Science Signaling
AAAS

Rules for Modeling Signal-Transduction Systems

William S. Hlavacek, James R. Faeder, Michael L. Blinov, Richard G. Posner, Michael Hucka and Walter Fontana (18 July 2006)

Sci. STKE **2006** (344), re6. [DOI: 10.1126/stke.3442006re6]

A Correction to the Review Titled "Rules for Modeling Signal-Transduction Systems"

(Published 25 July 2006)

The affiliations for two of the authors, James R. Faeder and Michael L. Blinov, were incorrectly listed as the "Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA." James R. Faeder is affiliated with the "Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA" and Michael L. Blinov is affiliated with the "Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT 06030, USA."

The corrected full text online and a corrected PDF are available (http://stke.sciencemag.org/cgi/content/full/sigtrans;2006/344/re6) starting 25 July 2006.

The original, uncorrected PDF can be accessed from http://stke.sciencemag.org/cgi/reprint/sigtrans;2006/345/er7.

Citation: A Correction to the Review Titled "Rules for Modeling Signal-Transduction Systems" by W. S. Hlavacek *et al. Sci. STKE* **2006**, er7 (2006).

Rules for Modeling Signal-Transduction Systems

William S. Hlavacek,[1]* James R. Faeder,[2] Michael L. Blinov,[3] Richard G. Posner,[4] Michael Hucka,[5] Walter Fontana[6]

(Published 18 July 2006)
(Revised 25 July 2006)

Formalized rules for protein-protein interactions have recently been introduced to represent the binding and enzymatic activities of proteins in cellular signaling. Rules encode an understanding of how a system works in terms of the biomolecules in the system and their possible states and interactions. A set of rules can be as easy to read as a diagrammatic interaction map, but unlike most such maps, rules have precise interpretations. Rules can be processed to automatically generate a mathematical or computational model for a system, which enables explanatory and predictive insights into the system's behavior. Rules are independent units of a model specification that facilitate model revision. Instead of changing a large number of equations or lines of code, as may be required in the case of a conventional mathematical model, a protein interaction can be introduced or modified simply by adding or changing a single rule that represents the interaction of interest. Rules can be defined and visualized by using graphs, so no specialized training in mathematics or computer science is necessary to create models or to take advantage of the representational precision of rules. Rules can be encoded in a machine-readable format to enable electronic storage and exchange of models, as well as basic knowledge about protein-protein interactions. Here, we review the motivation for rule-based modeling; applications of the approach; and issues that arise in model specification, simulation, and testing. We also discuss rule visualization and exchange and the software available for rule-based modeling.

Introduction

Many diseases are caused by molecular changes that affect signal-transduction systems, and some of these diseases, such as chronic myelogenous leukemia, can now be treated by drugs that target signaling proteins, such as kinases (*1*). Thus, in addition to our curiosity about the fascinating mechanisms that cells use to respond to signals, there is practical motivation to better understand the processes of cellular signaling, in which protein-protein interactions play a central role. Indeed, we would like to make accurate predictions about the functional roles of proteins and the effects of modifying the interactions of proteins in particular systems. Our ability to make such predictions is a measure of our basic understanding of molecular cell biology and is likely to have practical consequences in drug discovery (*2*, *3*).

The behavior of a signal-transduction system depends on dynamic interactions among its proteins (*4*, *5*). The combined effects of these interactions are difficult to predict from intuition alone. When intuition is insufficient, a mathematical model is often useful for acquiring a quantitative and predictive understanding of a complex dynamical system, and mathematical modeling is being increasingly used to aid in studies of cellular signaling (*6*, *7*). Models have now been developed and tested for signaling events mediated by several well-studied cell surface receptors, including the epidermal growth factor receptor (EGFR) (*8*) and two of the antigen-recognition receptors of the immune system (*9*).

However, current models are still far from capturing all of the relevant mechanistic details of signal-transduction systems that must be considered to provide realistic and complete pictures of how these systems work (*10*). In particular, models often fail to account for the complexities of protein-protein interactions, such as how these interactions depend on contextual details at the level of protein sites. Also, few models account for the enormous number of possible posttranslational covalent modifications of proteins and for formation of all the possible protein complexes (*11*). A major reason for this shortcoming is the strain on conventional modeling approaches caused by the combinatorial potential of protein-protein interactions. New modeling approaches that address this problem involve the use of rules to represent protein-protein interactions. (Rules are also useful for representing other types of biomolecular interactions, but we will focus on protein-protein interactions for purposes of discussion.) The introduction of rules greatly eases the task of specifying a model that incorporates details at the level of protein sites. A rule—such as "ligand binds receptor with rate constant k whenever ligand and receptor have free binding sites"—describes the features of reactants that are required for a particular type of chemical transformation to take place. Rules simplify the specification of a model when the reactivity of a component in a system is determined by only a subset of its possible features. Here, we review rule-based modeling of signal-transduction systems.

Modeling Goals and Challenges

What do we expect from a model? A model should incorporate enough details to make testable predictions about quantities that can be measured and controlled in experiments. Whenever possible, the parameters of a model should have a physical rather than phenomenological basis, such that parameters are independent of system behavior. A model should provide insights and guide experimentation by revealing the logical consequences of knowledge and assumptions about the mechanistic details of a system, including the proteins in the system and their binding sites, enzymatic activities, and sites of posttranslational modification. The development of models at this level of resolution is

[1]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [2]Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [3]Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT 06030, USA. [4]Translational Genomics Research Institute, Phoenix, AZ 85004, USA. [5]Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA. [6]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

*Corresponding author. E-mail, wish@lanl.gov

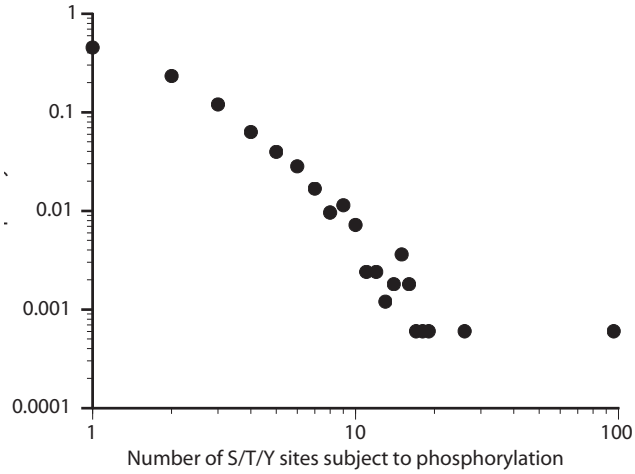


Fig. 1. Frequency distribution of multisite protein phosphorylation. This plot is based on information about phosphorylation of serine (S), threonine (T), and tyrosine (Y) residues of 1663 proteins documented in version 3.0 of the Phospho.ELM database (http://phospho.elm.eu.org/). The vertical axis indicates the fraction of proteins; the horizontal axis shows the number of phosphorylation sites. More than half the proteins are phosphorylated at two or more sites.

justified by current and emerging capabilities for characterizing systems at the level of protein sites and for monitoring the dynamics of multiple readouts of protein interactions (*12*–*16*). Thus, although other levels of abstraction may prove useful, we are most interested in models with physicochemical parameters that capture details about proteins and their interactions at the level of protein sites, which we take to include motifs, domains, and subunits.

General features of proteins and protein-protein interactions. Protein-protein interactions triggered by a signal can lead to various covalent posttranslational protein modifications, such as phosphorylation, ubiquitination, acetylation, or methylation (*4*, *5*, *17*). The effect of such modification, which is often reversible, may be a change in a protein's enzymatic or binding activity, location, or rate of turnover. Signal-regulated protein-protein interactions also mediate the assembly of heterogeneous protein complexes. A common effect of complex formation is an increase in the activity and specificity of an enzyme through colocalization of the enzyme with a substrate (*18*, *19*). Signal-induced protein complexes are also known to regulate enzymatic activity through allosteric mechanisms (*20*, *21*).

The enzymatic and binding activities of proteins involved in signaling tend to be localized to modular domains (*19*, *22*), such as the protein tyrosine kinase (PTK) and Src homology 2 (SH2) domains of a Src family kinase. Proteins may also contain smaller parts, like immunoreceptor tyrosine-based activation motifs (ITAMs) (*23*), that are sites of modification or binding or both (*24*). A signaling protein generally has multiple domains (*25*, *26*) and binding sites (and binding partners, which may have overlapping specificities), as well as multiple sites subject to posttranslational modifications (which may include the binding sites themselves) (*27*) (Fig. 1). Such multiplicity may permit many combinations of modification and binding events, which can be difficult to track in a model (*11*, *28*, *29*).

Combinatorial complexity. As a modeler considers more proteins and protein sites of a signal-transduction system, the number of possible protein complexes and combinations of protein modifications tends to increase exponentially. Moreover, hundreds to thousands of chemical species may be generated by the interactions of only a few proteins (*28*, *30*–*33*). Thus, a signal-transduction system can be quite large when viewed as a chemical reaction network. We refer to this potential for large size as "combinatorial complexity."

One source of combinatorial complexity, to which we have already alluded, is multisite protein modification (*27*). Consider the EGFR. According to a recent review (*34*), at least nine tyrosines in EGFR are phosphorylated during signaling (Fig. 2). Phosphorylation of these tyrosines is mediated by a nonreceptor PTK, Src, or EGFR itself when the PTK domain of one EGFR in a ligand-induced receptor dimer catalyzes phosphorylation of the paired EGFR in the dimer. As a simplification, let us assume that only one of nine tyrosines of EGFR can be phospho-

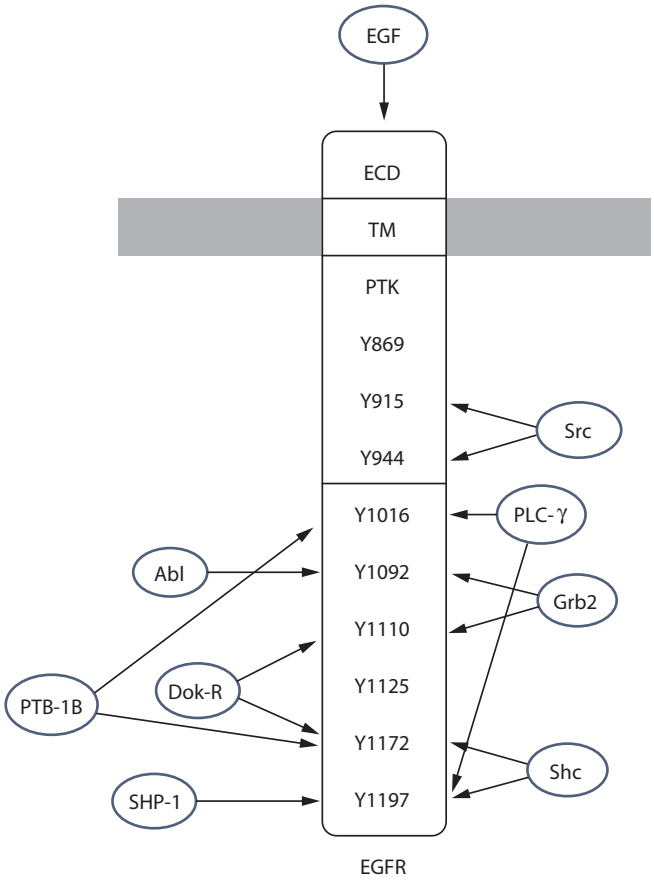
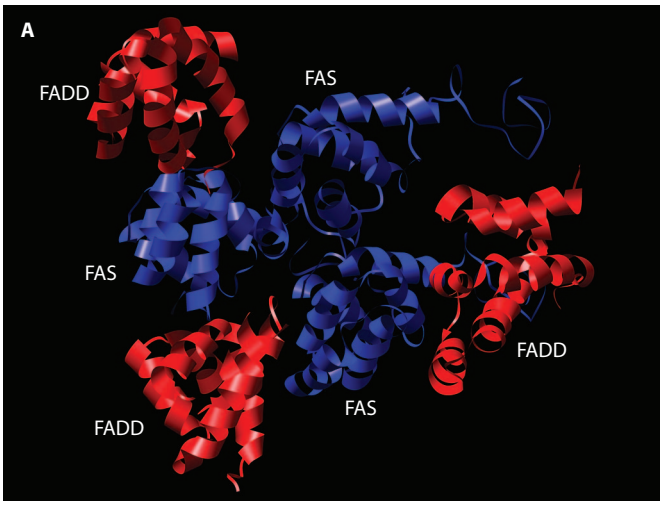


Fig. 2. Multisite phosphorylation of EGFR. The nine tyrosine residues indicated here are phosphorylated by EGFR or Src (*34*). Phosphorylation of particular sites regulates interactions of EGFR with intracellular binding partners, eight of which are indicated in the figure [including PLC-γ (phospholipase C–γ)]. These binding partners bind EGFR through their SH2 and/or phosphotyrosine binding (PTB) domains. Other tyrosine residues of EGFR are also subject to phosphorylation, and additional binding partners of EGFR are known (*14*, *16*). Numbering of residues is based on the full-length sequence of EGFR.

rylated at a time. Under this assumption, there are still 10 different phosphoforms of EGFR and 55 distinct combinations of these phosphoforms for a receptor dimer. These numbers grow to 512 and 131,328 if more than one tyrosine of EGFR can be phosphorylated at a time. The number of phosphorylation states relevant for signaling in particular contexts is unknown, and it seems unlikely that all states are functionally relevant or even realized. Indeed, some phosphorylation states may be prohibited (*35*). However, a modeler may wish to consider the full spectrum of possible phosphorylation states to assess their relevance in an unbiased manner. The states included in a model will depend on the questions asked and the opportunities a modeler may have to make meaningful simplifications.



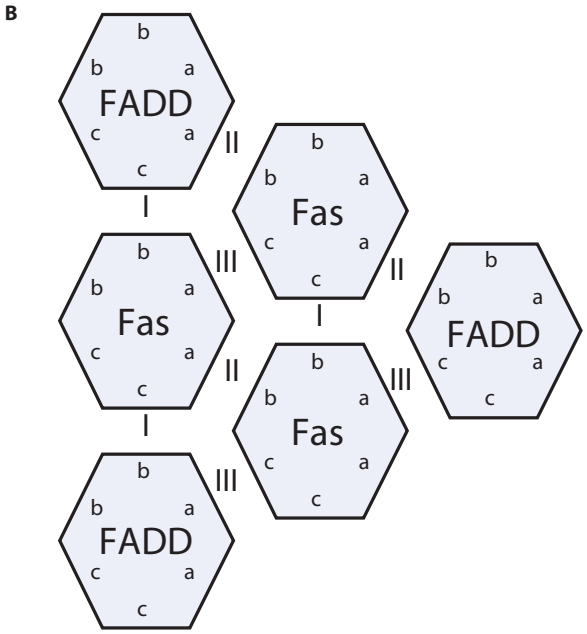


Fig. 3. Multivalent binding of the DD components of Fas and FADD. (**A**) View of a three-dimensional model of a 3:3 heterohexamer of the DDs derived from knowledge-based docking of homology models of the individual DDs (*38*) [reproduced by C.-S. Tung, Los Alamos National Laboratory]. The configuration of DDs is planar in this complex. (**B**) A schematic model of the same complex indicating that the DDs interact through three types of interfaces. The type I and II interfaces are seen in crystal structures of other protein complexes composed of members of the DD superfamily (*39*, *40*). These interfaces were used to guide docking. The type III interface is a prediction of the model (*38*). Binding of Fas and FADD DDs through the different interfaces can produce a variety of protein complexes.

Another source of combinatorial complexity is multivalent binding. Consider the death domain (DD), a protein interaction domain found in many proteins (*36*, *37*). A model of a complex of DDs of the death receptor Fas (also called CD95) and the Fas-associated protein with death domain (FADD) are shown in Fig. 3 (*38*). The model, which is based on interfaces seen in crystal structures (*39*, *40*), predicts that the Fas and FADD DDs form a hexamer through three interfaces. In fact, interactions through the three interfaces can generate a distribution of homo- and heterooligomers: 6 homodimers, 6 heterodimers, 20 homotrimers, 66 heterotrimers, and so on. Without regard for physicochemical limitations, the reaction network implied by the DD-DD interactions is of infinite size. Of course, the actual distribution of oligomers will be limited by a number of factors, such as protein copy numbers, binding affinities, steric effects, and cooperativity. However, to elucidate how these factors shape the distribution, a modeler may wish to account for all possibilities.

Dynamical models that account comprehensively for the possible species in a system can help determine which of these species are relevant for signaling, under what conditions, and why. In an analysis of such a model (*41*), Faeder *et al*. (*42*) found that only a small fraction of the possible species in the model is effectively populated, although there are enough proteins to populate all species in the model at a nontrivial level. When parameters that affect system dynamics (such as protein concentrations) change, the populated species can shift dramatically. These shifts are not predicted by reduced models that omit species. Other theoretical analyses also indicate that the populated species in a system are influenced by parameters that affect system dynamics (*43*–*45*). Thus, if certain species of a system appear to be favored, as in temporal ordering of phosphorylation events (*35*), a dynamical explanation might be suggested and evaluated with the aid of a model that has the capability to predict how the favored species will be affected by perturbations of system dynamics. A model that has such a capability is one that accounts for all possible species.

Approaches to Model Specification

Combinatorial complexity raises a series of problems when one is modeling signal-transduction systems as chemical reaction networks. One problem is model specification, the task of stating one's knowledge and assumptions about a system in a way that enables mathematical analysis.

Conventional approach to model specification. The textbook approach to modeling a biochemical system is to draw a reaction-scheme diagram depicting the chemical species and reactions in the system and then to translate this diagram, which is essentially an organized layout of a list of reactions, manually into a set of equations (*46*), such as a system of coupled ordinary differential equations (ODEs). Reaction-scheme diagrams that have been drawn for signal-transduction systems, some of which are quite large (*47*, *48*), are based mostly on knowledge and as-

sumptions about protein-protein interactions. An example of a simple scheme is provided in Fig. 4. A drawback of a reaction-scheme diagram is that it obscures the underlying protein-protein interactions by not explicitly representing them. Still, a diagram is far easier to interpret than the corresponding equations, and the readability of a diagram can be improved by using iconographic annotation. For example, Fig. 5 shows how the scheme of Fig. 4 can be made more readable with the notations of Kitano *et al*. (*49*) and Faeder *et al*. (*50*). In some cases, a reaction scheme can serve the purpose of model specification well.

Software tools have been developed to help biologists to draw reaction-scheme diagrams, to translate these diagrams into mathematical equations, and to perform model-based calculations by using standard methods of scientific computing (http://sbml.org/). Examples include CellDesigner (*51*) and the Virtual Cell (*52*). However, these tools are useful only if the textbook approach to model specification can be applied. In many cases, combinatorial complexity makes this approach prohibitively time-consuming and error-prone, even with the aid of powerful software tools for drawing reaction-scheme diagrams (*53*).

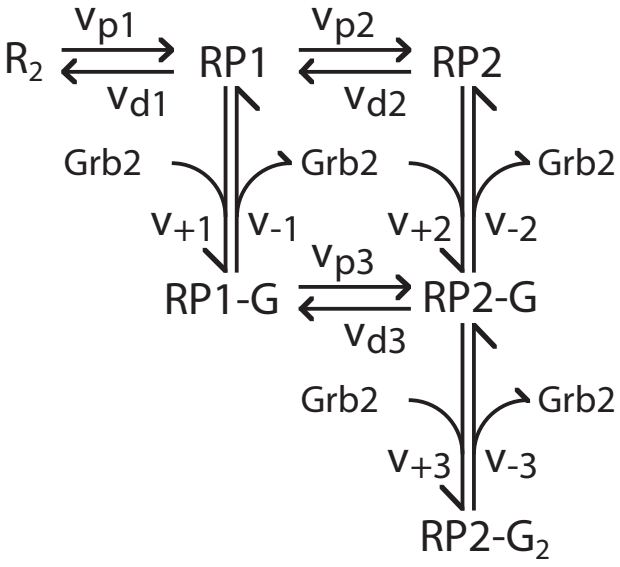


Fig. 4. A reaction scheme. The scheme includes seven chemical species, labeled R_2, RP1, RP2, Grb2, RP1-G, RP2-G, and RP2-G_2, and 12 reactions, labeled v_{p1}, v_{p2}, v_{p3}, v_{d1}, v_{d2}, v_{d3}, v_{+1}, v_{-1}, v_{+2}, v_{-2}, v_{+3}, and v_{-3}. Such notation is typical. Each label must be defined before this reaction scheme can be understood. Also, to derive a mathematical model from this scheme, a rate law must be assigned to each reaction.

Limits of conventional modeling. Although the conventional approach to model specification is problematic for signal-transduction systems, which are marked by combinatorial complexity, it is nevertheless the approach most often used to specify models for these systems, but at a cost. Models derived in this way are invariably based on assumptions, which may be difficult to justify, that limit the chemical species and reactions considered to a fraction of those possible. An example is the model of Kholodenko *et al*. (*54*) for EGFR signaling, which has been extended by a number of researchers (*55*–*58*). This model is based on mechanistic assumptions that result in a selective focus on only a fraction of the protein complexes and phosphorylation states that could potentially arise from the protein-protein interactions considered in the model. For example, one assumption is that ligand-induced dimers of EGFR are unable to dissociate when receptors are phosphorylated, which seems unlikely. This assumption arises from a description of signaling events as an ordered pathway, which is consistent with the way this system is presented in typical diagrammatic interaction maps, but inconsistent with rapidly reversible reactions and multiple branching possibilities. Lifting this and other assumptions causes a combinatorial explosion in the number of possible reactions and species (*59*), which makes manual model specification impractical.

Rule-based approach to model specification. Given that protein-protein interactions can generate large reaction networks, what can be done to capture the essence of these interactions without ignoring their combinatorial complexity? To address this question, a number of research groups have suggested, in one form or another, a new starting point for model specification. The basic idea is to specify protein-protein interactions as rules that serve as generators of chemical reactions (or reaction events) and species. A rule specifies the features of proteins that are required for or affected by a particular protein-protein interaction. A rule can be viewed as a definition of a reaction class, a generalized reaction. The relevant features of an interaction can often be identified, at least to a first approximation, because of the modularity of protein catalytic and interaction domains (*19*).

In one approach to rule-based modeling (*60*, *61*), which is implemented in the BioNetGen software package, rules are used as generators of chemical species and reactions as follows. A rule comprises patterns for recognizing reactants, a mapping of reactants to products, and a rate law. Given a set of initial chemical species, represented with text strings or graphs, each rule is used to identify, through pattern matching, the species that have features required to undergo the transformation from reactants to products specified in the rule. Each transformation is performed to obtain product species and the transformation is assigned a rate law, the one associated with the corresponding rule. By convention, it is assumed that the interaction represented in a rule is independent of features not explicitly indicated. Thus, multiple species may qualify as reactants in a type of reaction defined by a rule, and multiple reactions may be generated that have the same characteristic rate law, although parameters of the rate law may need to be adjusted for a variety of contextual reasons (*60*, *62*). The exact number of reactions generated by a rule depends, in general, on the entire set of rules in which the rule of interest is embedded and also on the set of species to which rules are initially applied. The species and reactions may either be generated in advance of a simulation, the "generate-first" approach, or during the course of a simulation, the "on-the-fly" approach (*60*, *62*, *63*). With the generate-first approach, a network of species and reactions is generated through iterative application of the rules until a specified, arbitrary termination condition is satisfied or no new reactions are generated (*60*). With the on-the-fly approach, reactions are generated as new species become populated, which may be advantageous when the network is large or unbounded, as is the case when rule application is nonterminating in the absence of an arbitrary halting condition (*60*, *62*). An example of a rule for which rule evaluation is nonterminating is provided in Fig. 6.

Applications of rule-based modeling and first software tools for model specification. To date, rule-based modeling has been applied to only a handful of systems, but there is an increasing awareness of the need for such models. In considering whether a particular application represents an example of a rule-based model, we ask the following question: Are the chemical species and reactions in the model drawn from a fixed user-specified list or are they generated automatically in some way by a computer algorithm from a set of rules? The latter represents what we call rule-based models. The programs used to generate rule-based models can be divided into two categories: those that are specific to a particular application and those that generalize to a range of problems.

Rule-based modeling has been applied to bacterial chemotaxis mediated by the Tar-receptor complex [for reviews of the biology, see (*64*, *65*)]. The complexity of models for receptor-mediated phosphorylation events (*66*) and the aggregation of receptors (*67*) led Bray and co-workers to develop two general-purpose computer programs, OLIGO (*68*) and STOCHSIM (*69*). These programs divide the modeling problem into two parts—formation of multisubunit signaling complexes, and phosphorylation cascades.

OLIGO uses graphs to represent multiprotein complexes, with each node representing a protein and each edge representing a noncovalent interaction between two proteins. A user draws a contact graph with the help of a graphical user interface (GUI). OLIGO then generates a reaction network by disassembling the complex in all possible ways. For example, the six-subunit complex that is active in Tar-mediated signaling is decomposed into its three protein constituents, and in the process, a network of 14 reversible binding reactions is generated. A subsequent program, OLIGO-D (*43*), generates a network in the opposite order, starting with a list of proteins and binding sites. Both programs generate networks on the basis of the assumption that rings form whenever possible. This simplification prevents polymerization through chain propagation. Although this assumption may be valid in some cases, there are cases where extensive oligomerization of signaling proteins occurs (*70*). A further and more severe limitation of these software tools is that regulation of protein interactions through covalent posttranslational modifications, such as phosphorylation, cannot be considered.

Use of OLIGO-D provided a qualitative explanation for how signaling networks can exhibit a decrease in signaling when complex-forming proteins are overexpressed, a phenomenon we will refer to as high-dose inhibition. Bray and Lay (*43*) found that bridging subunits that connected separate parts of a complex were particularly likely to exhibit such effects, but also showed that any subunit with multiple bonds to a complex could exhibit high-dose inhibition if the binding constants in the network were optimized for such behavior. Levchenko *et al*. (*44*) modeled high-dose inhibition for the particular contact geometry of a scaffold protein that binds kinases of a mitogen-activated protein (MAP) kinase cascade. Although this model was constructed manually, a Mathematica-embedded biochemical modeling package called Cellerator (*71*), an early example of software for rule-based modeling, was later used to generate automatically and to solve the differential equations for a generalized model of kinase-scaffold interactions (*72*). Specific functions (rules) were written to generate the species and reactions for scaffolds and scaffold dimers with various numbers of binding sites and other properties. The effects of previous assumptions made to restrict network complexity, such as a rapid second phosphorylation step for kinase activation, were investigated and found to have a quantitative, but not qualitative, effect on the observed behavior of the system. In particular, the scaffold concentration for optimal downstream activation was found to be insensitive to modeling assumptions.

STOCHSIM (*73*, *74*), the other main modeling package developed by Bray and co-workers, is complementary to OLIGO in that it allows detailed modeling of the kinetics of covalent modifications, but has fairly severe limitations for modeling complex formation. Further discussion of STOCHSIM's design and capabilities is provided below. An interesting feature of StochSim is that it can be used to consider nearest-neighbor interactions on regular two-dimensional lattices. This capability was developed to study lateral interactions among receptors in a static configuration (*75*, *76*).

Goldstein and co-workers have developed a detailed model of early events in signaling by FcεRI (*41*), the high-affinity receptor for immunoglobulin E (IgE) antibody, which plays an important role in allergic reactions. The model considers phosphorylation and binding events initiated by receptor aggregation and encompasses 354 chemical species and 3680 chemical reactions, but it is based on only 25 parameters (21 rate constants and 4 protein concentrations). This economy of parameters is achieved by using reaction rules describing conditional protein binding and phosphorylation events to generate the species and reactions that arise in the network model. Each reaction generated by a given rule uses the same rate constant. Thus, reaction rates are assumed to depend on a subset of the features of reactant species. A typical assumption is that the rate constant for ligand-receptor binding is not affected by the cytoplasmic state of the receptor. Analysis of this model has provided a number of insights into how the rate constants and affinities that govern particular interactions within receptor aggregates affect the outcome of signaling (*41*). For example, it was found that the multiple requirements for transphosphorylation within a receptor aggregate enable the system to discriminate effectively between ligands that bind with short and long half-lives (*9*), an effect known as kinetic proofreading (*77*).

Chakraborty and co-workers have developed two distinct models of early events in T cell receptor (TCR) signaling that have led to interesting insights and hypotheses about the effects of spatial organization (*45*) and cooperative interactions (*78*) on signal amplification. In the model of Lee *et al*. (*45*), the junction between a T cell and a stimulatory antigen-presenting cell is represented by a three-dimensional lattice on which various proteins involved in early signaling events diffuse and react. A set of interaction rules based on a detailed reaction scheme is used to determine which biochemical events can occur between a given pair of particles. At each simulation step, one of three possible update moves is selected at random with equal probability: diffusion of a molecule or complex, association or dissociation involving a pair of molecules chosen at random, or a state transition involving a pair of molecules chosen at random. This method avoids the problem of explicitly generating species and reactions but is restrictive in that the computational cost scales with the number of particles squared, limiting simulations to relatively small numbers. A more severe problem that is specific to this model is that although the model includes both reaction and diffusion effects, little attempt is made to simulate the cor-

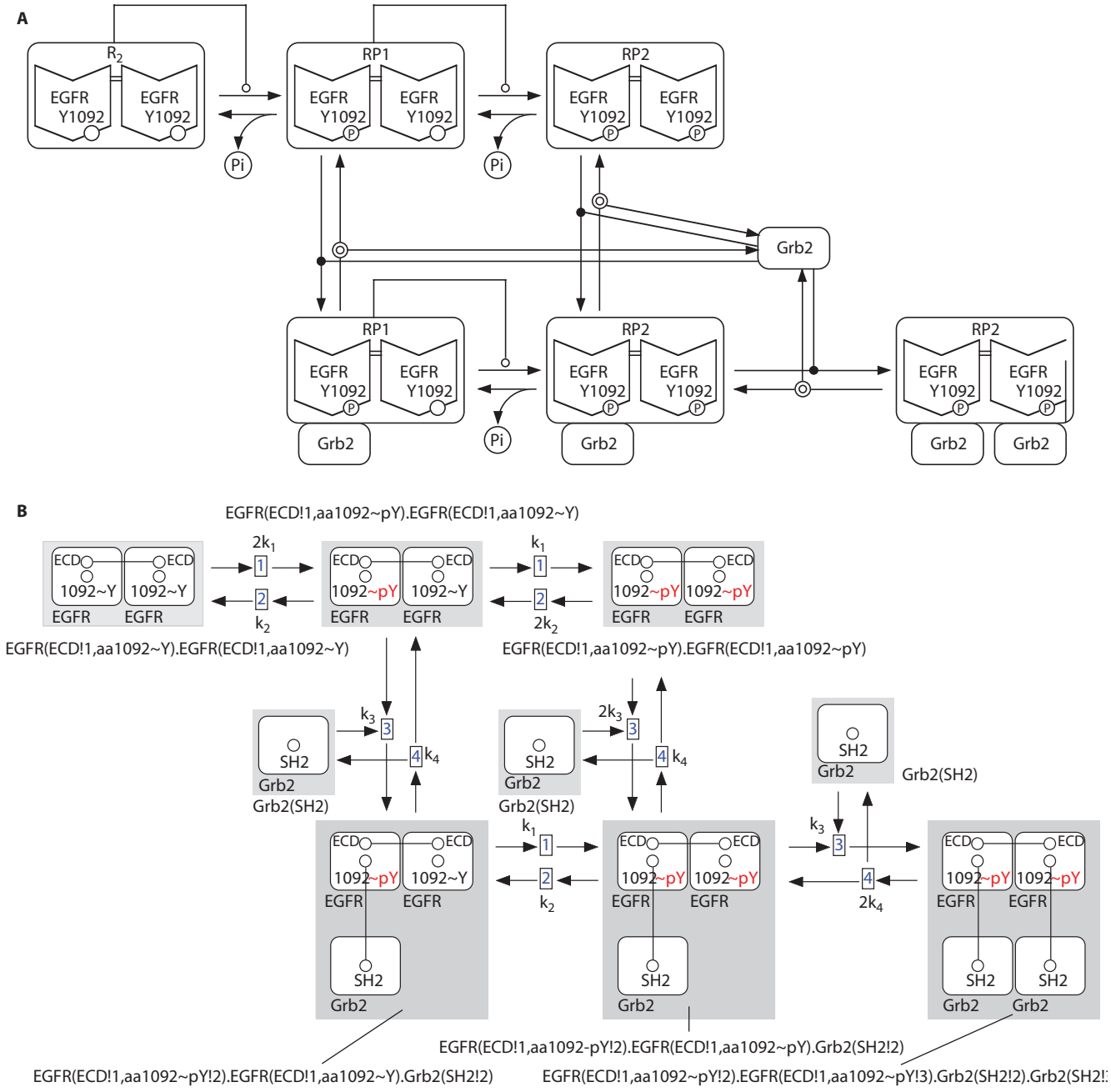


Fig. 5. Elaborations of the reaction scheme shown in Fig. 4. (**A**) A process diagram drawn according to the conventions of Kitano *et al*. (*49*). (**B**) The same scheme drawn according to the conventions of Faeder *et al*. (*50*) and Blinov *et al*. (*61*) for graph-based representation of chemical species and reactions. Below each chemical species graph, a corresponding definition in the BioNetGen Language (BNGL) is given. The meaning of the scheme when drawn as in either of these two panels is more self-evident than when drawn in the conventional way, as in Fig. 4.

rect rates of reaction and diffusion relative to each other or the correct absolute rates of reactions. Because quantitative simulations can be performed by using physically correct and computationally tractable methods (*79*), it might be interesting to do so to see if conclusions inferred from the original model change.

A different, nonspatial model was used by Li *et al*. (*78*) to model cooperative effects involving the TCR, its coreceptor CD4, and its ligands [peptide–major histocompatibility complex (MHC) complexes]. The model was designed to test the hypothesis that low-affinity binding between MHC molecules bound to endogenous (self) peptides could amplify signaling by a small number of high-affinity peptide-MHC complexes containing antigenic (foreign) peptides. The basic mechanism for cooperative enhancement was proposed to be the formation of a pseudo-

dimer complex, in which a high-affinity bond between antigenic peptide-MHC and a TCR acts as a scaffold to recruit a CD4 molecule, its constitutively associated molecule of the kinase Lck, and a second peptide-MHC molecule, most likely bound to endogenous peptide. The pseudo-dimer complex serves as an efficient machine for TCR activation, because TCRs that transiently bind the free peptide-MHC in the complex come into close proximity with the Lck, which phosphorylates TCR as the initial step in TCR activation. Reaction rules are given that describe the association and dissociation of these components, as well as phosphorylation and dephosphorylation, up to hexameric complexes (the size of the pseudodimer loaded with TCR), generating a network of 741 distinct reactions. A special-purpose program was used to generate the network, and dynamics were simulated by using the Gillespie algorithm (*80*, *81*). The model is constructed and parameter values are chosen so as to favor the formation of pseudo-dimer complexes only when high-affinity peptide-MHC is present. Under these conditions, cooperative enhancement of TCR signaling is observed at low densities of antigenic peptide, and the strength of this enhancement depends strongly on the affinity of interactions between TCR and the low-affinity (endogenous) peptide-MHC, which suggests a possible biological role for positive selection of TCRs that exhibit a relatively high affinity for endogenous MHC. On the basis of new experimental evidence, Krogsgaard *et al*. (*82*) have proposed a modification of the pseudodimer model in which the second TCR is recruited to the complex through its interaction with CD4, and it would be interesting to see how this change might affect the behavior of the network model.

Woolf and Linderman (*83*) used rules to model binary interactions among sets of G protein–coupled receptors (GPCRs), which they took to be in active or inactive states. Using Monte Carlo methods, they determined how rules for GPCR dimerization and changes of these rules affected the spatial distribution of receptors on a two-dimensional grid and receptor signaling. Using their rule-based models, they were able to make predictions consistent with experimental measurements of receptor clustering and second messenger signaling.

Haugh *et al*. (*84*) used a series of rule-based models to investigate mechanisms by which protein tyrosine phosphatases (PTPs) might regulate signaling through their association with membrane-bound receptors. Several cytosolic PTPs contain SH2 domains that allow them to bind directly to phosphorylated receptors. These interactions can increase PTP activity through at least three mechanisms: (i) direct allosteric activation, (ii) indirect allosteric activation through phosphorylation of the PTP by a receptor-associated kinase, and (iii) proximity to receptor phosphotyrosine substrates. Haugh *et al*. (*84*) found that the relation between PTP activity and receptor activation could be tuned by alterations in the relative expression levels of proteins that competed with PTPs for binding sites, leading to high-dose inhibition (with respect to the degree of receptor activation) under some conditions. The differential equations of the model were generated by a special-purpose program. The model also included algebraic equations, because Michaelis-Menten rate laws were used to characterize the kinetics of phosphorylation and dephosphorylation reactions with multiple substrates. These equations can be avoided by using elementary reactions of the Michaelis-Menten mechanism to model catalytic steps (*85*), at the cost of increasing the size of the reaction network (*86*).

An important application of rule-based modeling has been to examine the validity and implications of assumptions that have traditionally been made to limit the extent of combinatorial complexity. Conzelmann *et al*. (*87*) and Blinov *et al*. (*59*) have both developed rule-based versions of the EGFR signaling model originally developed by Kholodenko *et al*. (*54*). Because the primary focus of

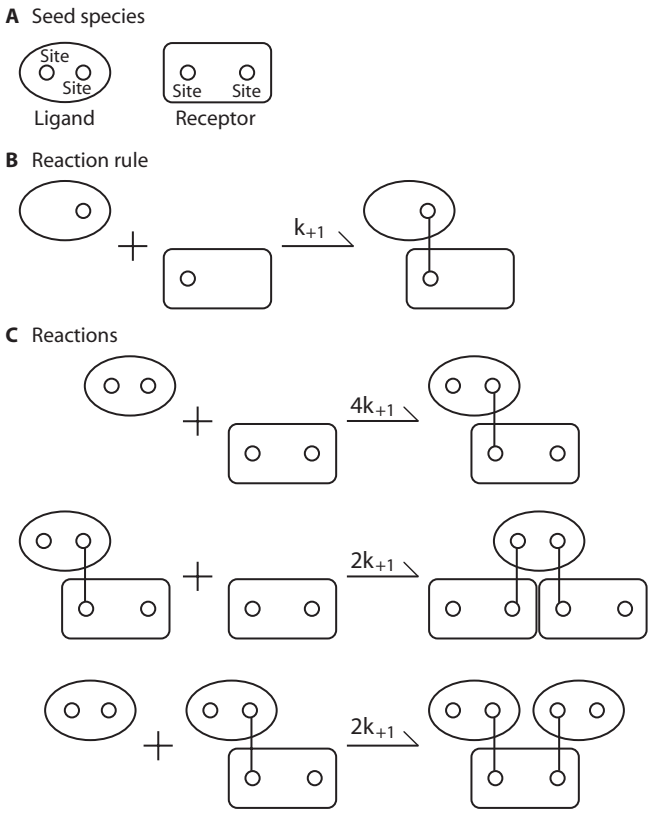


Fig. 6. An example of a BioNetGen model specification, made up of a single rule and set of two seed species, for which the iterative rule evaluation procedure of Faeder *et al*. (*60*) and Blinov *et al*. (*61*) is nonterminating. (**A**) The seed species, a free ligand (oval) and a free receptor (box with rounded corners). Both the ligand and receptor have two identical binding sites (circles). The species are represented as graphs according to the conventions of Faeder *et al*. (*50*) and Blinov *et al*. (*61*). (**B**) The rule, which is a graph-rewriting rule, defines a class of irreversible bimolecular association reactions in which the products are polymer-like chains of ligands and receptors. The rule indicates that a bond can form between a ligand with a free binding site and a receptor with a free binding site if the ligand and receptor are not already associated directly or indirectly. The plus symbol in the rule serves to specify a molecularity of two for reactions generated by the rule. (**C**) Reactions generated after two rounds of rule evaluation. The first reaction is generated after one round of rule evaluation. The product of this reaction is a 1:1 ligand-receptor aggregate, a new species. The next round of rule evaluation generates two new reactions, the products of which are two new species (1:2 and 2:1 ligand-receptor aggregates). Each subsequent round of rule evaluation generates at least one new reaction in which a product is a new ligand-receptor aggregate composed of more molecules than any aggregate previously generated. Thus, generation of all possible reactions implied by a rule for a protein-protein interaction is not always possible.

Conzelmann *et al*. (*86*) was model reduction, which we discuss elsewhere, we focus here on the efforts of Blinov *et al*. (*59*).

The rule-based version of the EGFR model was developed by generalizing the individual reactions in the Kholodenko model with the BioNetGen software package (*88*). Each reaction was assumed to be a specific instance of an interaction that could be expressed more generally as a reaction rule by using the rate constant of the original reaction. Thus, there is a one-to-one correspondence between reactions in the original model and reaction rules in the expanded model, and no new rate parameters are introduced. As an example, in the original model, dimerization of EGFR bound to the epidermal growth factor (EGF) is described by a single reaction with forward and reverse rate constants k_{+2} and k_{-2}, respectively. In the expanded model, this reaction is converted to a reaction rule that specifies that a molecule of EGFR that is bound to EGF but has a free dimerization site can bind to another EGFR molecule meeting the same criteria with the same forward and reverse rate constants as in the original model. Because EGFR has several additional binding and phosphorylation sites, this rule applies to a large number of EGFR-containing species and generates a large number of potential dimerization reactions. In addition, this rule lifts an assumption implicit in the original model that dimers of EGFR could only break up if both receptors were unphosphorylated. This assumption precluded the existence of phosphorylated EGFR monomers, which have been shown to be significant contributors to signaling through EGFR under certain conditions (*89*).

For the same values of the rate parameters, the original and expanded models were found to predict nearly identical behavior for the outputs examined by Kholodenko *et al*. (*54*), but the expanded model makes new predictions that can be tested experimentally. For example, the expanded model predicts different phosphorylation kinetics for the two primary EGFR tyrosines included in the model. Recent proteomic measurements indicate that different phosphorylation sites on EGFR exhibit different kinetics and that the kinetics depend on the interaction partners at the various sites (*14*). These results provide motivation for considering individual tyrosines in models. Further motivation is provided by results indicating that small-molecule kinase inhibitors have different effects at different tyrosines (*90*). In summary, recasting a heuristic model of a signaling pathway increased the computational complexity, but did not require additional model parameters and also facilitated understanding of the model by unveiling hidden assumptions in the original model. The rule-based model better reflects our understanding of protein-protein interactions, is more easily extendable, and generates additional predictions.

Model Simulation, Reduction, and Checking

In addition to model specification, a number of other problems arise from the large number of possible reactions in a signal-transduction system. In this section, we briefly review how rules have been used to improve the efficiency of procedures for simulating the dynamics of a system. We also discuss systematic methods for reducing a model to a simpler one and for verifying that the dynamical behavior of a model is consistent with specified properties.

Simulation. The cost of simulating the kinetics of a reaction network depends on the size of the network. The exact scaling of cost with network size depends on a number of factors and varies from problem to problem, but the scaling is nonlinear for practical methods. Thus, if a network is large, the simulation cost can be expensive in terms of computation time and/or computer memory. To address this problem, Lok and Brent (*62*) proposed a method of simulation that takes advantage of rules, the on-the-fly method mentioned earlier, which is similar to approaches that have been used to simulate chemical systems (*91*–*93*). In this method, rule evaluation is embedded in a discrete-event Monte Carlo simulation of reaction kinetics, and reactions are generated only when a species is first populated during a simulation. Thus, parts of a network unconnected to populated species are ignored, which may speed calculations and reduce memory requirements. This technique of on-the-fly network generation provides a practical benefit only when the populated fraction of a network is sufficiently small and branching is limited. For a highly branched network, the cost of updating the list of reactions connected to populated species can be an important factor even if few of the possible species are populated (Fig. 7).

The problem of simulating a highly branched network is addressed by another Monte Carlo method that takes advantage of rules (*69*, *94*) and is implemented in the StochSim software package (*73*). In the StochSim algorithm, rules are used during a simulation to generate discrete reaction events (that is, to execute reactions), rather than to generate reactions. A list of reactions is unnecessary, and the cost of generating one is avoided. Thus, the StochSim algorithm may provide a computational cost savings when many reactions are possible but only a fraction actually occur. However, its applicability is limited, because StochSim rules explicitly represent only state changes of proteins. Representing changes of connectivity can be problematic. Thus, for example, StochSim cannot be used in a straightforward way to model the system of Fig. 6.

New simulation capabilities are needed for rule-based modeling, because reaction networks derived from rules are unusually large, which is problematic for conventional simulation procedures. It may be possible to develop additional methods, like those mentioned above, that take advantage of the underlying rules of a rule-derived reaction network to speed simulation. For example, it may be possible to extend on-the-fly methodology from stochastic simulations to ODE-based simulations (*95*, *96*), which would likely be more efficient. Current on-the-fly methods essentially consider a species or set of species to be relevant if it is populated by just one molecule, which triggers reaction generation. It would be desirable to refine these methods to incorporate more stringent definitions of relevance such that the trigger for reaction generation can be tuned for efficiency. Finally, we wonder whether the method of StochSim could be made to work with more general rules or could be implemented in hardware (*97*, *98*). A general conclusion we draw from our experiences with simulation to date is that, because there are trade-offs involved in all of the methods that have been developed, it is highly desirable for a software platform to provide access to multiple simulation approaches through the same interface.

Model reduction. Another way to address the issue of combinatorial complexity is through model reduction. Some researchers have considered ways to obtain a model that is smaller than one needed to represent all microscopic details, yet behaviorally equivalent to it (*42*, *87*, *99*–*101*). For example, Borisov *et al*. (*99*) suggested lumping microscopic species of a model together into new variables to obtain an equivalent dynamical model in terms of coarse-grained variables. Recently, Conzelmann *et al*. (*101*) have developed a systematic method for obtaining reduced models in this way. Exact time courses of the macroscopic variables are ob-

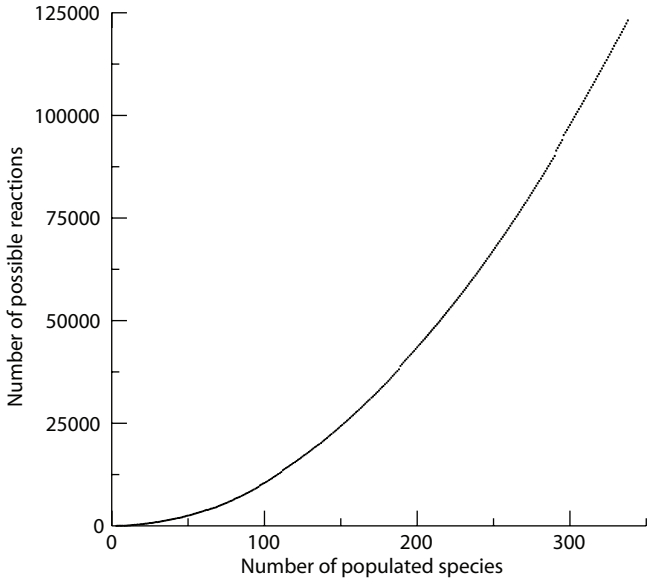


Fig. 7. The number of potential reactions versus the number of populated species when trivalent ligands interact with bivalent cell surface receptors. These results are generated through on-the-fly network generation with BioNetGen. Reactions are generated by six rules, which are similar to the rule shown in Fig. 6. The parameters of the simulation are as follows. The total ligand and receptor concentrations are 7 and 0.5 nM, respectively. Rate constants associated with elementary rate laws for three of the rules are k_{+1} = 0.0004 $nM^{-1}s^{-1}$, $k_{+2} = k_{+3}$ = 0.04 $nM^{-1}s^{-1}$, where k_{+1} is the rate constant for binding of a receptor site to a site on a free ligand, k_{+2} is the rate constant for binding of a receptor site to a site on a ligand bound once, and k_{+3} is the rate constant for binding of a receptor site to a site on a ligand bound twice. The corresponding reverse rate constants for the elementary rate laws associated with the three remaining rules are each 0.01 s^{-1}. As can be seen, a few hundred species have the potential to participate in hundreds of thousands of distinct reactions. Thus, simulation methods that rely on reaction generation, including on-the-fly methods, will not be computationally feasible for large-scale networks in some cases.

tained by solving a reduced set of differential equations that includes both the macroscopic variables and possibly a set of auxiliary (mesoscopic) variables. The extent of model compression depends on the degree of independence among the protein sites being modeled, because correlation among sites requires tracking of additional auxiliary variables. A number of factors, such as cooperative binding or enzyme-substrate relations within a complex, can give rise to coupling among sites. For example, the phosphorylation states of two sites in a receptor are correlated if binding of a cytosolic PTK to one site depends on phosphorylation of the site, and the PTK, once recruited to the receptor, catalyzes phosphorylation of the second site. The variable transformation approach has so far only been applied to the case of a single scaffold protein containing multiple sites of modification and binding. More work is needed to extend the approach to more complex situations. In addition, it would be desirable to develop a systematic method for developing approximate reduced models (for example, by dropping some of the auxiliary variables), which might be helpful when exact compression is not possible because of correlations between sites (*100*). The problem of model reduction through aggregation of variables is as difficult as it is necessary, in particular for systems that operate on several spatial and temporal scales.

Model checking. Model checking (*102*) refers to a suite of techniques deployed to prove that a model of a system behaves in a specified manner. The techniques originate in computer science and engineering, where they are also known as program verification. For example, an objective of program verification is to prove that a highly complex piece of software meets design requirements. A useful step in this direction is to construct a simplified program (a model) that preserves essential design characteristics, and then proving (or disproving) its compliance with specifications. This is typically done before the actual piece of software is built, precisely to avoid early design flaws. In the present biological context, the "simplified program that preserves essential design characteristics" is a rule-based model of a signal-transduction system. Various researchers have sought to apply the powerful tools developed in the context of program verification to biological models, particularly when they can be cast in the form of rules that are executed stochastically on the fly. Consider a volume in which molecular species collide randomly and react when an appropriate rule becomes applicable, much as in STOCHSIM described above. Unlike the unique deterministic trajectories of ODEs, a stochastic simulation will only yield a particular history at each run. Even when many runs are gathered to perform a statistical analysis, observing a time series of concentrations (or molecule numbers) does not necessarily lead to an understanding of the model (and the system it is intended to elucidate). Moreover, a detailed time series may not be the most appropriate representation to match against empirical observations that are qualitative or expressed in terms of salient events. Here is where model checking becomes useful.

The model checking process requires two inputs: (i) a suitable representation of the signal-transduction model and (ii) a property whose truth within the model we wish to check. In the process, a rule-based reaction network is translated (automatically) into an automaton. An automaton is a graph whose nodes represent global states of the system (a global state is determined by the numbers of all molecular species) and whose edges represent transitions between states due to possible reactions. An automaton so defined represents, at once, all possible trajectories the system can take and may therefore be very large, even infinite. Yet, it does not have to be held in computer memory in its entirety. As with on-the-fly methods described previously, the automaton graph can be expanded when needed by computing the partial graph of accessible states (nodes) from the current state. As the nodes of the automaton are generated, they are traversed systematically (breadth- or depth-first); at the same time, the automaton is checked in a divide-and-conquer fashion (that is, by recursively breaking the problem down into simpler subproblems) to determine whether the given property holds. This procedure reflects a strategy similar to that implemented in programs that play chess. Most important, when a property fails to hold, the procedure will reveal why the property fails by exhibiting the transition that led to its violation. Model checking is not a simulation, but a clever exhaustive search of the state space of a model. Naturally, model checking has limitations with regard to the type of properties that can be effectively checked.

Model checking requires properties to be expressed as formulas in a logical framework, such as standard first-order logic

(predicate calculus) augmented with operations that capture temporal relations between events ("before," "after"); temporal modalities ("sometimes," "always," "next time," "until"); constraints; and perhaps even structural properties (such as graph connectedness).

The application of model checking to biological systems is in its infancy. At present, biological applications constitute more an exploration of methods than breakthroughs in biological insight. An illustrative example is the analysis of Kohn's map of the cell cycle (*103*). The properties checked for the Kohn map were static assertions, such as, "The activation of CDC25C is necessary to generate the CDK1-cyclin B complex," as well as dynamic ones, such as oscillations in the abundance of cyclin A (which translates into, "There is a path such that whenever cyclin A is present it eventually disappears, and whenever it is absent, it eventually appears"). These formulas do not push the boundaries of biological insight. Yet, they illustrate the potential of queries for consistency checks, behavioral comparisons across models, and compliance with empirical observations.

Rule-based systems may be translated into different types of automata that reflect various levels of resolution. The class of hybrid automata defines a particularly important level. A hybrid automaton represents a mixture of continuous dynamics and discrete events, as in a system that is well described by a differential equation model until a condition is reached that triggers a discrete event altering the dynamics. For example, the net rate of production of a protein may switch from negative to positive when the concentration of another protein falls below a threshold value. In the class of piecewise affine systems, the dynamical equations are linear with an offset. Such automata can serve as qualitative approximations of complicated nonlinear reaction systems. They effectively transform a nonlinear system into a linear one, but with a structure that can be switched. This type of model yields a representation of a reaction network that facilitates model checking. Batt *et al*. (*104*) provide an interesting and predictive application of model checking to a hybrid-automaton representation of the nutritional stress response in *Escherichia coli*. The queries have illuminated the role of mutual inhibition between the transcription factors FIS and CRP, which are involved in regulating metabolism. Mishra and co-workers have applied hybrid automata techniques and model checking to a stylized Delta-Notch signaling circuit (*105*). We expect to see increasing activity in this area.

Software for Rule-Based Modeling

General-purpose software is needed to enable rapid rule-based modeling of diverse systems. Problem-specific programs can be difficult to reproduce independently, and the time and effort required to write and debug such a program can be expensive. Also, modifications of rules may require reprogramming, which discourages the consideration of alternative hypotheses about mechanisms of signaling or extensions of a model. In recent years, various general-purpose software tools and corresponding methods based on rules for protein-protein interactions have been developed. Software capabilities and methodology are advancing quickly, and it seems likely that rule-based modeling will soon be a practical, mature, and accessible method of computational systems biology. Below, we discuss several specific software tools and end with a short summary of the lessons learned from the development of these tools.

STOCHSIM 1.4. The salient feature of this tool (ftp://ftp.cds.caltech.edu/ pub/dbray/) is the representation of proteins as computational objects or agents, which have states (*73*). Model specification is accomplished by using a wizard-like GUI, which helps a user create a set of formatted input files defining proteins, their states, individual reactions, and rules for generalized reactions that change the states of proteins. A protein is represented by a list of binary flags encoding the modification or binding states of its sites. In the state-change rules of STOCHSIM, state flags are specified as 0, 1, or "?," which is a wildcard, and plus (+) and minus (–) symbols are used to direct the switching of flags from 0 to 1 and vice versa. A limitation of STOCHSIM is its representation of complexes. A complex is either represented implicitly in terms of protein states, in which case the topology of a complex may be difficult to track, or it is represented explicitly, in which case the molecular composition must be declared manually beforehand in an input file.

An interesting feature of STOCHSIM is its method for discrete-event Monte Carlo simulation of reaction kinetics. The method involves picking two agents and then checking to see whether they react. The method is approximate because time is incremented in fixed steps (*69*, *94*). Thus, a check for accuracy, against the results of exact Monte Carlo methods (*80*, *81*) or a STOCHSIM run with a smaller time step, is necessary. The procedure may be slow because, for correct results, the time step must be small enough such that pairs of reactants most often do not react. However, as discussed above, the method avoids generating reactions, which can be an advantage essential for computability in some situations. Shimizu and Bray (*94*) discuss the scenario of a protein with n sites of modification and m binding partners. This protein has up to 2^n modified forms and may participate in as many as $m2^n$ distinct bimolecular association reactions. For such cases and cases like that of Fig. 7, methods that avoid the cost of reaction generation, such as the STOCHSIM method, may be essential for computability.

Moleculizer 1.0. As with STOCHSIM, a set of input files is used to specify a model in Moleculizer 1.0 (http://www.molsci.org/~lok/moleculizer/) (*62*); these files are written in XML. Template input files are provided that correspond to a fixed set of rule types (reaction generators), which are coded as separate software modules. The templates are used to define individual rules and their parameters, such as rate constants. Although the types of rules available to a modeler are fixed, they provide sufficient flexibility to represent a wide array of protein-protein interactions. A feature of Moleculizer, not available in STOCHSIM, is the ability to represent the topology of a protein complex explicitly in terms of pairwise connections of binding sites. Also, the connectivity of a complex can be considered in a rule definition.

As can STOCHSIM, Moleculizer can be used to perform a discrete-event Monte Carlo simulation of reaction kinetics, but an exact method, Gillespie's direct method (*80*, *81*), is used. This method relies on a list of reactions, which Moleculizer generates on the fly during a simulation. When a species is first populated, rules are evaluated and new reactions involving this species are generated if necessary (that is, if they are not already generated and stored in memory). Given the parameters that govern a simulation of network dynamics, Moleculizer provides a principled means for identifying the relevant portion of the reaction network (that is, the populated species and reactions connected to these species). Moleculizer provides other

simulation methods as well, such as a τ-leap method for stochastic simulations (*106*). A new simulation feature is "look ahead" rounds of reaction generation (*107*). If the number of look-ahead rounds is set to a large enough number, then Moleculizer can generate a network without simulating it. This method is then equivalent to the generate-first method mentioned earlier.

BioNetGen 2.0 and BNGL. Earlier versions of this tool (1.0 and 1.1), BioNetGen (http://cellsignaling.lanl.gov/bionetgen/), which are based on the use of term or string rewriting rules to represent protein-protein interactions, are discussed elsewhere (*60*, *88*). We will focus on version 2.0, which is based on the use of graphs to represent proteins and protein complexes and the use of graph rewriting rules to represent protein-protein interactions (*50*, *61*). For historical perspective and a review of applications of graph transformation in molecular biology, see Rosselló and Valiente (*108*). Graph rewriting rules are generalizations of term rewriting rules. Unlike STOCHSIM and Moleculizer, BioNetGen interprets a formal model-specification language, called the BioNetGen Language (BNGL). An advantage of introducing a language is that model specification becomes independent of software implementations and therefore portable. In subsequent sections, we will discuss several languages for rule-based modeling that have been proposed. Interesting languages that we will not discuss include Bioglyphics (http://www.bioglyphics.org/) (*109*), Dynamical Grammars (http://www.arxiv.org/abs/ cs.AI/0511073), and "little b" (http://www.littleb.org/).

In BNGL, chemical species are represented by using graphs, which are encoded as structured strings. The basic unit of representation is a molecule string, which may contain component strings enclosed in parentheses. In a model specification, molecule strings actually serve several purposes, one of which is definition of the types of molecules included in a model. An example of a string used for this purpose is EGFR(ECD,aa1092~Y~pY), which defines a molecule called EGFR containing components called ECD (ectodomain) and aa1092. The string also indicates, by the prefix "~" for state labels, that the latter component has two possible internal states called Y and pY. This definition corresponds to one type of molecule represented in the reaction network of Fig. 5B, which also shows how molecule strings are concatenated to represent complexes.

In BNGL, rules for protein-protein interactions are represented by using text-encoded graph rewriting rules, which contain graph matching patterns. Patterns composed of (incomplete) molecule strings explicitly indicate the features required of reactants and implicitly indicate, through omission, the features that are irrelevant for a reaction. An example of a rule is Grb2(SH2)+EGFR(aa1092~pY) → Grb2(SH2!1).EGFR(aa1092~pY!1). This rule indicates that the adaptor protein Grb2 and EGFR can associate if the SH2 domain of Grb2 is free and residue 1092 in EGFR is free and phosphorylated. By omission, the rule also indicates that association of Grb2 and EGFR is independent of the bound state of the ECD of EGFR. The "!" character is a prefix for bond labels, which indicate how components of proteins are connected in a complex. The bond labels in this rule indicate that the SH2 domain of Grb2 binds pY1092 in EGFR. The rule for Grb2-EGFR association and others are illustrated in Fig. 8 by using the conventions of Faeder *et al*. (*50*).

A model is specified as a set of graph-rewriting rules, which are associated with rate laws, and a set of seed-species graphs to which the rules are initially applied. Using the iterative algorithm for processing rules outlined above (*60*, *61*), BioNetGen can generate a parameterized reaction network (meaning that the reactions are assigned rate laws) without performing a simulation of the network dynamics. Once a network has been generated, it can serve as the basis for ODE-based or stochastic simulations (*60*). Pregenerating a network and then performing an ODE-based simulation, if possible, may be more efficient than the stochastic simulation methods of STOCHSIM and Moleculizer, because ODE-based calculations are usually less expensive than Monte Carlo calculations for a "small" model (*63*). If the size of a network is too large to permit ODE-based calculations, then BioNetGen also provides the capability to generate a network on the fly during a simulation, as Moleculizer does.

One problem that BioNetGen attempts to solve is how to automatically adjust the rate law of a rule to account for contextual differences among reactions in the class of reactions defined by the rule. Contextual effects on rate laws can be speci-

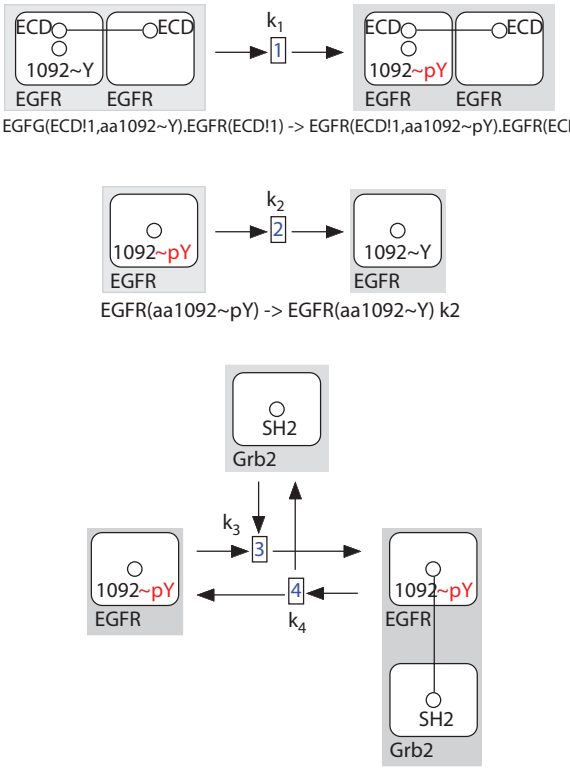


Fig. 8. Representation of protein-protein interactions by using graphical reaction rules. The reactions illustrated in Figs. 4 and 5 can be generated by rules for the underlying protein-protein interactions, which are illustrated here with the conventions of Faeder *et al*. (*50*) and Blinov *et al*. (*61*). The rules define generalized reactions for 1, autophosphorylation of EGFR; 2, dephosphorylation of EGFR mediated by a phosphatase assumed to be present in excess; 3, association of Grb2 and EGFR, which depends on phosphorylation of Y1092; and 4, dissociation of Grb2 and EGFR. Below each graph-rewriting rule, a corresponding definition in BNGL is given.

fied explicitly in an expanded set of rules, but this approach may be laborious, and the problem arises often enough that general automatic solutions merit pursuit. The contextual factors that may necessitate modifications of a rate law for a class of reactions are numerous: (i) Collision frequency can vary (compare A + A→ and A + B→ reactions); (ii) reaction path degeneracy (that is, the number of distinct reaction paths from reactants to products) can vary, giving rise to different statistical factors (compare A.A→A.B and A→B reactions); (iii) the turnover frequency of an enzymatic reaction may depend on the numbers of enzymes and substrates in a complex; (iv) a factor equal to a volume ratio may arise for reactions in different compartments; (v) diffusion-limited reactions may be affected by masses (or equivalently, diffusivities) of reactants (*60*, *62*); and so on. BioNetGen handles statistical factors, for example, by considering the symmetries of graphs in reaction rules and the reactions generated by rules (*61*). Assigning correct statistical factors is essential for a self-consistent model. Figure 6C illustrates reactions of the same class that have different statistical factors.

BioSPI and other tools of process algebra. BioSPI interprets a model-specification language, a variant of the π-calculus, and provides the capability to perform a stochastic simulation (http://www.wisdom.weizmann.ac.il/~biopsi/) (*110*). Similar tools, which have also been used to specify models for biological systems, include SPiM (*111*) and the PEPA Workbench (*112*). The π-calculus is a general and minimalist model-specification language originally designed to capture essential features of concurrent and distributed systems in computer science (*113*). It is one of many process calculi (algebras) studied in this field (*114*, *115*). These languages are axiomatic, which allows properties of a model specification to be formally proven and facilitates certain types of model checking (for example, the observational equivalence of two model specifications can be determined).

Use of π-calculus to model protein-protein interactions was suggested by Regev *et al*. (*116*), who provided, loosely speaking, the following translations of biochemistry to π. Proteins are mobile processes (meaning agents that exchange messages that can affect their behavior); protein sites are communication channels (ports through which messages are passed from senders to receivers); and protein-protein interactions are communications. The channels of a system and the messages that can be sent and received along these channels are essentially rules for protein-protein interactions. The proteins in a complex are linked by a backbone (that is, by colocation in a communication compartment), and the topology of a complex is indicated by pair-wise communications. In this framework, signal transduction can be viewed as asynchronous concurrent computation and studied by using pertinent methods from computer science (*117*–*119*).

Because π-calculus is a minimalist language and some of its notational features are irrelevant and even inappropriate for biological applications (for example, communication has directionality, whereas physical association does not), researchers have sought to develop a more congruent higher-level language for modeling biochemical systems that retains the mathematical formality of π. One proposal is κ-calculus (*120*). In this language, a protein is represented as a collection of sites, which can be visualized as a box with nodes, representing sites, on the border. Coincidentally, the appearance of such a box is much like that of a state node in a process diagram (*49*). Also, protein complexes are represented as graphs, in which edges connect interacting sites of proteins. For simulation purposes, the κ-calculus can be reduced to π-calculus.

Although process algebra has been used to specify biological processes in some detail, such as cell adhesion (*121*, *122*), it is unclear at this time whether methods of process algebra can reach fundamentally beyond the level of understanding provided by ODEs and stochastic simulations. Process algebra may, however, provide a formal foundation for rule-based modeling that enables principled mathematical reasoning about system behavior and its dependency on interaction capabilities of system components.

Pathway Logic Assistant. Pathway Logic Assistant provides an interface to "Pathway Logic" models of signal-transduction systems (http://www.csl. sri.com/users/clt/PLweb/pl.html) (*123*) defined by using the Maude specification language (http://maude.cs.uiuc.edu/). Pathway Logic models can formally represent protein-protein interactions at different levels of resolution, ranging from details about abstract protein states (*124*) to details about functional domains and binding sites (*125*, *126*). Eker *et al*. (*127*) proposed conventions for using Maude to specify term rewriting rules for protein-protein interactions. Later, conventions for specifying graphs and graph rewriting rules were introduced (*126*), which allow the topology of a protein complex to be explicitly considered. The models obtained with Maude are essentially unparameterized chemical reaction networks, meaning that rate laws are not specified for reactions. With the addition of rate information, a reaction network can be converted to ODEs, for example, but this capability is not native to Pathway Logic Assistant. Instead, this tool, through the formalism of Petri nets, enables a qualitative analysis of a reaction network, such as visualization and interactive exploration of the network. Also, a modeler can specify a formula in linear temporal logic (LTL) that defines a putative property, and then evaluate this formula in a model-checking query to determine whether the property holds for the system. Properties that can be specified in LTL are qualitative and relate to stable states of paths. For example, a query can be specified to determine whether a particular protein complex is potentially present in a signal-transduction system and to find a sequence of reactions leading to it. Of course, the value of qualitative analysis is limited, as the dynamics of protein-protein interactions are important for system behavior, for example, dynamics influence which protein complexes are populated during signaling (*42*). Still, it will be interesting to see what biological insights can be obtained from qualitative analysis alone, because quantitative rate parameters for protein-protein interactions are often unavailable or uncertain.

BIOCHAM 2.4. This tool, BIOCHAM 2.4, interprets a model-specification language in which term-rewriting rules are used to represent protein-protein interactions (http://contraintes.inria.fr/BIOCHAM/) (*128*). Structured strings are used to represent chemical species. Rules, which may contain string-matching patterns, indicate how strings representing reactants are rewritten to obtain strings representing products. As is also true for other term-rewriting approaches (*60*, *88*, *127*), the topology of a protein complex cannot be explicitly considered in a rule, which can be a limitation. Unlike the π-calculus or the general Maude programming language, but like BNGL, the κ-calculus, and Pathway Logic, the BIOCHAM language has been de-

signed specifically for the purpose of modeling biological systems. The language is simple by design but still expressive enough to specify meaningful models. For example, a model has been specified for the cell cycle–control system described by Kohn (*32*) (*103*).

Like BioNetGen, BIOCHAM can process a set of rules associated with rate laws to generate a parameterized reaction network and the corresponding ODEs, and then perform and analyze ODE-based simulations. In addition, like Pathway Logic Assistant, BIOCHAM can evaluate model-checking queries, which can be specified by using Computation Tree Logic (CTL) as well as LTL. CTL and LTL are closely related; however, CTL queries can be specified that cannot be expressed in LTL and vice versa (*129*). Model checking in the framework of BIOCHAM is provided through an interface to the NuSMV model checker (http://nusmv.irst.itc.it/). BIOCHAM essentially converts a chemical reaction network to a NuSMV input file. Thus, model-checking capabilities are likewise accessible, in principle, from any rule-based modeling tool that generates a reaction network. An intriguing feature of BIOCHAM is a machine-learning system for modifying rules or parameters to obtain consistency with behavioral constraints specified in CTL or LTL (*130*).

Lessons learned. Many of the software tools discussed above are based on a formal model-specification language, which has the desirable property of being application-independent. A standard language has yet to emerge, but the idea of using rewriting rules to represent protein-protein interactions has been suggested multiple times. Likewise, in several approaches, graphs are used to represent the contact map of protein complexes. One needs to consider the configuration of a complex when connectivity affects reactivity, as exemplified in Fig. 9. Chemical reaction network models derived from rules tend to be large. It may therefore be nontrivial to confirm that the model produced by a set of rules is the one intended. A high standard of software reliability, including careful consideration of physical chemistry, is essential. To cope with the complexity of models, we need software tools that assist in reasoning about a model and that provide means for automatically monitoring system properties at a quantitative and qualitative level. For example, BioNetGen provides output rules for calculating properties of sets of species, such as a sum of concentrations of species containing a particular protein. Finally, rules can be used in different ways to study a system. The different approaches now available appear to be complementary, and new approaches may be useful.

Extension of SBML for Rules

The Systems Biology Markup Language (SBML) (*131*), which is based on XML, is a popular standardized format for the electronic storage, exchange, and reuse of mathematical models of biochemical systems. A number of software packages are now available that import or export models specified in SBML (http://sbml.org/), and a public repository for annotated SBML-encoded models has been established (*132*, *133*). However, SBML, like most SBML-aware software, is based on the assumption that a model can be specified adequately in terms of a reaction scheme, which is not likely to hold for a model of a signal-transduction system because of the context-sensitivity and combinatorial complexity of protein-protein interactions. Versions of SBML up through Level 2 Version 2, which is presently being finalized, do not provide direct support for compactly representing multiple protein complexes or multiple states of proteins. Likewise, there is no support for defining rules for protein-protein interactions or other generalized reactions. To specify a model in strict SBML, one must enumerate all of the individual chemical species and reactions included in the model. Rules used to derive these species and reactions could actually be included in a SBML file as annotation, but such annotation would be nonstandard. Thus, for interpretability, SBML can only be used to represent a rule-based model as a list of rule-derived reactions, which means that SBML encodings of rule-based models tend to be exceedingly verbose and difficult to comprehend or modify (*50*, *88*). In light of this limitation, several extensions of SBML that incorporate rules have been proposed

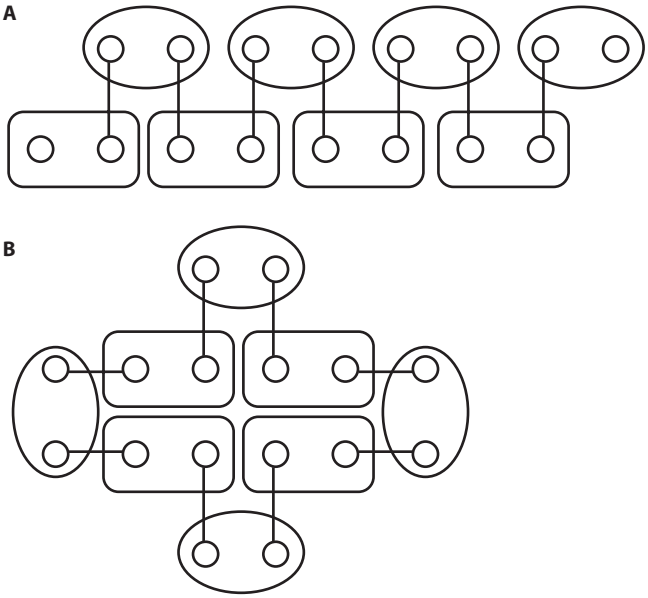


Fig. 9. Two protein complexes with identical composition but different connectivity. (**A**) A chain of bivalent ligands, represented as ovals, and bivalent receptors, represented as boxes with rounded corners. Circles represent ligand-binding and receptor-binding sites, and lines connect sites that are bound to each other. (**B**) The ring formed through closure of the chain. The reactivities of the chain and ring differ. The chain can close through intramolecular binding of ligand and receptor sites at the chain ends or elongate by binding either a ligand or receptor with a free site. In contrast, the ring cannot react with ligands or receptors. It can only break apart through the opening of one of its ligand-receptor bonds. Thus, tracking the connectivity of proteins in a complex can be important for modeling protein-protein interactions.

(*1342–137*). It is anticipated that such extensions will be available with the SBML update to Level 3 (*138*, *139*).

SBML Level 3 is anticipated to introduce a modular language extension capability, which will allow different language features to be added to a common language core, which will be based on SBML Level 2. Language extensions will add syntax and semantics for software tools that share a common theme, such as the tools for rule-based modeling reviewed here. Models specified in SBML Level 3 will include a declaration of the feature sets (beyond the core Level 3 features) required for proper interpretation. The presence of a feature tag will inform a

software tool reading the model that the model uses that particular feature and will permit the tool to quit gracefully if it does not have the necessary interpretive capabilities.

The latest proposed Level 3 extensions for rule-based modeling (*136*, *137*), like several of the software tools reviewed here, are based on the concept of using graphs to represent proteins and graph-rewriting rules to represent protein-protein interactions. As discussed above, such rules are expressive enough to convey information about the contextual constraints on a protein-protein interaction and the parts of proteins responsible for and affected by an interaction. The same is not true of the standardized exchange format widely used today to store information about protein-protein interactions in electronic databases (*140*). Thus, extension of SBML to incorporate rules for protein-protein interactions has the potential to improve not only our ability to model the dynamics of protein-protein interactions, but also our ability to archive basic knowledge about them. Rules for large numbers of protein interactions might be specified on the basis of high-throughput assays (*141*), data extracted with literature-mining tools (*142*), or statistical analysis of large data sets (*143*). An important next step in the development of SBML is software for converting a putative SBML Level 3 specification of a rule-based model into Level 2 form.

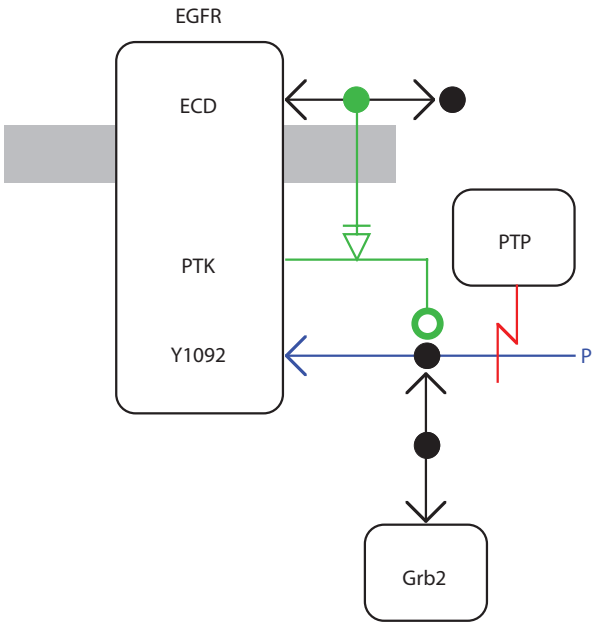


Fig. 10. A diagrammatic interaction map drawn according to the conventions of Kohn *et al*. (*145*). This map illustrates protein-protein interactions underlying the reaction scheme illustrated in Figs. 4 and 5.

Visualization of Rules

A mathematical model provides at least two benefits. Its implications, no matter how hidden from intuition, can usually be revealed through computer-aided analysis. A model also constitutes a precise statement of a modeler's understanding and assumptions about a system. Thus, it is a powerful tool for summarizing knowledge about a system, although a model may be difficult to comprehend. If it is represented as a list of equations or reactions, the formalism of the model obscures the underlying grammar of protein interactions that generates the chemical species and reactions included in the model. A widely used and more comprehensible way of representing knowledge about a signal-transduction system is a diagrammatic interaction map in which proteins and their interactions (or the functional consequences of these interactions) are indicated by labeled cartoons and arrows. Most maps are ambiguous and lack a mathematical interpretation. However, formal conventions have been proposed for drawing interaction maps such that they have precise meanings (*144*, *145*). An example of a simple map drawn according to these conventions is provided in Fig. 10. It illustrates interactions of EGFR and Grb2. No automatic procedures for translating this type of map into a mathematical model are yet available (*145*), although software intended to provide such a capability has been developed on the basis of similar diagrammatic formalisms (*146*–*149*). Nevertheless, automatically translating a map like that of Fig. 10 into a mathematical model remains an open problem when combinatorial complexity is present.

An alternative to representing a model as a list of reactions or equations is to express a model as a set of rules, which resemble reactions in many ways but are highly compressive. If the rules are graph-based, then they can also be visualized, and therefore, they provide a new means of visualizing knowledge of protein-protein interactions. An example of rules that represent interactions considered in Fig. 10 is shown in Fig. 8. Similar illustrations would be obtained by using any of the other graph-based model specification approaches discussed above. An advantage of visual rules over interaction maps is that the rules can be evaluated automatically to obtain a mathematical model. When the rules of Fig. 8 are applied to a seed set of species composed of unphosphorylated dimeric EGFR and cytosolic Grb2, they generate the reaction network illustrated in Figs. 4 and 5. Rules can also be made more readable by enhancing them with graphical notation developed for annotating protein-protein interactions (*49, 53*) (http://www.sbgn.org/).

Outlook

New technologies are enabling an increasingly quantitative characterization of signal transduction events at the level of protein sites. The resultant data streams challenge our capacity to interpret on the basis of intuition alone. Understanding the dynamics of signaling systems hinges on appropriate mathematical methods and models. A mathematical model tells a story that converts facts into putative knowledge by animating facts with dynamical behavior, making assumptions explicit, and enabling formal manipulations whose correctness can be assessed. Presumed knowledge generated in this way is more useful, predictive, and empirically testable. A model is a bookkeeping device that helps us to assess whether we know enough about a system.

The multiplicity of sites at which a signaling protein can be modified and a protein's potential for assembling into diverse complexes result in a combinatorial explosion of molecular states that stymies conventional modeling approaches. Models of signal-transduction therefore require software tools grounded in a mathematical framework that represents proteins as grammatical structures with specific context-dependent actions. Well-known challenges to modeling include the problem of

compositionality, that is, the need for a framework for growing models in a cumulative fashion by plugging together independently developed models of subsystems. A different compositionality problem arises from the need to juxtapose systems that operate at different spatial and temporal scales, or at vastly different molecular densities. Some of these challenges are beginning to be addressed with the introduction of rules for protein-protein interactions and the development of methods for rule-based modeling, like those reviewed here.

Rule-based modeling is a fundamental departure from traditional dynamical systems based on ordinary differential equations (*117*). Dynamical systems require all possible interactions to be explicitly specified ahead of time. In contrast, rule-based approaches can be used to spawn reactions and molecular states only if the dynamics of the system generates the appropriate context (on-the-fly operation). Unlike differential equations, rule-based approaches represent molecules with an internal structure that encodes potential behavior that may never be triggered in a given system but that could be triggered if the system were changed. This aspect of rules is what makes rule-based approaches extensible on the go, much like dropping a new molecule into a reaction mixture alters the intrinsic possibilities of that mixture. (Note that a modeler should avoid specifying rules that are overly general, because such rules may generate unintended reactions when a model is extended.) Rule-based approaches do not jettison differential equations. In their simplest mode of operation, they enable the automatic generation of ODE systems that are too large to be written down by hand. Although rules have been used to formalize interactions among objects of various types for over two decades (*150*), they only recently have been applied to signal-transduction systems.

General-purpose tools for rule-based modeling would benefit from the adoption of a standard specification language for defining models in software-independent ways. SBML now serves as such a standard, but it conceives of a model as an exhaustive list of reactions. Because this view fails to recognize the linguistic character of complex proteins and protein complexes, it becomes quickly impractical. In response to this shortcoming, extensions of SBML consider the adoption of rules for representing protein interactions. Much remains to be done to devise languages capable of expressing more fine-grained aspects of protein structure and contextual constraints, including those of a geometric kind. Early examples of capturing spatial effects in a rule-based fashion were phenomenological models of morphogenesis (*151*) and models of virus capsid assembly (*152*, *153*).

There is a clear need for making the modeling process more accessible to nonspecialists. Likewise, it is highly desirable to invite exploration of the possible, such as neighboring signaling networks that are accessible from actual ones by a few shuffling operations. Deriving mechanistic network models from empirical observations, evolving or designing networks to behave in specified ways (*154*), and predicting the consequences of interventions are all tasks that depend on efficient ways of exploring alternative network models. These objectives require the partial automation of the modeling process. Such automation would be critically enabled by the adoption of a formal language to express properties of proteins and their interactions or properties of whole systems, as extracted from models (via model checking) or empirical observations. This language could be used to annotate the behavior of a protein or a small system and could be stored alongside other information in databases. A statement in a formal language has a well-defined semantics, that is, a clear and precise interpretation, which could perhaps be quickly grasped through visualization. At the same time, such a statement would lead a double life as a codelet, a fragment of computer code, which can be downloaded and used as a component in a model of any system in which the corresponding protein or protein-protein interaction plays a role. Formal statements about system behavior (originating in empirical observations or trusted models) could be used as test-suites for new models or refinements. We believe the progress in rule-based modeling summarized in this review may lead ultimately to a common language for systems biology, much like the four-letter code for DNA sequences provides a common language for bioinformatics.

References and Notes

1. C. Sawyers, Targeted cancer therapy. *Nature* **432**, 294–297 (2004).
2. E. C. Butcher, E. L. Berg, E. J. Kunkel, Systems biology in drug discovery. *Nat. Biotechnol.* **22**, 1253–1259 (2004).
3. P. Rajasethupathy, S. J. Vayttaden, U. S. Bhalla, Systems modeling: A pathway to drug discovery. *Curr. Opin. Chem. Biol.* **9**, 400–406 (2005).
4. T. Hunter, Signaling—2000 and beyond. *Cell* **100**, 113–127 (2000).
5. T. Pawson, Specificity in signal transduction: From phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* **116**, 191–203 (2004).
6. N. J. Eungdamrong, R. Iyengar, Computational approaches for modeling regulatory cellular networks. *Trends Cell Biol.* **14**, 661–669 (2004).
7. J. A. Papin, T. Hunter, B. O. Palsson, S. Subramaniam, Reconstruction of cellular signaling networks and analysis of their properties. *Nat. Rev. Mol. Cell Biol.* **6**, 99–111 (2005).
8. H. S. Wiley, S. Y. Shvartsman, D. A. Lauffenburger, Computational modeling of the EGF-receptor system: A paradigm for systems biology. *Trends Cell Biol.* **13**, 43–50 (2003).
9. B. Goldstein, J. R. Faeder, W. S. Hlavacek, Mathematical and computational models of immune-receptor signaling. *Nat. Rev. Immunol.* **4**, 445–456 (2004).
10. R. Breitling, D. Hoeller, Current challenges in quantitative modeling of epidermal growth factor signaling. *FEBS Lett.* **579**, 6289–6294 (2005).
11. W. S. Hlavacek, J. R. Faeder, M. L. Blinov, A. S. Perelson, B. Goldstein, The complexity of complexes in signal transduction. *Biotechnol. Bioeng.* **84**, 783–794 (2003).
12. J. C. D. Houtman, M. Barda-Saad, L. E. Samelson, Examining multiprotein signaling complexes from all angles. *FEBS J.* **272**, 5426–5435 (2005).
13. C. Witt, S. Raychaudhuri, A. K. Chakraborty, Movies, measurement, and modeling: The three Ms of mechanistic immunology. *J. Exp. Med.* **201**, 501–504 (2005).
14. W. X. Schulze, L. Deng, M. Mann, Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol. Syst. Biol.* doi:10.1038/msb4100012 (2005).
15. Y. Zhang, A. Wolf-Yadlin, P. L. Ross, D. J. Pappin, J. Rush, D. A. Lauffenburger, F. M. White, Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics* **4**, 1240–1250 (2005).
16. R. B. Jones, A. Gordus, J. A. Krall, G. MacBeath, A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439**, 168–174 (2006).
17. C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto Jr., Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* **44**, 7342–7372 (2005).
18. M. Ptashne, A. Gann, Imposing specificity on kinases. *Science* **299**, 1025–1027 (2003).
19. T. Pawson, P. Nash, Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445–452 (2003).
20. W. A. Lim, The modular logic of signaling proteins: Building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol.* **12**, 61–68 (2002).
21. J. E. Dueber, B. J. Yeh, R. P. Bhattacharyya, W. A. Lim, Rewiring cell signaling: The logic and plasticity of eurkaryotic protein circuitry. *Curr. Opin. Struct. Biol.* **14**, 690–699 (2004).
22. I. Letunic, R. R. Copley, S. Schmidt, F. D. Ciccarelli, T. Doerks, J. Schultz, C. P. Ponting, P. Bork, SMART 4.0: Towards genomic data integration. *Nucleic Acids Res.* **32**, D142–D144 (2004).

23. J. C. Cambier, Antigen and Fc receptor signaling: The awesome power of the immunoreceptor tyrosine-based activation motif (ITAM). *J. Immunol.* **155**, 3281–3285 (1995).
24. P. Puntervoll, R. Linding, C. Gemund, S. Chabanis-Davidson, M. Mattingsdal, S. Cameron, D. M. Martin, G. Ausiello, B. Brannetti, A. Costantini, F. Ferre, V. Maselli, A. Via, G. Cesareni, F. Diella, G. Superti-Furga, L. Wyrwicz, C. Ramu, C. McGuigan, R. Gudavalli, I. Letunic, P. Bork, L. Rychlewski, B. Kuster, M. Helmer-Citterich, W. N. Hunter, R. Aasland, T. J. Gibson, ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* **31**, 3625–3630 (2003).
25. C. Chothia, J. Gough, C. Vogel, S. A. Teichmann, Evolution of the protein repertoire. *Science* **300**, 1701–1703 (2003).
26. C. A. Orengo, J. M. Thornton, Protein families and their evolution—a structural perspective. *Annu. Rev. Biochem.* **74**, 867–900 (2005).
27. X.-J. Yang, Multisite protein modification and intramolecular signaling. *Oncogene* **24**, 1653–1662 (2005).
28. D. Endy, R. Brent, Modelling cellular behaviour. *Nature* **409**, 391–395 (2001).
29. D. Bray, Genomics. Molecular prodigality. *Science* **299**, 1189–1190 (2003).
30. C. Wofsy, C. Torigoe, U. M. Kent, H. Metzger, B. Goldstein, Exploiting the difference between intrinsic and extrinsic kinases: Implications for regulation of signaling by immunoreceptors. *J. Immunol.* **159**, 5984–5992 (1997).
31. C. J. Morton-Firth, "Stochastic simulation of cell signaling pathways," thesis, University of Cambridge (1998).
32. K. W. Kohn, Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell* **10**, 2703–2734 (1999).
33. B. Goldstein, J. R. Faeder, W. S. Hlavacek, M. L. Blinov, A. Redondo, C. Wofsy, Modeling the early signaling events mediated by FcεRI. *Mol. Immunol.* **38**, 1221–1228 (2002).
34. R. N. Jorissen, F. Walker, N. Pouliot, T. P. J. Garrett, C. W. Ward, A. W. Burgess, Epidermal growth factor receptor: Mechanisms of activation and signalling. *Exp. Cell Res.* **284**, 31–53 (2003).
35. E. N. Kersh, A. S. Shaw, P. M. Allen, Fidelity of T cell activation through multistep T cell receptor ζ phosphorylation. *Science* **281**, 572–575 (1998).
36. C. H. Weber, C. Vincenz, The death domain superfamily: A tale of two interfaces? *Trends Biochem. Sci.* **26**, 475–481 (2001).
37. A. Kohl, M. G. Grütter, Fire and death: The pyrin domain joins the death-domain superfamily. *C. R. Biol.* **327**, 1077–1086 (2004).
38. C. H. Weber, C. Vincenz, A docking model of key components of the DISCcomplex: Death domain superfamily interactions redefined. *FEBS Lett.* **492**, 171–176 (2001).
39. H. Qin, S. M. Srinivasula, G. Wu, T. Fernandes-Alnemri, E. S. Alnemri, Y. Shi, Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. *Nature* **399**, 549–557 (1999).
40. T. Xiao, P. Towb, S. A. Wasserman, S. R. Sprang, Three-dimensional structure of a complex between the death domains of Pelle and Tube. *Cell* **99**, 545–555 (1999).
41. J. R. Faeder, W. S. Hlavacek, I. Reischl, M. L. Blinov, H. Metzger, A. Redondo, C. Wofsy, B. Goldstein, Investigation of early events in FcεRI-mediated signaling using a detailed mathematical model. *J. Immunol.* **170**, 3769–3781 (2003).
42. J. R. Faeder, M. L. Blinov, B. Goldstein, W. S. Hlavacek, Combinatorial complexity and dynamical restriction of network flows in signal transduction. *Syst. Biol.* **2**, 5–15 (2005).
43. D. Bray, S. Lay, Computer-based analysis of the binding steps in protein complex formation. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13493–13498 (1997).
44. A. Levchenko, J. Bruck, P. W. Sternberg, Scaffold proteins biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5818–5823 (2000).
45. K. H. Lee, A. R. Dinner, C. Tu, G. Campi, S. Raychaudhuri, R. Varma, T. N. Sims, W. R. Burack, H. Wu, J. Wang, O. Kanagawa, M. Markiewicz, P. M. Allen, M. L. Dustin, A. K. Chakraborty, A. S. Shaw, The immunological synapse balances T cell receptor signaling and degradation. *Science* **302**, 1218–1222 (2003).
46. E. O. Voit, *Computational Analysis of Biochemical Systems* (Cambridge Univ. Press, Cambridge, 2000).
47. K. Oda, Y. Matsuoka, A. Funahashi, H. Kitano, A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.* doi:10.1038/msb4100014 (2005).
48. G. Altan-Bonnet, R. N. Germain, Modeling T cell antigen discrimination based on feedback control of digital ERK responses. *PLoS Biol.* **3**, e356 (2005).
49. H. Kitano, A. Funahashi, Y. Matsuoka, K. Oda, Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* **23**, 961–966 (2005).
50. J. R. Faeder, M. L. Blinov, W. S. Hlavacek, Graphical rule-based representation of signal-transduction networks, in *Proc. 2005 ACM Symp. Appl. Computing*, L. M. Liebrock, Ed. (ACM Press, New York, 2005), pp. 133–140.
51. A. Funahashi, M. Morohashi, H. Kitano, N. Tanimura, CellDesigner: A process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO* **1**, 159–162 (2003).
52. B. M. Slepchenko, J. C. Schaff, I. Macara, L. M. Loew, Quantitative cell biology with the Virtual Cell. *Trends Cell Biol.* **13**, 570–576 (2003).
53. M. L. Blinov, J. Yang, J. R. Faeder, W. S. Hlavacek, Depicting signaling cascades. *Nat. Biotechnol.* **24**, 137–138 (2006).
54. B. N. Kholodenko, O. V. Demin, G. Moehren, J. B. Hoek, Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.* **274**, 30169–30181 (1999).
55. B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, G. Müller, Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.* **20**, 370–375 (2002).
56. H. Resat, J. A. Ewald, D. A. Dixon, H. S. Wiley, An integrated model of epidermal growth factor receptor trafficking and signal transduction. *Biophys. J.* **85**, 730–743 (2003).
57. M. Hatakeyama, S. Kimura, T. Naka, T. Kawasaki, N. Yumoto, M. Ichikawa, J. H. Kim, K. Saito, M. Saeki, M. Shirouzu, S. Yokoyama, A. Konagaya, A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling. *Biochem. J.* **373**, 451–463 (2003).
58. I. V. Maly, H. S. Wiley, D. A. Lauffenburger, Self-organization of polarized cell signaling via autocrine circuits: Computational model analysis. *Biophys. J.* **86**, 10–22 (2004).
59. M. L. Blinov, J. R. Faeder, B. Goldstein, W. S. Hlavacek, A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems* **83**, 136–151 (2006).
60. J. R. Faeder, M. L. Blinov, B. Goldstein, W. S. Hlavacek, Rule-based modeling of biochemical networks. *Complexity* **10**, 22–41 (2005).
61. M. L. Blinov, J. Yang, J. R. Faeder, W. S. Hlavacek, Graph theory for rule-based modeling of biochemical networks, in *Proc. BioCONCUR 2005*, San Francisco, CA, 27 August 2005.
62. L. Lok, R. Brent, Automatic generation of cellular reaction networks with Moleculizer 1.0. *Nat. Biotechnol.* **23**, 131–136 (2005).
63. M. L. Blinov, J. R. Faeder, J. Yang, B. Goldstein, W. S. Hlavacek, 'On-the-fly' or 'generate-first' modeling? *Nat. Biotechnol.* **23**, 1344–1345 (2005).
64. J. S. Parkinson, P. Ames, C. A. Studdert, Collaborative signaling by bacterial chemoreceptors. *Curr. Opin. Microbiol.* **8**, 116–121 (2005).
65. M. D. Baker, P. M. Wolanin, J. B. Stock, Signal transduction in bacterial chemotaxis. *Bioessays* **28**, 9–22 (2006).
66. D. Bray, R. B. Bourret, M. I. Simon, Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Mol. Biol. Cell* **4**, 469–482 (1993).
67. D. Bray, R. B. Bourret, Computer analysis of the binding reactions leading to a transmembrane receptor-linked multiprotein complex involved in bacterial chemotaxis. *Mol. Biol. Cell* **6**, 1367–1380 (1995).
68. S. Lay, D. Bray, A computer program for the analysis of protein complex formation. *Comput. Appl. Biosci.* **13**, 439–444 (1997).
69. C. J. Morton-Firth, D. Bray, Predicting temporal fluctuations in an intracellular signalling pathway. *J. Theor. Biol.* **192**, 117–128 (1998).
70. S. Bhattacharjya, P. Xu, M. Chakrapani, L. Johnston, F. Ni, Polymerization of the SAM domain of MAPKKK Ste11 from the budding yeast: Implications for efficient signaling through the MAPK cascades. *Protein Sci.* **14**, 828–835 (2005).
71. B. E. Shapiro, A. Levchenko, E. M. Meyerowitz, B. J. Wold, E. D. Mjolsness, Cellerator: Extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics* **19**, 677–678 (2003).
72. B. E. Shapiro, A. Levchenko, E. Mjolsness, Automatic model generation for signal transduction with applications to MAP kinase pathways, in *Foundations of Systems Biology*, H. Kitano, Ed. (MIT Press, Cambridge, MA, 2001), chap. 1.
73. N. Le Novère, T. S. Shimizu, STOCHSIM: Modelling of stochastic biomolecular processes. *Bioinformatics* **17**, 575–576 (2001).
74. C. Firth, N. Le Novère, T. Shimizu, *STOCHSIM, The Stochastic Simulator* (manual distributed with version 1.4 of the software, 2003), (ftp://ftp.cds.caltech.edu/pub/dbray).
75. T. S. Shimizu, N. Le Novère, M. D. Levin, A. J. Beavil, B. J. Sutton, D. Bray, Molecular model of a lattice of signalling proteins involved in bacterial chemotaxis. *Nat. Cell Biol.* **2**, 792–796 (2000).
76. T. S. Shimizu, S. V. Aksenov, D. Bray, A spatially extended stochastic model of the bacterial chemotaxis signalling pathway. *J. Mol. Biol.* **329**, 291–309 (2003).

77. T. W. McKeithan, Kinetic proofreading in T-cell receptor signal transduction. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5042–5046 (1995).
78. Q. J. Li, A. R. Dinner, S. Qi, D. J. Irvine, J. B. Huppa, M. M. Davis, A. K. Chakraborty, CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse. *Nat. Immunol.* **5**, 791–799 (2004).
79. K. Takahashi, S. N. Arjunan, M. Tomita, Space in systems biology of signaling pathways—towards intracellular molecular crowding in silico. *FEBS Lett.* **579**, 1783–1788 (2005).
80. D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976).
81. D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
82. M. Krogsgaard, Q.-J. Li, C. Sumen, J. B. Huppa, M. Huse, M. M. Davis, Agonist/endogenous peptide-MHC heterodimers drive T cell activation and sensitivity. *Nature* **434**, 238–243 (2005).
83. P. J. Woolf, J. J. Linderman, An algebra of dimerization and its implications for G-protein coupled receptor signaling. *J. Theor. Biol.* **229**, 157–168 (2004).
84. J. M. Haugh, I. C. Schneider, J. M. Lewis, On the cross-regulation of protein tyrosine phosphatases and receptor tyrosine kinases in intracellular signaling. *J. Theor. Biol.* **230**, 119–132 (2004).
85. C.-R. Yang, B. E. Shapiro, E. D. Mjolsness, G. W. Hatfield, An enzyme mechanism language for the mathematical modeling of metabolic pathways. *Bioinformatics* **21**, 774–780 (2005).
86. J. M. Haugh and J. R. Faeder, personal communication
87. H. Conzelmann, J. Saez-Rodriguez, T. Sauter, E. Bullinger, F. Allgöwer, E. D. Gilles, Reduction of mathematical models of signal transduction networks: Simulation-based approach applied to EGF receptor signalling. *IEE Syst. Biol.* **1**, 159–169 (2004).
88. M. L. Blinov, J. R. Faeder, B. Goldstein, W. S. Hlavacek, BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* **20**, 3289–3291 (2004).
89. A. R. Reynolds, C. Tischer, P. J. Verveer, O. Rocks, P. I. H. Bastiaens, EGFR activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation. *Nat. Cell Biol.* **5**, 447–453 (2003).
90. F. Taguchi, Y. Koh, F. Koizumi, T. Tamura, N. Saijo, K. Nishio, Anticancer effects of ZD6474, a VEGF receptor tyrosine kinase inhibitor, in gefitinib ("Iressa")-sensitive and resistant xenograft models. *Cancer Sci.* **95**, 984–989 (2004).
91. R. J. Bagley, J. D. Farmer, W. Fontana, Evolution of a metabolism. In *Artificial Life II*, C. G. Langton, C. Taylor, J. D. Farmer, S. Rasmussen, Eds. (Addison Wesley, Redwood City, CA, 1991), pp. 141–158.
92. J.-L. Faulon, A. G. Sault, Stochastic generator of chemical structure. 3. Reaction network generation. *J. Chem. Inf. Comput. Sci.* **41**, 894–908 (2001).
93. G. Benkö, C. Flamm, P. F. Stadler, A graph-based toy model of chemistry. *J. Chem. Inf. Comput. Sci.* **43**, 1085–1093 (2003).
94. T. S. Shimizu, D. Bray, Computational cell biology—the stochastic approach, in *Foundations of Systems Biology*, H. Kitano, Ed. (MIT Press, Cambridge, MA, 2001), chap. 10.
95. J. D. Farmer, S. A. Kauffman, N. H. Packard, Autocatalytic replication of polymers. *Physica D.* **22**, 50–67 (1986).
96. J. D. Farmer, A Rosetta stone for connectionism. *Physica D.* **42**, 153–187 (1990).
97. J. F. Keane, C. Bradley, C. Ebeling, A compiled accelerator for biological cell signaling simulations, in *Proc. 2004 ACM/SIGDA 12th Int. Symp. Field Programmable Gate Arrays*, R. Tessier, H. Schmit, Eds. (ACM Press, New York, 2004), pp. 233–241.
98. L. Salwinski, D. Eisenberg, In silico simulation of biological network dynamics. *Nat. Biotechnol.* **22**, 1017–1019 (2004).
99. N. M. Borisov, N. I. Markevich, J. B. Hoek, B. N. Kholodenko, Signaling through receptors and scaffolds: Independent interactions reduce combinatorial complexity. *Biophys. J.* **89**, 951–966 (2005).
100. N. M. Borisov, N. I. Markevich, J. B. Hoek, B. N. Kholodenko, Trading the micro-world of combinatorial complexity for the macro-world of protein interaction domains. *Biosystems* **83**, 152–166 (2006).
101. H. Conzelmann, J. Saez-Rodriguez, T. Sauter, B. N. Kholodenko, E. D. Gilles, A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinform.* **7**, 34 (2006).
102. E. M. Clarke Jr., O. Grumberg, D. A. Peled, *Model Checking* (MIT Press, Cambridge, MA, 1999).
103. N. Chabrier-Rivier, M. Chiaverini, V. Danos, F. Fages, V. Schächter, Modeling and querying biomolecular interaction networks. *Theor. Comput. Sci.* **325**, 25–44 (2004).
104. G. Batt, D. Ropers, H. de Jong, J. Geiselmann, R. Mateescu, M. Page, D. Schneider, Validation of qualitative models of genetic regulatory networks by model checking: Analysis of the nutritional stress response in *Escherichia coli*. *Bioinformatics* **21**, I19–I28 (2005).
105. C. Piazza, M. Antoniotti, V. Mysore, A. Policriti, F. Winkler, B. Mishra, Algorithmic algebraic model checking I: Challenges from systems biology. *Lect. Notes Comput. Sci.* **3576**, 5–19 (2005).
106. D. T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **115**, 1716–1733 (2001).
107. L. Lok, personal communication.
108. F. Roselló, G. Valiente, Graph transformation in molecular biology. *Lect. Notes Comput. Sci.* **3393**, 116–133 (2005).
109. B. R. Franza, From play to laws: Language in biology. *Sci. STKE* **2004**, pe9 (2004).
110. C. Priami, A. Regev, E. Shapiro, W. Silverman, Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Inform. Process. Lett.* **80**, 25–31 (2001).
111. A. Phillips, L. Cardelli, A correct abstract machine for the stochastic π-calculus. In *Proc. BioCONCUR 2004*, London, 30 August 2004.
112. M. Calder, S. Gilmore, J. Hillston, Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA, in *Proc. BioCONCUR 2004*, London, 30 August 2004.
113. R. Milner, *Communicating and Mobile Systems: The π-calculus* (Cambridge University Press, Cambridge, 1999).
114. W. Fokkink, *Introduction to Process Algebra* (Springer, Berlin, 2000).
115. J. C. M. Baeten, A brief history of process algebra. *Theor. Comput. Sci.* **335**, 131–146 (2005).
116. A. Regev, W. Silverman, E. Shapiro, Representation and simulation of biochemical processes using the π-calculus process algebra, in *Pac. Symp. Biocomput. 2001*, R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale, T. E. Klein, Eds. (World Scientific, Singapore, 2001), pp. 459–470.
117. W. Fontana, L. W. Buss, The barrier of objects: From dynamical systems to bounded organizations. In *Boundaries and Barriers*, J. Casti, A. Karlqvist, Eds. (Addison-Wesley, Reading, MA, 1996), pp. 56–116.
118. A. Regev, E. Shapiro, Cells as computation. *Nature* **419**, 343 (2002).
119. C. Priami, P. Quaglia, Modelling the dynamics of biosystems. *Brief. Bioinform.* **5**, 259–269 (2004).
120. V. Danos, C. Laneve, Formal molecular biology. *Theor. Comput. Sci.* **325**, 69–110 (2004).
121. P. Lecca, C. Priami, P. Quaglia, B. Rossi, C. Laudanna, G. Constantin, A stochastic process algebra approach to simulation of autoreactive lymphocyte recruitment. *Simulation* **80**, 273–288 (2004).
122. D. D'Ambrosio, P. Lecca, G. Constantin, C. Priami, C. Laudanna, Concurrency in leukocyte vascular recognition: Developing the tools for a predictive computer model. *Trends Immunol.* **25**, 411–416 (2004).
123. C. Talcott, D. L. Dill, The pathway logic assistant, in *Third International Workshop on Computational Methods in Systems Biology*, G. Plotkin, Ed., Edinburgh, 3 to 5 April 2005, pp. 228–239.
124. S. Eker, M. Knapp, K. Laderoute, P. Lincoln, C. Talcott, Pathway Logic: Executable models of biological networks. *Electron. Notes Theor. Comput. Sci.* **71**, 125–142 (2004).
125. M. G. Sriram, Modelling protein functional domains in signal transduction using Maude. *Brief. Bioinform.* **4**, 236–245 (2003).
126. C. Talcott, S. Eker, M. Knapp, P. Lincoln, K. Laderoute, Pathway logic modeling of protein functional domains in signal transduction, in *Pac. Symp. Biocomput. 2004*, R. B. Altman, A. K. Dunker, L. Hunter, T. A. Jung, T. E. Klein, Eds. (World Scientific, Singapore, 2004), pp. 568–580.
127. S. Eker, M. Knapp, K. Laderoute, P. Lincoln, J. Meseguer, K. Sonmez, Pathway logic: Symbolic analysis of biological signaling, in *Pac. Symp. Biocomput. 2002*, R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale, T. E. Klein, Eds. (World Scientific, Singapore, 2002), pp. 400–412.
128. F. Fages, S. Soliman, N. Chabrier-Rivier, Modelling and querying interaction networks in the biochemical abstract machine BIOCHAM. *J. Biol. Phys. Chem.* **4**, 64–73 (2004).
129. M. Y. Vardi, Branching vs. linear time: Final showdown. *Lect. Notes Comput. Sci* **2031**, 1–22 (2001).
130. L. Calzone, N. Chabrier-Rivier, F. Fages, L. Gentils, S. Soliman, Machine learning bio-molecular interactions from temporal logic properties, in *Third International Workshop on Computational Methods in Systems Biology*, G. Plotkin, Ed., , Edinburgh, 3 to 5 April 2005.
131. M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, and the rest of the SBML Forum: A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang, The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).

132. N. Le Novère, A. Finney, M. Hucka, U. S. Bhalla, F. Campagne, J. Collado-Vides, E. J. Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J. L. Snoep, H. D. Spence, B. L. Wanner, Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* **23**, 1509–1515 (2005).
133. N. Le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J. L. Snoep, M. Hucka, BioModels Database: A free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* **34**, D689–D691 (2006).
134. A. Finney, Possible extension to the Systems Biology Markup Language: Complex species and species graphs. (2001) (http://sbml.org/wiki/SBML_Level_3_Efforts).
135. N. Le Novère, T. S. Shimizu, A. Finney, Systems Biology Markup Language (SBML) Level 3 proposal: Multistate features. (2002) (http://sbml.org/wiki/SBML_Level_3_Efforts).
136. A. Finney, Systems Biology Markup Language (SBML) Level 3 proposal: Multi-component species features. (2004) (http://sbml.org/wiki/SBML_Level_3_Efforts).
137. M. L. Blinov, J. R. Faeder, B. Goldstein, A. Finney, W. S. Hlavacek, Rule-based modeling of multi-component species: Proposal for SBML Level 3. (2004) (http://sbml.org/wiki/SBML_Level_3_Efforts).
138. A. Finney, M. Hucka, Systems biology markup language: Level 2 and beyond. *Biochem. Soc. Trans.* **31**, 1472–1473 (2003).
139. M. Hucka, A. Finney, B. J. Bornstein, S. M. Keating, B. E. Shapiro, J. Matthews, B. L. Kovitz, M. J. Schilstra, A. Funahashi, J. C. Doyle, H. Kitano, Evolving a lingua franca and associated software infrastructure for computational systems biology: The Systems Biology Markup Language (SBML) project. *IEE Proc. Syst. Biol.* **1**, 41–53 (2004).
140. H. Hermjakob, L. Montecchi-Palazzi, G. Bader, J. Wojcik, L. Salwinski, A. Ceol, S. Moore, S. Orchard, U. Sarkans, C. von Mering, B. Roechert, S. Poux, E. Jung, H. Mersch, P. Kersey, M. Lappe, Y. Li, R. Zeng, D. Rana, M. Nikolski, H. Husi, C. Brun, K. Shanker, S. G. N. Grant, C. Sander, P. Bork, W. Zhu, A. Pandey, A. Brazma, B. Jacq, M. Vidal, D. Sherman, P. Legrain, G. Cesareni, I. Xenarious, D. Eisenberg, B. Steipe, C. Hogue, R. Apweiler, The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat. Biotechnol.* **22**, 177–183 (2004).
141. A. H. Y. Tong, B. Drees, G. Nardelli, G. D. Bader, B. Brannetti, L. Castagnoli, M. Evangelista, S. Ferracuti, B. Nelson, S. Paoluzi, M. Quondam, A. Zucconi, C. W. V. Hogue, S. Fields, C. Boone, G. Cesareni, A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* **295**, 321–324 (2002).
142. L. J. Jensen, J. Saric, P. Bork, Literature mining for the biologist: From information retrieval to biological discovery. *Nat. Rev. Genet.* **7**, 119–129 (2006).
143. R. Riley, C. Lee, C. Sabatti, D. Eisenberg, Inferring protein domain interactions from databases of interacting proteins. *Genome Biol.* **6**, R89 (2005).
144. M. I. Aladjem, S. Pasa, S. Parodi, J. N. Weinstein, Y. Pommier, K. W. Kohn, Molecular interaction maps—a diagrammatic graphical language for bioregulatory networks. *Sci. STKE* **2004**, pe8 (2004).
145. K. W. Kohn, M. I. Aladjem, J. N. Weinstein, Y. Pommier, Molecular interaction maps of bioregulatory networks: A general rubric for systems biology. *Mol. Biol. Cell* **17**, 1–13 (2006).
146. H. Kurata, N. Matoba, N. Shimizu, CADLIVE for constructing a large-scale biochemical network based on a simulation-directed notation and its application to yeast cell cycle. *Nucleic Acids Res.* **31**, 4071–4084 (2003).
147. R. Maimon, S. Browning, Diagrammatic notation and computational structure of gene networks, in *Proc. 2nd Int. Conf. Syst. Biol.* T.-M. Yi, M. Hucka, M. Morohashi, H. Kitano, Eds. (Omnipress, Madison, WI, 2001), pp. 311–317.
148. R. Maimon, Computational theory of biological function I—the kinematics of molecular trees. (2005) (http://www.arXiv.org/abs/q-bio.MN/0503028).
149. S. V. Aksenov, personal communication.
150. P. Dittrich, J. Ziegler, W. Banzhaf, Artificial chemistries: A review. *Artif. Life* **7**, 225–275 (2001).
151. P. Prusinkiewicz, A. Lindenmayer, *The Algorithmic Beauty of Plants* (Springer-Verlag, New York, 1990).
152. B. Berger, P. W. Shor, L. Tucker-Kellogg, J. King, Local rule-based theory of virus shell assembly. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7732–7736 (1994).
153. T. Zhang, R. Schwartz, Simulation study of the contribution of oligomer/oligomer binding to capsid assembly kinetics. *Biophys. J.* **90**, 57–64 (2006).
154. T. Pawson, R. Linding, Synthetic modular systems—reverse engineering of signal transduction. *FEBS Lett.* **579**, 1808–1814 (2005).

155. We are grateful to all those who have contributed to our synthesis of ideas about rule-based modeling, especially J. Cavenaugh, J. Colvin, A. M. Evangelisti, M. L. Fanning, G. M. Fricke, J. Kozdon, N. Lemons, F. Mu, A. N. Singh, A. Trehan, A. Vandenberg, and J. Yang. We also thank C.-S. Tung for providing Fig. 3, L. Lok for informing us of the look-ahead feature of Moleculizer, and S. V. Aksenov, V. Danos, B. Goldstein, B. N. Kholodenko, P. M. Loriaux, and C. Talcott for their comments about the manuscript. Our work has been supported by NIH grants RR18754, GM35556, and AI35997 and DOE contract W-7405-ENG-36.

Citation: W. S. Hlavacek, J. R. Faeder, M. L. Blinov, R. G. Posner, M. Hucka, W. Fontana, Rules for modeling signal-transduction systems. *Sci. STKE* **2006**, re6 (2006).