Less Is More in Modeling Large Genetic Networks

Stefan Bornholdt

Science **310**, 449 (2005);
DOI: 10.1126/science.1119959

SYSTEMS BIOLOGY

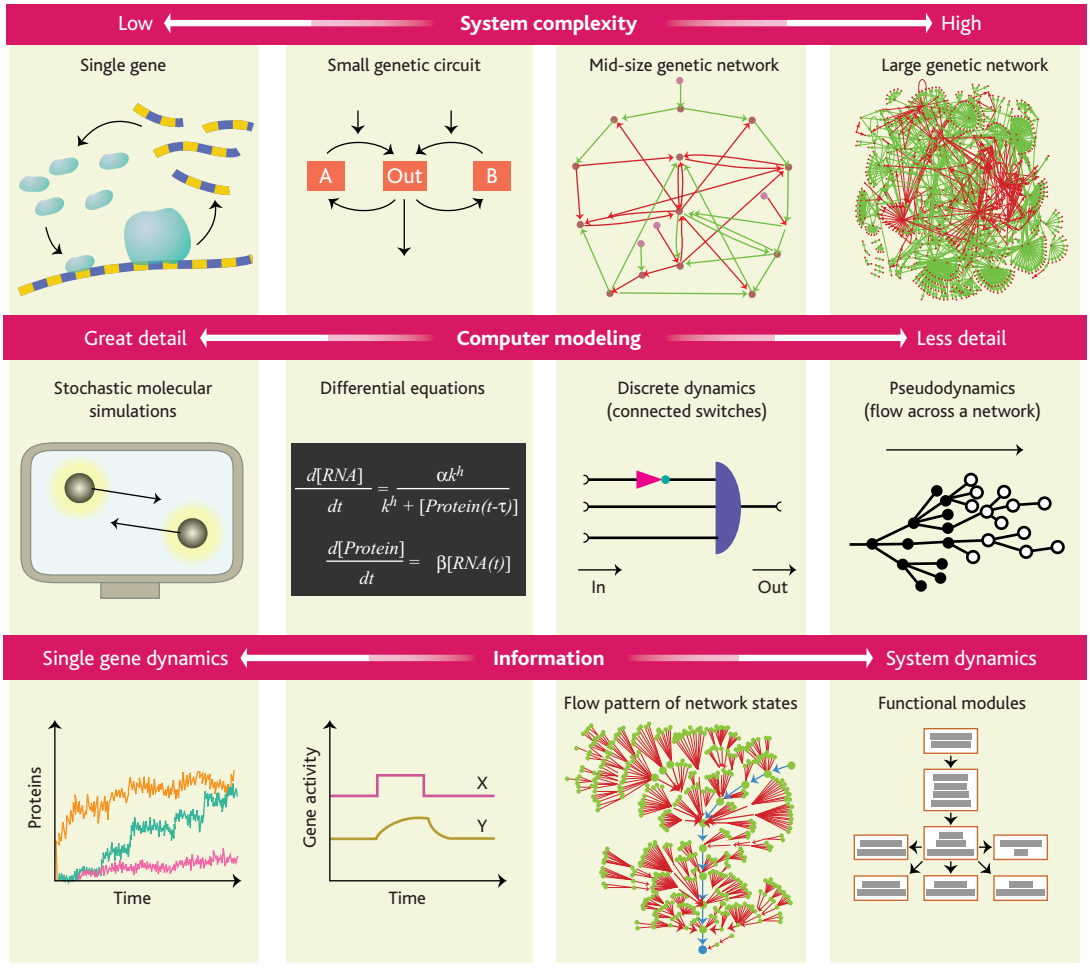
Less Is More in Modeling Large Genetic Networks

Stefan Bornholdt

As the biology of information processing in the living cell shifts from the study of single signal transduction pathways to increasingly complex regulatory networks, mathematical models become indispensable tools. Detailed predictive models of large genetic networks could revolutionize how researchers study complex diseases, yet such models are not yet within reach. One reason is that experimental data for large genetic systems are incomplete; another is that large genetic systems are difficult to model. Extrapolating the standard differential equations model of a single gene (with its several kinetic parameters) to large systems would render the model prohibitively complicated. One possible way to simplify such models would be to find a "coarse-grained" level of description for genetic networks; that is, to focus on the system behavior of the network while neglecting molecular details wherever possible (see the figure). Such an approach exists for other fields of science—for example, the concept of molecular orbitals in organic chemistry, which mercifully spares us from the details of the underlying quantum physics. On page 496 in this issue, Brandman *et al.* (*1*) points to the possibility of simplifying large genetic network models. Using a standard differential equations approach, the authors find that the intricate internal dynamics of a frequent cellular subcircuit exhibits a simple bistable "ON/OFF" behavior, and thus could be modeled by something much simpler than differential equations—something as simple as a switch.

A first level of coarse-graining in genetic regulation already exists in the standard approach of modeling protein and RNA concentrations with specific equations called "ordinary" differential equations. These equations nicely summarize the molecular interactions that make up the cellular machinery that regulates the activity of a gene. When at least a few tens of molecules are involved in regulating a gene, details of the interactions can usually be neglected, and interaction rates can be used instead of tracking the single molecular binding events (*2*).

With large networks involving thousands of regulatory genes (genes that encode proteins that regulate other genes), the number of differential equations needed to describe the system can become huge. The sheer number of parameters (such as decay rates, production rates, and interaction strengths) in this mathematical model poses a chal-



The different levels of description in models of genetic networks. Whereas single genes can be modeled in molecular detail with stochastic simulations (**left column**), a differential equation representation of gene dynamics is more practical when turning to circuits of genes (**center left column**). Approximating gene dynamics by switchlike ON/OFF behavior allows modeling of mid-sized genetic circuits (**center right column**) and still faithfully represents the overall dynamics of the biological system. Large genetic networks are currently out of reach for predictive simulations. However, more simplified dynamics, such as percolating flows across a network structure, can teach us about the functional structure of a large network (**right column**).

The author is with the Institute for Theoretical Physics, University of Bremen, Otto-Hahn-Allee, D-28359 Bremen, Germany. E-mail: bornholdt@itp.uni-bremen.de

Published by AAAS

lenge, both for experiment and theory. A central question is what the right level of description is when constructing quantitative models of large or even systemwide genetic networks (see the figure). Is coarse-graining of genetic network models possible?

A number of general building blocks identified in genetic networks at least indicate that robust simplified models are possible. Modules such as autoregulatory excitatory (positive) feedback loops (which can convert a transient signal into a sustained signal and thus serve as "storage" devices), inhibitory feedback loops (which suppress instability due to noise), or feed-forward loops (which may enhance responsiveness of a gene) represent different kinds of robust switching elements. Brandman *et al*. describe another such building block—the dual positive-feedback loop, which is frequently found in subnetworks of larger cellular and genetic networks. But why would cells have evolved two positive feedback loops when one is enough to create a switch? Brandman *et al*. find that the combination of the two loops can make genetic switching faster and, at the same time, reduce signal noise. A slow loop creates robustness in the signal, whereas a fast loop allows for switching speed. Given the quite complex cellular machinery that is needed to run this dual positive feedback circuit with biochemical means, its dynamic behavior is intriguingly simple. It functions as a particularly robust, yet fast switch that is reminiscent of the robustly designed electronic building blocks used to build modern computers.

This observation provides support for discrete models of genetic networks in which genes are modeled as switchlike dynamic elements that are either ON or OFF. The first such models, generated about 36 years ago, were random networks of discrete dynamical elements, as few data about regulatory genetic networks were available at the time (*3*). These models were long considered to be merely a speculative analogy. However, recent advances in modeling combined with the first opportunities to validate genetic network models with data from living cells show that simplified network models, such as those representing a regulatory gene as a binary (ON/OFF) switch, can indeed predict the overall dynamical trajectory of a biological genetic circuit. For example, the trajectory of the segment polarity network in the fly *Drosophila melanogaster* has been predicted solely on the basis of discrete binary model genes (*4*). Similarly, a dynamic binary model of the genetic network that controls the yeast cell cycle was constructed (*5*). In both systems, the dynamics converge to so-called attractors (states or sequences of states of the genes) and for these, the models match the biological dynamics. These dynamical attractors seem to depend not so much on the details of the kinetic constants, as on the circuit wiring. Insensitivity to biochemical kinetic parameters indicates that for understanding the dynamics of these circuits, it's their wiring that is most important (*6*). This seems to be why large genetic networks can be represented as networks of discrete dynamic elements, without the tuning of parameters. Simplified models on even larger scales are encouraged.

Modeling of large cellular networks is often hampered by incomplete knowledge of the full circuitry, despite a wealth of data. An example of how simplification of the dynamics of single elements enables us to gain valuable information about a system's function is presented in the recent article by Ma'ayam *et al*. (*7*). Here, discrete "pseudo-dynamics" of binary states simply percolate through the known part of a 1500-node mammalian cellular network and give a rough but informative estimate of the property of the regulatory information flow through the system. The thousands of parameters required to generate a standard differential equations model of all the relevant biochemical interactions has been neglected here in favor of a statistical perspective that provides valuable information about the global architecture of a cellular network. It is not a direct representation of the biochemical dynamics and does not allow a detailed dynamic simulation of the network. However, it is an analog of the potential propagation of a signal and therefore useful to determine the global signaling structure of an overall network. This approach is error tolerant and gives a robust picture of the overall global modular structure of a network.

The simple dynamics of the building blocks points to an interesting perspective for our further understanding of genetic networks. Distinguishing between the robust effective dynamics of a genetic or regulatory switch and the biochemical means to practically run it shows that, to understand the system, we do not have to retrace all the details of the biochemistry. Characterizing the circuit wiring seems to be the most important consideration, and when going "dynamic," a clever way to throw away details may be the most important part of model building.

References

1. O. Brandman, J. E. Ferrell Jr., R. Li, T. Meyer, *Science* **310**, 496 (2005).
2. H. H. McAdams, A. Arkin, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814 (1997).
3. S. A. Kauffman, *J. Theor. Biol.* **22**, 437 (1969).
4. R. Albert, H. G. Othmer, *J. Theor. Biol.* **223**, 1 (2003).
5. F. Li, T. Long, Y. Lu, Q. Ouyang, C. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4781 (2004).
6. A. Wagner, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11775 (2005).
7. A. Ma'ayan *et al*., *Science* **309**, 1078 (2005).

10.1126/science.1119959

CHEMISTRY

The Renaissance of Natural Products as Drug Candidates

Ian Paterson and Edward A. Anderson

Around half of the drugs currently in clinical use are of natural product origin (*1*, *2*). Despite this statistic, pharmaceutical companies have embraced the era of combinatorial chemistry, neglecting the development of natural products as potential drug candidates in favor of high-throughput synthesis of large compound libraries (*3*). Perhaps it is time to reassess this prevailing dogma for chasing quantity over quality.

Cancer chemotherapy, in particular, presents an ideal opportunity for natural product–inspired drug discovery and development. Unfortunately, many of the most promising natural lead compounds are available only in extremely small quantities, especially those from marine organisms such as sponges. The reluctance of industry to pursue such bioactive natural products as potential drugs lies primarily in the perceived supply problem. This leaves organic synthesis as a key option for sourcing these important drug candidates for preclinical and clinical studies. However, the academic-style approach to "hot target molecules" usually results in lengthy synthetic routes owing to their often exquisitely complicated architectures, with long development times, low overall yields, and impracticality of scale-up and provision of diverse structural analogs.

An alternative approach to drug discovery, which has been embraced by the phar-

The authors are in the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: ip100@cam.ac.uk