of cells affect their fates and phenotypes. The molecular state of a cell can be perturbed and potentially rationally manipulated. For example, pharmacological inhibition of a receptor tyrosine kinase might be used to impact cell-cycle control and reduce proliferation. However, the effect of a perturbation must generally propagate through a network of regulatory systems to bring about a change in phenotype, and these systems are rife with interconnections and feedback loops (Tyson *et al.* 2003; Brandman and Meyer 2008). The effects of some perturbations can be intuited. Others can be predicted with greater accuracy on the basis of simple or coarse-grained models or on the basis of mechanistically detailed but circumscribed models (Tyson and Novák 2010). However, at some point, to obtain deep understanding and/or exquisite control, we will seek to develop models that are large in scope and heavy in detail. Such models may only promise minor improvements in understanding or controllability, but even minor improvements, given the potential benefits, could provide strong motivation to pursue ambitious modeling efforts. In the United States, the annual cost of treating cancer is estimated to now be greater than US $100 billion and rising (Mariotto *et al.* 2011). This cost could perhaps

be reduced and/or care could be improved with greater predictive capabilities with regard to the effects of perturbations on cell signaling.

The second reason we need large models is that a cell signaling system typically comprises a large number of proteins and other signaling molecules (e.g. lipids), as well as a larger number of molecular interactions. For example, documentation of a cell signaling system in NetPath (netpath.org) includes, on average, information about >80 different molecules (counts from 'Molecules Involved' fields) and >150 molecular interactions (counts from 'Physical Interactions' and 'Enzyme Catalysis' fields) (Kandasamy *et al.* 2010). This problem can be dramatically exacerbated when one attempts to translate biological knowledge of a cell signaling system into mathematical form because the number of formal elements in a conventional, or traditional, model (e.g. variables or equations) is typically much larger than the number of molecules or interactions considered in the model. For example, a recent model encompassing only about 30 proteins is composed of nearly 500 ODEs (Chen *et al.* 2009).

Models that are far more comprehensive than available models, such as models that couple receptor signaling to cell cycle control, can be developed, evaluated, and incrementally improved through an iterative process, including parameter estimation, experimental tests of model predictions, and model refinement. This iterative process is often discussed, but the process is difficult to implement. It will continue to be difficult to implement until certain types of resources and conventions are developed to aid modelers. In addition, methods are needed to make modeling more automatic, such as methods for data-driven model specification [for an example of inference of network structure in the context of cell signaling, see Ciaccio *et al.* (2010)], automated comparison of model predictions to known, formalized system properties (Batt *et al.* 2005; Calzone *et al.* 2006; Gong *et al.* 2010), and automated model-guided design of experiments (Bongard and Lipson 2007). The resources I have in mind would provide the type of information available at the Tyson Lab website (mpf.biol.vt.edu), where for example, one can find an impressively long list of budding yeast mutants and their cell cycle phenotypes. This collection of information, which would be time consuming to reproduce, serves to document the basis of the model of Chen *et al.* (2004) and also provides a set of known system properties, which can be used to validate any new model for the budding yeast cell cycle or to compare alternative models. There is a wealth of information in the primary literature about system properties, which are mostly qualitative in nature. This information could perhaps be better used by modelers to (tightly) constrain the values of parameters in models.

The information available about a cell signaling system is often so abundant that only a fraction of the available information can be held in memory at a time. In fact, it may be impossible to even enter into memory all the available information. Available information must be systematized (ideally in the form of reusable and extensible models), reporting of new knowledge must be made more formal, and protocols and standardized conventions for accomplishing these tasks are needed. Translating available knowledge into formal elements of a model specification or into benchmarks for validating a model is currently a rate-limiting step in the modeling process, and it requires an unusual combination of skills. (It also requires a rare temperament.) With the adoption of conventions for reporting observed system properties, this burden could be lifted from modelers. A standard language for specifying dynamical system properties could perhaps be developed on the basis of a temporal logic (Calzone *et al.* 2006; Gong *et al.* 2010). An important element of such a language will be a capacity to represent system perturbations, such as mutations. Progress on formal representation of perturbations has been made by Danos *et al.* (2009).

It should be noted that a common concern about the development of a large model is the difficulty of estimating the many parameters in such a model. Recently, Apgar *et al.* (2010) demonstrated, via an assumed ground truth model and synthetic data, that the parameters of a model can be identified via fitting to measurements from an unexpectedly small number of experiments if the experiments are well chosen and under the assumptions that model structure is correct and all variables of the model are observable in the form of time-series data. Despite these caveats, this study suggests cautious optimism. The parameters in large models may be more identifiable than conventional wisdom suggests, and new statistical methods could perhaps be

developed for better addressing the problem of parameter estimation in the the context of large models for cell signaling systems. One need in this area is the development of parameter estimation methods that respect thermodynamic constraints, which are easily enforced in the case of a small model but not so easily enforced in the case of a large model (Yang *et al.* 2006; Ederer and Gilles 2007, 2008; Vlad and Ross 2009; Jenkinson *et al.* 2010).

1.4   The challenge of parts with parts

The parts of cell signaling systems (e.g. proteins) themselves have parts (e.g. sites of post-translational modification). To faithfully represent the mechanisms of cell signaling, we need models that account for the site-specific details of molecular interactions (Hlavacek *et al.* 2006; Hlavacek and Faeder 2009; Mayer *et al.* 2009). A signaling protein is typically composed of multiple functional components or sites. Tracking information about these sites is critical for modeling the system-level dynamics of molecular interactions because protein interactions are context sensitive, i.e. they depend on site-specific details. For example, many protein–protein interactions are regulated by phosphorylation (Pawson and Kofler 2009). It is difficult to document and/or visualize the contextual dependencies of molecular interactions in a way that is both clear and precise. Online resources, such as NetPath (Kandasamy *et al.* 2010), Phospho.ELM (Dinkel *et al.* 2011) and Reactome (Matthews *et al.* 2009), provide (limited) information about the site-specific details of protein–protein interactions through combinations of illustrations, lists (e.g. a list of sites of phosphorylation), comments, and literature citations. It is also difficult to incorporate the site-specific details of molecular interactions into a conventional or traditional model (e.g. ODEs) (Hlavacek *et al.* 2003). Nevertheless, our mechanistic understanding of site-specific details is fairly advanced and such details can be monitored experimentally (Gstaiger and Aebersold 2009; Choudhary and Mann 2010; Kolch and Pitt 2010), which encourages a consolidation of this knowledge in the form of models.

As mentioned above, ODEs are traditionally used to model the kinetics of (bio)chemical reaction systems, and many models of cell signaling systems take the form of a system of coupled ODEs (Kholodenko *et al.* 1999; Schoeberl *et al.* 2002). Another commonly used modeling approach involves using Gillespie's method, or kinetic Monte Carlo (KMC), to execute a stochastic simulation of the kinetics of a given list of reactions (Gillespie 2007; Voter 2007). This approach is typically applied in preference to the ODE-based approach when one is concerned about stochastic fluctuations that arise from small population sizes (McAdams and Arkin 1999). Both approaches require a listing of the reactions that are possible and the chemical species that can be populated in a system. The approach of using ODEs to represent the kinetics of chemical reaction systems was first applied by Harcourt and Esson (1865), long before we understood how cells process information. Cell signaling systems have features that distinguish these systems from other reaction systems, such as many reaction systems of importance in the chemical processing industry. In a typical commercial reactor or test tube, there are large numbers of molecules that can populate a relatively small number of chemical species. In contrast, in a cell signaling system, because of the combinatorial potential of protein interactions, relatively small numbers of molecules populate a tiny fraction of a large, possibly astronomical or even super astronomical number of possible chemical species (Endy and Brent 2001; Bray 2003; Hlavacek *et al.* 2003, 2006). This feature of cell signaling systems has been called combinatorial complexity. Combinatorial complexity hinders application of traditional modeling approaches, because in these approaches, one must write an equation for each chemical species that could be populated in a system, or the equivalent.

To overcome the problem of combinatorial complexity, the rule-based modeling approach has been developed (Hlavacek *et al.* 2006). In this approach, graphs are used to model proteins, and graph-rewriting rules are used to model protein interactions. Rules can be specified using a formal language, such as the BioNet-Gen language (BNGL) (Faeder *et al.* 2009) or Kappa (Feret *et al.* 2009; Harmer *et al.* 2010), which is closely

related to BNGL. The two basic ideas underlying the rule-based modeling approach are to treat molecules as the building blocks of chemical species and to model the molecular interactions that generate chemical species as rules rather than to specify a list of reactions connecting the chemical species that can be populated, which is equivalent to specifying a set of ODEs under the assumption of mass action kinetics. A rule can be viewed as implicitly defining a set of individual reactions, which involve a common transformation. A rule includes a specification of the necessary and sufficient conditions that a chemical species must satisfy to undergo a transformation defined by the rule and a rate law for the transformation. If the conditions are minimal, the number of individual reactions implied by the rule is maximal. For example, if ligand–receptor binding is independent of receptor signaling, then a rule can be specified that captures ligand–receptor inter-action without considering the interaction of the receptor with cytosolic signaling proteins. However, there is no free lunch. The cost of the rule-based modeling approach is that the reactions implied by a rule are all assumed to be characterized by the same rate law. Thus, the approach relies on a coarse-grained description of reaction kinetics, because each individual reaction in a system may truly be characterized by a unique rate law. Although it is important to keep this limitation of rule-based modeling in mind, the reliance on coarse-grained reaction kinetics is not onerous, as the granularity of a rule can be adjusted as needed. At the finest level of granularity, a rule implies only a single reaction, i.e. a rule is equivalent to a reaction in the traditional sense. Thus, the rule-based modeling approach can be viewed as a generalization of conventional modeling approaches. Rule-based modeling provides a practical solution to the problem of combinatorial complexity.

A number of methods for simulating rule-based models have been developed and these methods have been implemented in various software tools (Blinov *et al.* 2004; Lok and Brent 2005; Meier-Schellersheim *et al.* 2006; Moraru *et al.* 2008; Colvin *et al.* 2009, 2010; Lis *et al.* 2009; Mallavarapu *et al.* 2009; Andrews *et al.* 2010; Gruenert *et al.* 2010; Ollivier *et al.* 2010; Sneddon *et al.* 2010). A few recently developed tools are briefly discussed below to point to newly available modeling capabilities and to identify capability gaps. The tools discussed below all implement statistical methods. Once a rule-based model has been specified, it can be simulated in a number of different ways. For example, the set of ODEs for mass action kinetics corresponding to the list of reactions implied by a model can be derived from the rules of the model (Blinov *et al.* 2004; Faeder *et al.* 2005), or a smaller set of ODEs for the dynamics of collective variables can be derived from the rules (Borisov *et al.* 2008; Conzelmann and Gilles 2008; Koschorreck and Gilles 2008; Feret *et al.* 2009; Harmer *et al.* 2010). However, statistical methods are particularly well suited for simulation of a rule-based model, because tracking individual molecules in a stochastic simulation can be easier than tracking the population levels of chemical species in a deterministic simulation (Morton-Firth and Bray 1998; Danos *et al.* 2007; Yang *et al.* 2008; Colvin *et al.* 2009, 2010).

The stochastic simulation compiler (SSC) of Lis *et al.* (2009) implements a 'next subvolume' method for simulation of spatial rule-based models, in which subvolumes of a reaction volume are considered to be well mixed but transport between subvolumes is included to capture the effects of diffusion. Within a subvolume, KMC is used to simulate reaction kinetics, i.e. Gillespie's method is applied. SSC uses a constructive solid geometry approach, in which Boolean operators are used to combine primitive shapes/objects, to define surfaces and objects. A limitation of SSC is its reliance on the computationally expensive procedure of network generation, as SSC implements a 'generate-first' approach to simulation (Blinov *et al.* 2005). In this approach, a rule set is expanded into the corresponding reaction network (or equivalently, the list of reactions implied by the rules) before a simulation is performed. Network generation is expensive and impractical for many rule-based models. It is also unnecessary, as various 'network-free' approaches are available (Danos *et al.* 2007; Yang *et al.* 2008; Colvin *et al.* 2009, 2010). In these approaches, rules are used directly to advance a stochastic simulation. In fact, network-free simulation can be viewed as an extension of Gillespie's method in which a list of reactions is replaced with a list of rules. A number of software tools implement network-free simulation methods, including DYNSTOC (Colvin *et al.* 2009), RuleMonkey (Colvin *et al.* 2010) and NFsim (Sneddon *et al.* 2010), which are all compliant with BNGL. None of these tools have a spatial modeling

capability, i.e. the methods implemented in DYNSTOC, RuleMonkey and NFsim all rely on the assumption of a well-mixed reaction compartment. Network-free simulation tools compliant with Kappa with similar capabilities and limitations, such as KaSim, are available at kappalanguage.org/tools.

Another software tool for simulation of spatial rule-based models is Smoldyn (Andrews *et al.* 2010), which performs particle-based reaction-diffusion calculations. These types of calculations are well suited for dealing with complex geometrical constraints, such as a reconstructed 3D surface (Mazel *et al.* 2009). Smoldyn implements an 'on-the-fly' approach to simulation (Faeder *et al.* 2005; Lok and Brent 2005), which limits network generation by expanding a list of reactions as chemical species become populated during the course of a simulation. However, on-the-fly simulation still relies on network generation, and because of this, on-the-fly simulation is impractical for many rule-based models (Hlavacek *et al.* 2006; Yang *et al.* 2008). Interestingly, the network-free simulation approaches implemented in DYNSTOC, RuleMonkey, NFsim and KaSim are particle-based methods, and it should be straightforward to extend any of these methods to include molecule positions (spatial coordinates) and propagation of molecules in space, on or off a lattice. Powerful tools are available for specifying model geometry and visualizing results in 3D (Czech *et al.* 2009), and such tools could perhaps be coupled to rule-based simulation capabilities.

Finally, I wish to call attention to the simulation capability provided by the SRSim package (Gruenert *et al.* 2010), which couples BNGL (Faeder *et al.* 2009) to LAMMPS (Plimpton 1995), a package for performing molecular dynamics (MD) calculations. The MD simulation capability provided by SRSim allows molecules and their binding sites to be modeled as patchy particles and molecular interactions to be modeled in terms of force fields. This capability could be used to study the effects of orientation constraints on molecular interactions, molecular size (i.e. volume exclusion), and both translational and rotational diffusion. A wealth of structural information is available about the modular domains of signaling proteins, other molecules involved in signaling, and biomolecular complexes (Rose *et al.* 2011). Structures or models for protein signaling complexes could perhaps be built and used to identify potential steric clashes and other factors that potentially impact protein–protein interactions involved in signaling. Relating such structural information to function could perhaps be aided by system-level modeling of the dynamics of protein–protein interactions. A model for a signaling complex is static, whereas assembly of a signaling complex during the course of cell signaling is dynamic. Tools such as SRSim could provide a bridge for connecting structural data available about signaling proteins to system-level dynamics of protein–protein interactions, i.e. to protein function. The study of Monine *et al.* (2010) provides another example, but no general-purpose software, of how structural constraints on molecular interactions can be considered within the framework of rule-based modeling.

1.5   Closing remarks

Interest in modeling of cellular regulatory systems has increased dramatically in recent years, especially interest in stochastic modeling. In 2010, the classic paper of Gillespie (1977) was cited over 300 times, which is more than the number of times the paper was cited in the 20 years before Gillespie's method was applied by McAdams and Arkin (1997) to study gene regulation. Although modeling of cellular regulatory systems is more commonplace now than ever, the models being considered are usually small models, which also tend to incorporate little of the known mechanistic details (e.g. the site-specific details of protein–protein interactions). Much can be learned from small models, but cellular regulatory systems are large, and there are interesting questions that can be addressed with large and mechanistically detailed models. I am reminded of the saying attributed to Einstein, 'everything should be as simple as it can be, but not simpler.'

It will not be easy to develop useful large models of cell signaling systems for a variety of reasons. One reason is that there is no biological expert available to help the ambitious modeler, who may be a physicist or mathematician by training, with formalization of the known biology. I could be wrong, but I do not

believe there is a person alive today who has what can be considered a complete command of the known facts about any well-studied cell signaling system. One could be discouraged by this situation or motivated by it, as the systematization of knowledge provided by a model could help us make better use of available knowledge. Another difficulty for the ambitious modeler is the lack of support that can be expected from colleagues through the peer review process. Once an ODE-based model, for example, crosses some threshold of complexity, a modeler can expect little proofreading of equations, questioning of model assumptions, or helpful suggestions on how to improve parameter estimates. This situation points to a need to develop methods for better communicating the content of large models. Finally, one should keep in mind that models should be formulated to address specific questions, not for the sake of modeling. Finding the right questions to ask, the ones that can be answered with the help of large models, is the crux of the matter. Whether, or rather when, these questions can be found will determine when large modeling efforts become commonplace in the future.

References

Aldridge BB, Burke JM, Lauffenburger DA and Sorger PK 2006 Physicochemical modelling of cell signalling pathways. *Nat. Cell Biol.* **8**, 1195–1203.

Alon U 2006 *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall/CRC, Boca Raton, FL.

Andrews SS, Addy NJ, Brent R and Arkin AP 2010 Detailed simulations of cell biology with Smoldyn 2.1. *PLoS Comput. Biol.* **6**, e1000705.

Apgar JF, Witmer DK, White FM and Tidor B 2010 Sloppy models, parameter uncertainty, and the role of experimental design. *Mol. Biosyst.* **6**, 1890–1900.

Bashor CJ, Horwitz AA, Peisajovich SG and Lim WA 2010 Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu. Rev. Biophys.* **39**, 515–537.

Batt G, Ropers D, de Jong H, Geiselmann J, Mateescu R, Page M and Schneider D 2005 Validation of qualitative models of genetic regulatory networks by model checking: analysis of the nutritional stress response in *Escherichia coli*. *Bioinformatics* **21**, i19–i28.

Blinov ML, Faeder JR, Goldstein B and Hlavacek WS 2004 BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* **20**, 3289–3291.

Blinov ML, Faeder JR, Yang J, Goldstein B and Hlavacek WS 2005 'On-the-fly' or 'generate-first' modeling? *Nat. Biotechnol.* **23**, 1344–1345.

Bongard J and Lipson H 2007 Automated reverse engineering of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **104**, 9943–9948.

Borisov NM, Chistopolsky AS, Faeder JR and Kholodenko BN 2008 Domain-oriented reduction of rule-based network models. *IET Syst. Biol.* **2**, 342–351.

Bradshaw JM 2010 The Src, Syk, and Tec family kinases: distinct types of molecular switches. *Cell. Signal.* **22**, 1175–1184.

Brandman O and Meyer T 2008 Feedback loops shape cellular signals in space and time. *Science* **322**, 390–395.

Bray D 2003 Molecular prodigality. *Science* **299**, 1189–1190.

Calzone L, Fages F and Soliman S 2006 BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics* **22**, 1805–1807.

Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B and Tyson JJ 2004 Integrative analysis of cell cycle control in budding yeast. *Mol. Biol. Cell* **15**, 3841–3862.

Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA and Sorger PK 2009 Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol. Syst. Biol.* **5**, 239.

Choudhary C and Mann M 2010 Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.* **11**, 427–439.

Ciaccio MF, Wagner JP, Chuu CP, Lauffenburger DA, and Jones RB 2010 Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nat. Methods* **7**, 148–155.

Conzelmann H and Gilles ED 2008 Dynamic pathway modeling of signal transduction networks: a domain-oriented approach. *Methods Mol. Biol.* **484**, 559–578.

Colvin J, Monine MI, Faeder JR, Hlavacek WS, Von Hoff DD and Posner RG 2009 Simulation of large-scale rule-based models. *Bioinformatics* **25**, 910–917.

Colvin J, Monine MI, Gutenkunst RN, Hlavacek WS, Von Hoff DD and Posner RG 2010 RuleMonkey: software for stochastic simulation of rule-based models. *BMC Bioinformatics* **11**, 404.

Czech J, Dittrich M and Stiles JR 2009 Rapid creation, Monte Carlo simulation, and visualization of realistic 3D cell models. *Methods Mol. Biol.* **500**, 237–287.

Danos V, Feret J, Fontana W and Krivine J 2007 Scalable simulation of cellular signaling networks. *Lect. Notes Comput. Sci.* **4807**, 139–157.

Danos V, Feret J, Fontana W, Harmer R and Krivine J 2009 Rule-based modelling and model perturbation. *Lect. Notes Comput. Sci.* **5750**, 116–137.

Deininger M, Buchdunger E and Druker BJ 2005 The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood* **105**, 2640–2653.

Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ and Diella F 2011 Phopho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.* **39**, D261–D267.

Ederer M and Gilles ED 2007 Thermodynamically feasible kinetic models of reaction networks. *Biophys. J.* **92**, 1846–1857.

Ederer M and Gilles ED 2008 Thermodynamic constraints in kinetic modeling: thermodynamic-kinetic modeling in comparison to other approaches. *Eng. Life Sci.* **8**, 467–476.

Edwards JS, Faeder JR, Hlavacek WS, Jiang Y, Nemenman I and Wall ME 2007 q-bio 2007: a watershed moment in modern biology. *Mol. Syst. Biol.* **3**, 148.

Endy D and Brent R 2001 Modelling cellular behaviour. *Nature* **409**, 391–395.

Faeder JR, Blinov ML, Goldstein B and Hlavacek WS 2005 Rule-based modeling of biochemical networks. *Complexity* **10**, 22–41.

Faeder JR, Blinov ML and Hlavacek WS 2009 Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol. Biol.* **500**, 113–167.

Feret J, Danos V, Krivine J, Harmer R and Fontana W 2009 Internal coarse-graining of molecular systems. *Proc. Natl. Acad. Sci. USA* **106**, 6453–6458.

Gillespie DT 1977 Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361.

Gillespie DT 2007 Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55.

Gonzalez-Angulo AM, Hennessy BT and Mill GB 2010 Future of personalized medicine in oncology: a systems biology approach. *J. Clin. Oncol.* **28**, 2777–2783.

Gong H, Zuliani P, Komuravelli A, Faeder JR and Clarke EM 2010 Analysis and verification of the HMGB1 signaling pathway. *BMC Bioinformatics* **11**, S10.

Gould CM, Diella F, Via A, Puntervoll P, Gemünd C, Chabanis-Davidson S, Michael S, Sayadi A, Bryne JC, Chica C, Seiler M, Davey NE, Haslam N, Weatheritt RJ, Budd A, Hughes T, Pas J, Rychlewski L, Travé G, Aasland R, Helmer-Citterich M, Linding R and Gibson TJ 2010 ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res.* **38**, D167–D180.

Gruenert G, Ibrahim B, Lenser T, Lohel M, Hinze T and Dittrich P 2010 Rule-based spatial modeling with diffusing, geometrically constrained molecules. *BMC Bioinformatics* **11**, 307.

Günberg R and Serrano L 2010 Strategies for protein synthetic biology. *Nucleic Acids Res.* **38**, 2663–2675.

Gstaiger M and Aebersold R 2009 Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat. Rev. Genet.* **10**, 617–627.

Hanahan D and Weinberg RA 2000 The hallmarks of cancer. *Cell* **100**, 57–70.

Harcourt AV and Esson W 1865 On the laws of connexion between the conditions of a chemical change and its amount. *Proc. R. Soc.* **14**, 470–474.

Harmer R, Danos V, Feret J, Krivine J and Fontana W 2010 Intrinsic information carriers in combinatorial dynamical systems. *Chaos* **20**, 037108.

Hilbert D 1902 Mathematical problems. *Bull. Am. Math. Soc.* **8**, 437–479.

Hlavacek WS 2009 How to deal with large models? *Mol. Syst. Biol.* **5**, 240.

Hlavacek WS and Faeder JR 2009 The complexity of cell signaling and the need for a new mechanics. *Sci. Signal.* **2**, pe46.

Hlavacek WS, Faeder JR, Blinov ML, Perelson AS and Goldstein B 2003 The complexity of complexes in signal transduction. *Biotechnol. Bioeng.* **84**, 783–794.

Hlavacek WS, Faeder JR, Blinov ML, Posner RG, Hucka M and Fontana W 2006 Rules for modeling signal-transduction systems. *Sci. STKE* **2006**, re6.

Hunter T 2000 Signaling—2000 and beyond. *Cell* **100**, 113–127.

Hunter T 2007 Treatment for chronic myelogenous leukemia: the long road to imatinib. *J. Clin. Invest.* **117**, 2036–2043.

Jenkinson G, Zhong X and Goutsias J 2010 Thermodynamically consistent Bayesian analysis of closed biochemical reaction systems. *BMC Bioinformatics* **11**, 547.

Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Sameer Kumar GS, Venugopal AK, Telikicherla D, Navarro JD, Mathivanan S, Pecquet C, Kanth Gollapudi S, Gopal Tattikota S, Mohan S, Padhukasahasram H, Subbannayya Y, Goel R, Jacob HKC, Zhong J, Sekhar R, Nanjappa V, Balakrishnan L, Subbaiah R, Ramachandra YL, Abdul Rahiman B, Keshava Prasad TS, Lin JX, Houtman JCD, Desiderio S, Renauld JC, Constantinescu SN, Ohara O, Hirano T, Kubo M, Singh S, Khatri P, Draghici S, Bader GD, Sander C, Leonard WJ and Pandey A 2010 NetPath: a public resource of curated signal transduction pathways. *Genome Biol.* **11**, R3.

Keasling JD 2010 Manufacturing molecules through metabolic engineering. *Science* **330**, 1355–1358.

Khalil AS and Collins JJ 2010 Synthetic biology: applications come of age. *Nat. Rev. Genet.* **11**, 367–379.

Kholodenko BN, Demin OV, Moehren G and Hoek JB 1999 Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.* **274**, 30169–30181.

Kholodenko BN 2006 Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* **7**, 165–176.

Kholodenko BN, Hancock JF and Kolch W 2010 Signalling ballet in space and time. *Nat. Rev. Mol. Cell Biol.* **11**, 414–426.

Kitano H (ed.) 2001 *Foundations of Systems Biology*. MIT Press, Cambridge, MA.

Kitano H 2002 Systems biology: a brief overview. *Science* **295**, 1662–1664.

Kitano H, Funahashi A, Matsuoka Y and Oda K 2005 Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* **23**, 961–966.

Klipp E, Wade RC and Kummer U 2010 Biochemical network-based drug-target prediction. *Curr. Opin. Biotechnol.* **21**, 511–516.

Kohn KW 1999 Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell* **10**, 2703–2734.

Kolch W and Pitt A 2010 Functional proteomics to dissect tyrosine kinase signalling pathways in cancer. *Nat. Rev. Cancer* **10**, 618–629.

Koschorreck M and Gilles ED 2008 ALC: automated reduction of rule-based models. *BMC Syst. Biol.* **2**, 91.

Kreeger PK and Lauffenburger DA 2010 Cancer systems biology: a network modeling perspective. *Carcinogenesis* **31**, 2–8.

Lander AD 2010 The edges of understanding. *BMC Biol.* **8**, 40.

Lemmon MA and Schlessinger J 2010 Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134.

Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villeger A, Boyd SE, Calzone L, Courtot M, Dogrusoz U, Freeman TC, Funahashi A, Ghosh S, Jouraku A, Kim S, Kolpakov F, Luna A, Sahle S, Schmidt E, Watterson S, Wu G, Goryanin I, Kell DB, Sander C, Sauro H, Snoep JL, Kohn K and Kitano H 2009 The systems biology graphical notation. *Nat. Biotechnol.* **27**, 735–741.

Lim WA 2010 Designing customized cell signalling circuits. *Nat. Rev. Mol. Cell Biol.* **11**, 393–403.

Lis M, Artyomov MN, Devadas S and Chakraborty AK 2009 Efficient stochastic simulation of reaction-diffusion processes via direct compilation. *Bioinformatics* **25**, 2289–2291.

Lok L and Brent R 2005 Automatic generation of cellular reaction networks with Moleculizer 1.0. *Nat. Biotechnol.* **23**, 131–136.

Mallavarapu A, Thomson M, Ullian B and Gunawardena J 2009 Programming with models: modularity and abstraction provide powerful capabilities for systems biology. *J. R. Soc. Interface* **6**, 257–270.

Mariotto AB, Robin Yabroff K, Shao Y, Feuer EJ and Brown ML 2011 Projections of the cost of cancer care in the United States: 2010–2020. *J. Natl. Cancer Inst.* **103**, 117–128.

Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L and D'Eustachio P 2009 Reactome knowledge base of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–D622.

Mayer BJ, Blinov ML and Loew LM 2009 Molecular machines or pleiomorphic ensembles: signaling complexes revisited. *J. Biol.* **8**, 81.

Mazel T, Raymond R, Raymond-Stintz M, Jett S and Wilson BS 2009 Stochastic modeling of calcium in 3D geometry. *Biophys. J.* **96**, 1691–1706.

McAdams HH and Arkin A 1997 Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* **94**, 814–819.

McAdams HH and Arkin A 1999 It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15**, 65–69.

Meier-Schellersheim M, Xu K, Angermann B, Kunkel EJ, Jin T and Germain RN 2006 Key role of local regulation in chemosensing revealed by a new molecular interaction-based modeling method. *PLoS Comput. Biol.* **2**, e82.

Monine MI, Posner RG, Savage PB, Faeder JR and Hlavacek WS 2010 Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. *Biophys. J.* **96**, 48–56.

Moraru II, Schaff JC, Slepchenko BM, Blinov ML, Morgan F, Lakshminarayana A, Gao F, Li Y and Loew LM 2008 Virtual Cell modelling and simulation software environment. *IET Syst. Biol.* **2**, 352–362.

Morton-Firth CJ and Bray D 1998 Predicting temporal fluctuations in an intracellular signalling pathway. *J. Theor. Biol.* **192**, 117–128.

Mukherji S and van Oudenaarden A 2009 Synthetic biology: understanding biological design from synthetic circuits. *Nat. Rev. Genet.* **10**, 859–871.

Novák B and Tyson JJ 2008 Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* **9**, 981–991.

Nurse P 2008 Life, logic and information. *Nature* **454**, 424–426.

Ollivier JF, Shahrezaei V and Swain PS 2010 Scalable rule-based modelling of allosteric proteins and biochemical networks. *PLoS Comput. Biol.* **6**, e1000975.

Pawson T and Kofler M 2009 Kinome signaling through regulated protein–protein interactions in normal and cancer cells. *Curr. Opin. Cell Biol.* **21**, 147–153.

Pawson T and Nash P 2003 Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445–452.

Plimpton SJ 1995 Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19.

Ptashne M and Gann A 2001 *Genes & Signals*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Rafelski SM and Marshall WF 2008 Building the cell: design principles of cellular architecture. *Nat. Rev. Mol. Cell Biol.* **9**, 593–602.

Rose PW, Beren B, Bi C, Bluhm WF, Dimitropoulos D, Goodsell DS, Prlić A, Quesada M, Quinn GB, Westbrook JD, Young J, Yukich B, Zardecki C, Berman HM and Bourne PE 2011 The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.* **39**, D392–D401.

Schoeberl B, Eichler-Jonsson C, Gilles ED and Müller G 2002 Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.* **20**, 370–375.

Scott JD and Pawson T 2009 Cell signaling in space and time: where proteins come together and when they're apart. *Science* **326**, 1220–1224.

Sneddon MW, Faeder JR and Emonet T 2010 Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat. Methods* doi:10.1038/nmeth.1546.

Tyson JJ and Novák B 2010 Functional motifs in biochemical reaction networks. *Annu. Rev. Phys. Chem.* **61**, 219–240.

Tyson JJ, Chen KC and Novak B 2003 Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231.

Vlad MO and Ross J 2009 Thermodynamically based constraints for rate coefficients of large biochemical networks. *WIREs Syst. Biol. Med.* **1**, 348–358.

Voter AF 2007 Introduction to the kinetic Monte Carlo method. In *Radiation Effects in Solids* (eds Sickafus KE, Kotomin EA, and Uberuaga BP), pp. 1–23. Springer, Dordrecht.

Wall ME, Hlavacek WS and Savageau MA 2004 Design of gene circuits: lessons from bacteria. *Nat. Rev. Genet.* **5**, 34–42.

Walsh CT, Garneau-Tsodikova S and Gatto Jr GJ 2005 Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* **44**, 7342–7372.

Xu AM and Huang PH 2010 Receptor tyrosine kinase coactivation networks in cancer. *Cancer Res.* **70**, 3857–3860.

Yang J, Bruno WJ, Hlavacek WS and Pearson JE 2006 On imposing detailed balance in complex reaction mechanisms. *Biophys. J.* **91**, 1136–1141.

Yang J, Monine MI, Faeder JR and Hlavacek WS 2008 Kinetic Monte Carlo method for rule-based modeling of biochemical networks. *Phys. Rev. E* **78**, 031910.