

Model-based design approaches in drug discovery: A parallel to traditional engineering approaches

B. Schoeberl
U. B. Nielsen
R. Paxson

Model-based design (MBD) has been successfully applied in the automotive, chemical, and aerospace industries. Here we discuss the possible application of engineering-based MBD approaches to drug discovery. One of the biggest challenges in drug discovery is the high attrition rate of new drugs in development: Many promising candidates prove ineffective or toxic in animal or human testing. More often than not, these failures are the result of a poor understanding of the molecular mechanisms of the biological systems they target. Recent advances in biological systems modeling make MBD an attractive approach to improve drug development. We elaborate on the view that the pharmaceutical industry should be able to use MBD to design new drugs more effectively. There are significant differences between drug discovery and traditional engineering that lead to specific MBD requirements. We delineate those differences and introduce suggestions to overcome them.

Introduction

One of the promises of systems biology is that the systems-level understanding of biological pathways and processes will allow for “smarter” drug development, as described by Aksenov et al. [1], Apic et al. [2], and Nielsen and Schoeberl [3]. The optimal therapeutic approach and drug design are selected on the basis of a detailed understanding of the drug mechanism of action, which results in targeted therapeutics with higher efficacy and fewer side effects. To achieve this goal, there is a need for model-based drug design that allows for the rapid *in silico* identification of the “optimal” target and identifies the best mechanism of action (MOA)—e.g., small molecule compared with antibody therapy, optimal inhibitor affinity, target ligand or receptor, possible induction of receptor internalization and degradation, and so forth.

In broad terms, successful model construction and rapid application by research teams requires systematic methods of building, annotating, analyzing, maintaining, and sharing these mathematical models. Further, successful models must bridge a virtual organizational gap by facilitating their use by experimentalists as well as by model creators.

Mathematical models of dynamic systems have long been used in the automobile, chemical, and other industries to support model-based component design and plant control. The day when the pharmaceutical industry can test a drug against a computational model with a high degree of confidence may arrive in the not-too-distant future. The computational approach is building momentum in the pharmaceutical industry and is already beginning to yield results. Mathematical models of relatively small signal transduction or metabolic networks

©Copyright 2006 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/06/\$5.00 © 2006 IBM

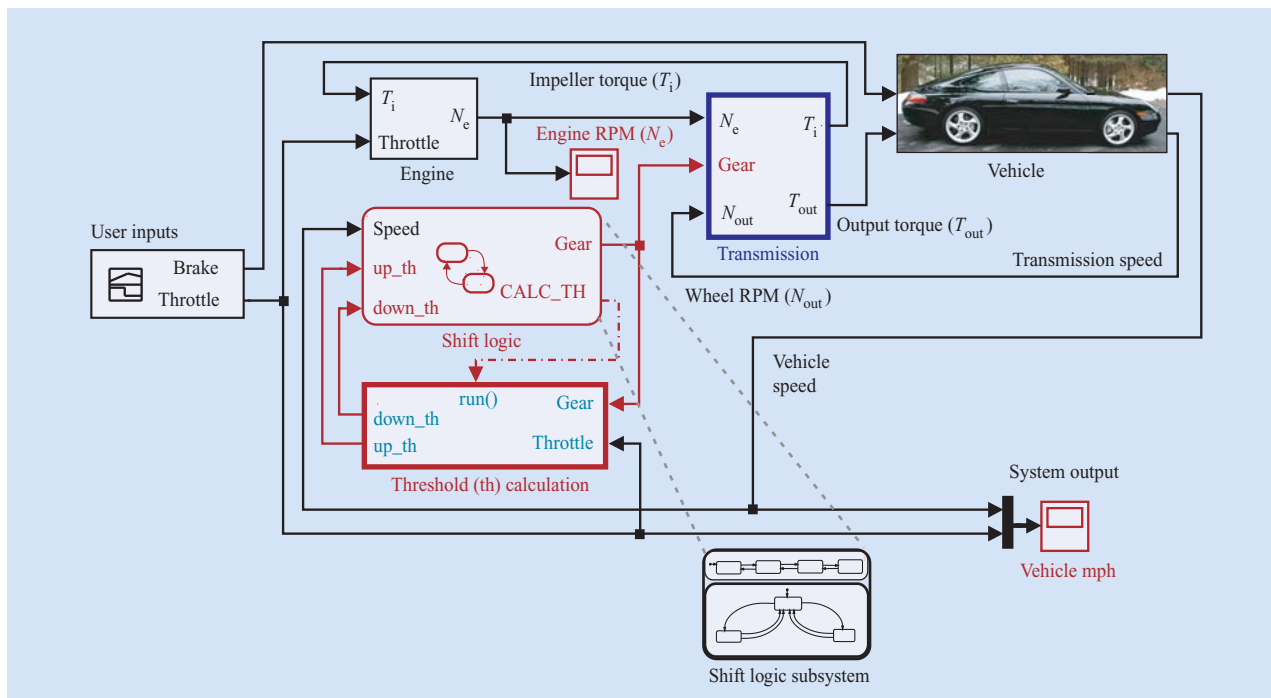


Figure 1

Model-based design scheme of an automatic transmission for an automobile. Black: Continuous sampling (differential equations); red: fastest discrete sample time in the model; blue: hybrid sample times (a submodel containing blocks running at multiple sample times).

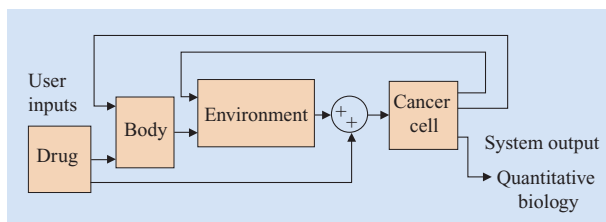


Figure 2

Model-based drug design scheme for an anticancer drug.

could be the first building blocks in the systems integration of many models into a more comprehensive, higher-fidelity model—possibly even a multicell, full-organ, or full-body model.

There are methodologies and tools that support model construction, parameterization, aggregation, validation, and analysis. The automotive, chemical, semiconductor, and aerospace industries have been leaders in the application of model-based design (MBD) tools and have experienced significant improvement in design and shortened product development cycles based on their use. However, there is a key fundamental difference between

modeling biological systems and modeling engineered systems: the former is a *reverse-engineering* task.

Here we examine the applicability of mathematical modeling methodologies and tools to the drug development process. Although the modeling approaches used in MBD have previously been applied in other industries, the challenges faced by drug companies using model-based design are more complex. One implication is that most of the kinetic parameters and species concentrations are unknown. More generally, the challenges include the extreme “stiffness” of these biochemical models, the high number of undetermined parameters, the difficulty of use for modelers and experimentalists, and the fact that model documentation and annotation based on prior knowledge from literature requires specialized software that combines the methodology and knowledge gained in other areas.

Model-based design and model-based drug design

In the early 1990s, industries such as automotive and aerospace began to use software tools that facilitated the construction and application of models. More than a decade later, a complete MBD methodology has become standard for the construction and application of models in various design and development activities.

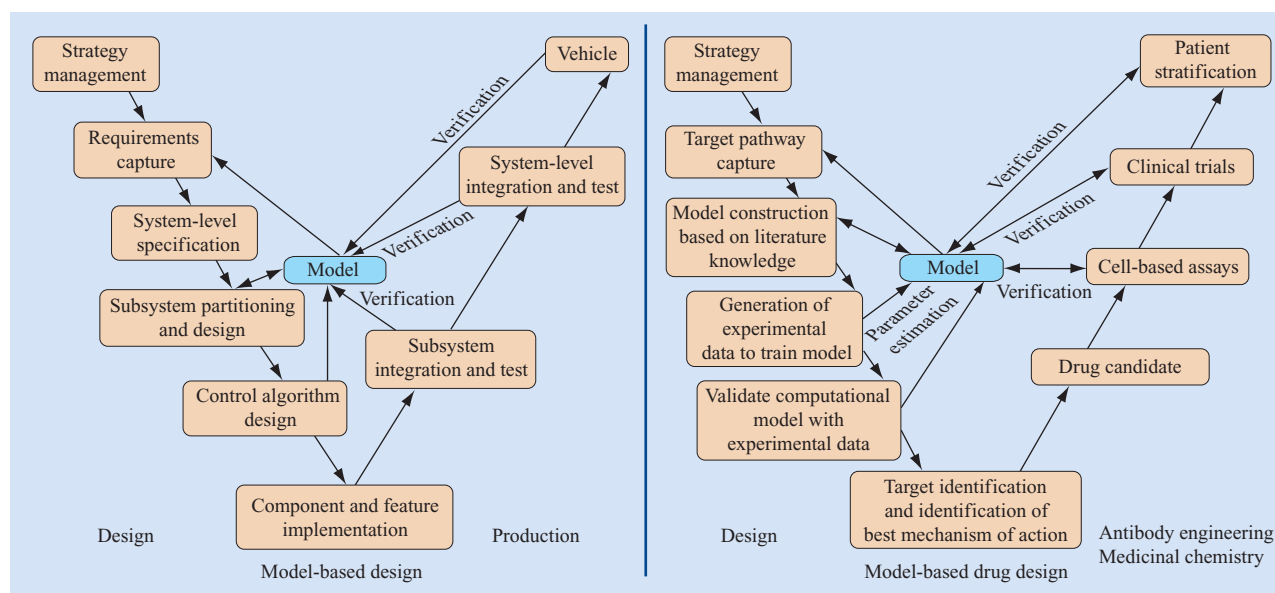


Figure 3

Comparison of model-based-design methodologies in engineering and drug discovery.

Figure 1 shows a model of an automobile transmission design using MBD. The model is composed of submodels that represent the *vehicle*, *transmission*, *engine*, and *controller*, which itself consists of the *shift logic* and *shift point* (depicted as threshold) calculation. On the basis of the driver's input (i.e., brake or accelerator), the model can be used to optimize system performance by shifting gears to maintain the operation of the engine at maximum fuel efficiency. Once the desired performance is obtained, the controller (shown in the figure as the *shift logic* and *threshold calculation* blocks) can be implemented automatically via code generation and tested, validated, and deployed on an embedded system.

In the case of model-based drug design (MBDD), the objective is to identify the protein that will be targeted by the novel drug and its best MOA. In contrast to the top-down approach typically used in engineering, a bottom-up strategy is favored in the early stages of the drug design process. The critical networks implicated in the disease of interest are initially identified and reverse-engineered (*pathway capture*). Well-known signaling or metabolic pathways can be reconstructed on the basis of literature knowledge, and network inference methods (e.g., Bayesian networks) can be applied to reverse-engineer protein interaction or gene networks from large datasets of protein or gene datasets, as elucidated by Kholodenko et al. [4], Sachs et al. [5], and Needham et al. [6]. The resulting model can then be used to identify drug targets and determine the optimal MOA. After an initial

validation with wet-lab experiments and model tuning, the submodel can be used in the assembly of a more encompassing model that accounts for pharmacokinetic and pharmacodynamic effects (PK/PD), as depicted in **Figure 2**. The final model, encompassing both signaling pathway models and PK/PD models, can later be applied in the clinic to identify responders (patient stratification). However, all-encompassing models for MBDD, such as the one shown in Figure 2, remain only a vision. Submodels such as PK/PD models and signaling pathway models are already being applied in the pharmaceutical industry [7, 8].

The MBD and MBDD design processes are shown in **Figure 3**. The first step in the construction of a model is to gather the model requirements. It is important to define the questions that the model should address and the information available to support the model construction. Throughout this activity, the model functions as a knowledge aggregator by accumulating the collective information known about the physical system and its decomposition into subunits. The model may be incomplete at this point, but through an iterative process between model simulations and experiments, the knowledge gaps can be filled during the design process. In Figure 3, the highly iterative process is depicted by double-ended arrows comparing the model with experiments in each step of the MBDD. This is in contrast to the traditional process of gathering requirements in written specifications and building

Table 1 Comparison of model-based design methodologies applied to engineering and to biology and drug discovery.

	<i>Engineering</i>	<i>Biology and drug discovery</i>
Modeling strategy	Forward-engineered and designed	Reverse-engineered and inferred from literature
Modularity and hierarchy	Modules are part of the design architecture and have clear input and output characteristics	• Modules are difficult to ascertain initially • Modules are discovered via model analysis
Model aggregation	Submodels are plug-and-play because of the clear application programming interface	Variability of inputs and outputs according to the comprehensiveness of the model makes model aggregation from submodels difficult
Level of fidelity	Level of fidelity is increased as needed to reach design goals	Key components initially require a high level of fidelity
Interdisciplinary level	Modules are typically built by specific engineering teams; control engineers work on control logic, while mechanical engineers work on vehicle model	Interdisciplinary teams are required throughout the design process because of the need for high fidelity
Modeling goal	Robust controllers and code generation for verification with embedded systems	Plant model that is able to determine the best MOA and best drug target
Model parameters	Most parameters are directly measurable	Many of the model parameters are typically unknown and estimated with experimental data
Analysis	Nonlinear models are used, but linear analysis methods are established and widely applied	Nonlinear systems analysis tools, bifurcation, sensitivities, global optimization
Graphical representation	Well-established notation	No standardized notation yet

simulations by writing software code. While written specifications serve as static pieces of information for the various stages of development, the model aggregates information and requirements across teams.

It has been well established that the MBD methodology has been successful in facilitating large-system design and in reducing development cycles. It has proven useful in concentrating system knowledge on a single model and in facilitating model aggregation through the reuse of component models. A comparison of activities in the engineering and the drug development processes, as summarized in **Table 1**, outlines ways in which the methodology can be applied and highlights areas in which new tools are needed to increase the applicability of MBD to problems and challenges in drug discovery.

Model decomposition—motifs and modularity

A first step in system architecture is to specify its hierarchical decomposition into submodels and to specify their interconnectivity. The modular approach allows large and complex systems to be modeled in a more manageable and transparent fashion. In addition, each submodel can be individually built, validated, integrated, and tested in the full system.

Using an automobile transmission as an example, the initial model can be built with rough (low-fidelity) submodels for each component, e.g., the engine, transmission, and controller, all of which have well-defined input and output characteristics. During the

development process, specialized teams improve and refine the submodels individually, increasing the fidelity of the overall model. Each submodel encapsulates the current state of the art and is updated with new knowledge or technology as it becomes available.

Because a submodel encapsulates a distinct part, it can be reused. For an automatic transmission, the controller unit can be used in the development process of another vehicle. The clear definition of the input and output characteristics of each component is a prerequisite for the “plug-and-play” approach to model building.

Because biological systems are characterized by large and complex networks, the modular approach appears to be appropriate. As depicted in Figure 2, this approach appears to be applicable on the macro level (tissue–organ–body) because the model components can be identified by defined inputs and outputs [9–11]. However, on the micro level, where the models account for protein or gene networks, it becomes difficult to break the system down into modules. This is the main difference between applying MBD to engineering and to biology.

To meet drug design goals, a high-fidelity model of a cancer cell and the tumor environment is needed, but little is known about these systems. Therefore, it is difficult to determine where low-fidelity models would be adequate. To learn how to adequately approximate the system behavior, a high-fidelity model is initially required. Currently, high-fidelity models are obtained by reverse-

engineering protein or gene networks. However, depending on the design goal, very detailed submodels might still be needed to appropriately model the active mechanism of a particular drug.

In contrast to the engineering case, in which specialized teams work on submodels, in MBDD interdisciplinary teams are needed throughout the design methodology because of the difficulty of defining modules in protein or gene networks and achieving the required high-level fidelity obtained by gathering experimental data and continuously refining the model.

Unlike engineering models, biological submodels do not have clear input and output characteristics; therefore, model reuse and aggregation becomes more complex. Rather than plug-and-play, the relationships among submodels must be manually edited after several models have been fused. For example, the mitogen-activated kinase (MAPK) cascade submodel is ubiquitous in many signaling pathways, but the nature of the interactions between it and other submodels—e.g., the phosphoinositide 3-kinases (PI3Ks) cascade—differs from pathway to pathway in connectivity and strength. Instead of plug-and-play, a strategy that uses a rule-based scheme to create the crosstalk interactions between the different submodels appears to be more adequate.

An additional complication of model aggregation arises from the multiple occurrences of network motifs (e.g., the same kinase or phosphatase) in different signaling pathways. Thus, the combination of several signaling pathway models is complicated by uncertainty as to whether these network motifs use the same protein pools or are spatially separated by, for example, scaffolding proteins. This is reflected in the choice of initial species concentrations (single or multiple pools) and the value of the kinetic parameters. Experimental data addressing both issues is needed in order to parameterize the model appropriately and to obtain high model fidelity.

As opposed to engineering models, in which most parameters can be directly measured, kinetic parameters in biological systems must often be inferred from large and “noisy” datasets. This step in model refinement is done with parameter estimation; however, because of the large number of unknown parameters and the uncertainty in their values, global optimization techniques are needed [12]—more so than in engineering.

Model analysis: Drug target identification

Existing techniques such as identification, gain quantification, sensitivity analysis, and optimal control are well-developed areas of control theory commonly applied in engineering. For engineering systems, even though the models are nonlinear, the operating regions

are generally known. This allows the system to be linearized and analyzed using linear methods. Because of the lack of well-characterized operating points in biological systems, nonlinear analysis methods must be used, and tools must be developed that can conveniently facilitate the exploration of model behavior across parameter space.

Sensitivity analysis is used as a bootstrap method during model tuning to determine which parameters should be estimated from experimental data. Given that these represent the most sensitive nodes of the network, the same analysis helps to identify the best drug targets in the system [13, 14]. Because of the broad range of its parameter values, the nonlinear system can display a variety of stable and unstable behaviors. Often bifurcation analysis is used to reveal regions of stability or instability in the network [15]. This is helpful because experiments can be specifically designed to observe interesting nonlinear phenomena predicted by the model. For instance, Hoffmann et al. [16] have developed a computational model that accounts for the temporal control of NF- κ B activation by the coordinated degradation and synthesis of I- κ B proteins. The activity of NF- κ B is controlled by three different isoforms of I κ B: α , β , and ϵ . The model predicted that I κ B α is responsible for the fast turn-off, and I κ B β and I κ B ϵ function to reduce the system oscillatory potential and stabilize the NF- κ B response. These model predictions were experimentally confirmed, and Hoffmann et al. showed that gene expression specificity is achieved by the signal characteristics (i.e., persistent vs. transient NF- κ B signals). This example shows how bifurcation analysis serves both to validate the model and to facilitate learning about unexpected system behavior.

However, it is foreseeable not only that engineering principles can be applied to biology, but that entirely new theoretical problems arise when questions in the field of systems biology are addressed with traditional control theory [17]. It may be that in the long run, MBD as it is applied to biological systems may advance the use of MBD in other industries, and vice versa.

Graphical representation: Pathway representation

The graphical language used to describe the automatic transmission model example (Figure 1) consists of signals flowing through the lines in the diagram between the various system components, represented by blocks. The mathematical operations represented by the blocks and the data carried by the signals together represent an unambiguous mathematical model of the system (such models typically result in large systems of nonlinear ordinary differential equations). In addition, the models are typically hierarchical. For instance, in Figure 1 the

block representing the state machine for the shifting logic itself encapsulates lines and blocks that perform its function. Thus, the modeler can quickly build models by graphically adding blocks that contain underlying mathematical models. This speeds up the model-building process and makes it less error-prone. The semantics of the graphical language are well defined and widely used in control engineering. Other engineering domains (e.g., mechanical, electrical, and hydraulic) also have well-defined graphical notation that is useful in specifying models and communicating model architecture. An important aspect of these semantics is that they are capable of fully and unambiguously describing the mathematical model underlying the diagram.

Biologists use less standardized symbols to describe biological pathways—i.e., lines are used to represent interactions between species, and blocks are used to represent species and their concentration. Combinatorial complexity arising from protein complex formation is not accounted for in traditional graphical representations. Conventions are beginning to appear (e.g., those published by Kohn [18], Maimon and Browning [19], and Kitano et al. [20]) that attempt to uniquely determine the interaction type and therefore possibly the underlying mathematics. However, these notations are still far from the standardized representation used in engineering disciplines. Current graphical representations of pathways help interdisciplinary teams to communicate but do not provide an executable model that can be used for system simulation.

The granularity and size of protein and gene networks, in addition to the fact that they are reverse-engineered from large datasets, call for automatic graphical layout. Similarly, such models are often composed of hundreds to thousands of species, making manual graphical construction and representation of the model almost impossible.

Summary

About ten years ago, the first mathematical models of signal transduction were published. These began with subsystems in order to obtain a quantitative understanding of processes such as receptor trafficking [21], the signal transfer behavior of the MAPK cascade [22], or larger models of signaling pathways [13, 23, 24]. The success of model construction and parameterization was demonstrated by their experimental validation and created interest in the drug discovery industry.

With the sequencing of the human genome and an increased knowledge of the individual proteins that make up cellular pathways, the pharmaceutical industry has focused increasingly on individual molecules as targets in their quest for targeted therapeutics. However, this approach has not yet resulted in the delivery of the first

systems biology drug, nor has it increased the number of drug approvals or reduced the toxicity of drug candidates. It is becoming evident that putative protein targets for drugs must be understood in the greater biological context in which they are active—molecular interactions, biochemical pathways, cellular compartments, tissues, and organs. To do this effectively, scientists must embrace the notion that biology is increasingly an engineering-based science and should take advantage of the proven tools developed in other engineering-based disciplines. The application of computational and engineering tools to life sciences research and drug discovery is beginning to provide a detailed quantitative understanding of the interactions occurring within biological systems; it promises to lead to the development of novel, targeted drugs offering improved efficacy and safety.

Acknowledgment

We thank Tad Stewart at Merrimack Pharmaceuticals for his critique of this paper.

References

1. S. V. Aksenov, B. Church, A. Dhiman, A. Georgieva, R. Sarangapani, G. Helmlinger, and I. G. Khalil, "An Integrated Approach for Inference and Mechanistic Modeling for Advancing Drug Development," *FEBS Lett.* **579**, No. 8, 1878–1883 (2005).
2. G. Apic, T. Ignjatovic, S. Boyer, and R. B. Russell, "Illuminating Drug Discovery with Biological Pathways," *FEBS Lett.* **579**, No. 8, 1872–1877 (2005).
3. U. B. Nielsen and B. Schoeberl, "Using Computational Modeling to Drive the Development of Targeted Therapeutics," *IDrugs* **8**, No. 10, 822–826 (2005).
4. B. N. Kholodenko, A. Kiyatkin, F. J. Bruggeman, E. Sontag, H. V. Westerhoff, and J. B. Hoek, "Untangling the Wires: A Strategy to Trace Functional Interactions in Signaling and Gene Networks," *Proc. Natl. Acad. Sci. USA* **99**, No. 20, 12841–12846 (2002).
5. K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, and G. P. Nolan, "Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data," *Science* **308**, No. 5721, 523–529 (2005).
6. C. J. Needham, J. R. Bradford, A. J. Bulpitt, and D. R. Westhead, "Inference in Bayesian Networks," *Nature Biotechnol.* **24**, No. 1, 51–53 (2006).
7. A. Hooker and P. Vicini, "Simultaneous Population Optimal Design for Pharmacokinetic–Pharmacodynamic Experiments," *AAPS J.* **7**, No. 4, E759–E785 (2005).
8. J. M. Gallo, P. Vicini, A. Orlansky, S. Li, F. Zhou, J. Ma, S. Pulfer, M. A. Bookman, and P. Guo, "Pharmacokinetic Model-Predicted Anticancer Drug Concentrations in Human Tumors," *Clin. Cancer Res.* **10**, No. 23, 8048–8058 (2004).
9. J. A. Papin, J. L. Reed, and B. O. Palsson, "Hierarchical Thinking in Network Biology: The Unbiased Modularization of Biochemical Networks," *Trends Biochem. Sci.* **29**, No. 12, 641–647 (2004).
10. A. Ma'ayan, R. D. Blitzer, and R. Iyengar, "Toward Predictive Models of Mammalian Cells," *Annu. Rev. Biophys. Biomolec. Struct.* **34**, 319–349 (2005).
11. O. Wolkenhauer, S. N. Sreenath, P. Wellstead, M. Ullah, and K.-H. Cho, "A Systems- and Signal-Oriented Approach to Intracellular Dynamics," *Biochem. Soc. Trans.* **33**, Pt. 3, 507–515 (2005).

12. C. G. Moles, P. Mendes, and J. R. Banga, "Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods," *Genome Res.* **13**, No. 11, 2467–2474 (2003).
13. M. Bentele, I. Lavrik, M. Ulrich, S. Stosser, D. W. Heermann, H. Kalthoff, P. H. Krammer, and R. Eils, "Mathematical Modeling Reveals Threshold Mechanism in CD95-Induced Apoptosis," *J. Cell Biol.* **166**, No. 6, 839–851 (2004).
14. J. J. Hornberg, B. Binder, F. J. Bruggeman, B. Schoeberl, R. Heinrich, and H. V. Westerhoff, "Control of MAPK Signalling: From Complexity to What Really Matters," *Oncogene* **24**, No. 36, 5533–5542 (2005).
15. H. Song, P. Smolen, E. Av-Ron, D. A. Baxter, and J. H. Byrne, "Bifurcation and Singularity Analysis of a Molecular Network for the Induction of Long-Term Memory," *Biophys. J.* (2006); see <http://www.biophysj.org/cgi/content/abstract/90/7/2309>.
16. A. Hoffmann, A. Levchenko, M. L. Scott, and D. Baltimore, "The I κ B–NF- κ B Signaling Module: Temporal Control and Selective Gene Activation," *Science* **298**, No. 5596, 1241–1245 (2002).
17. E. D. Sontag, "Some New Directions in Control Theory Inspired by Systems Biology," *Syst. Biol.* **1**, No. 1, 9–18 (2004).
18. K. W. Kohn, "Molecular Interaction Maps as Information Organizers and Simulation Guides," *Chaos* **11**, No. 1, 84–97 (2001).
19. R. Maimon and S. Browning, "Diagrammatic Notation and Computational Structure of Gene Networks," *Proceedings of the 2nd International Conference on Systems Biology*, 2001; see <http://www.gnsbiotech.com/papers/dac.pdf>.
20. H. Kitano, A. Funahashi, Y. Matsuoka, and K. Oda, "Using Process Diagrams for the Graphical Representation of Biological Networks," *Nature Biotechnol.* **23**, No. 8, 961–966 (2005).
21. C. Starbuck and D. A. Lauffenburger, "Mathematical Model for the Effects of Epidermal Growth Factor Receptor Trafficking Dynamics on Fibroblast Proliferation Responses," *Biotechnol. Progress* **8**, No. 2, 132–143 (1992).
22. C.-Y. F. Huang and J. E. Ferrell, Jr., "Ultrasensitivity in the Mitogen-Activated Protein Kinase Cascade," *Proc. Natl. Acad. Sci. USA* **93**, No. 19, 10078–10083 (1996).
23. B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek, "Quantification of Short Term Signaling by the Epidermal Growth Factor Receptor," *J. Biol. Chem.* **274**, No. 42, 30169–30181 (1999).
24. B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Muller, "Computational Modeling of the Dynamics of the MAP Kinase Cascade Activated by Surface and Internalized EGF Receptors," *Nature Biotechnol.* **20**, No. 4, 370–375 (2002).

Received November 7, 2005; accepted for publication January 13, 2006; Internet publication September 15, 2006

Birgit Schoeberl *Merrimack Pharmaceuticals, 101 Binney Street, Cambridge, Massachusetts 02142 (bschoeberl@merrimackpharma.com)*. Dr. Schoeberl is the director of Network Biology at Merrimack Pharmaceuticals. Together with Ulrik Nielsen, she helped formulate Merrimack's research and development platform, Network Biology. Dr. Schoeberl received an M.S. degree in chemical engineering from the University of Karlsruhe, Germany, and a Ph.D. degree from the Max-Planck-Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany.

Ulrik B. Nielsen *Merrimack Pharmaceuticals, 101 Binney Street, Cambridge, Massachusetts 02142 (unielsen@merrimackpharma.com)*. Dr. Nielsen is co-founder and vice president of research at Merrimack Pharmaceuticals. He directs all major functions and activities within the company's research and drug discovery programs. Dr. Nielsen received his Ph.D. degree in molecular biology from the University of Copenhagen, Denmark.

Ricardo Paxson *The MathWorks Inc., 3 Apple Hill Drive, Natick, Massachusetts 01760 (pax@mathworks.com)*. Mr. Paxson is head of the Systems Biology group at The MathWorks Inc. He manages a team of engineers developing tools for modeling and simulation of biological systems. Mr. Paxson received B.S. and M.S. degrees in aeronautical engineering from the Massachusetts Institute of Technology.