A Spectrum of Models of Signaling Pathways

Sharat J. Vayttaden, Sriram M. Ajay, and Upinder S. Bhalla*[a]

1. Introduction

Models in biology have a long and distinguished history. By far the most common form of biological model is the word model, where abstractions of many observations are simply stated in plain language. For example, the theory of evolution is stated as a word model. Despite this familiarity with word models, biologists have tended to be suspicious of putting the words into mathematical form. There are good reasons for such concern. The mathematical approach often brings with it a bias toward simple and general explanations of observations. Evolution, on the other hand, does not always choose simple solutions, and abounds in special cases. It takes special effort to develop models that are faithful to biology while drawing upon computational and mathematical tools.

This review describes a spectrum of such models for signaling pathways. At one end of the spectrum are word models arranged as the familiar signaling-pathway diagrams. At the other end there are carefully crafted computer simulations in which every reaction in the signaling cascade is experimentally defined and has an attached rate constant. Models across this spectrum have great promise for understanding signaling complexity. It is interesting, at this early stage in the field, to present a snapshot of current model development and to note how the models themselves have begun to evolve.

Why model signaling? The challenges are considerable, and at least the common perception is that such models have not been particularly useful. Despite these difficulties, there is a general consensus that modeling approaches are necessary. The motivation for this is the sheer complexity of signaling, backed up by a flood of raw data. The past decade has seen an exponential rise in high-throughput experimentation. The philosophy behind performing large-scale genome sequencing, gene expression profiling, and proteomic analysis can be summed up in a quote by Craig Venter: "If we hope to understand biology, instead of looking at one little protein at a time, which is not how biology works, we will need to understand the integration of thousands of proteins in a dynamically changing environment."[1] Modeling and analysis tools are critical to understanding these often non-intuitive interactions.[2] In this study we restrict ourselves to a subset of modeling studies that deal with signaling pathways. Protein structure, genetic interactions, and statistical analysis of large biological datasets are other major areas of systems biology analysis, which we will not consider.

In this review, we first categorize models by their degree of quantification. We analyze the level of quantification feasible or already accomplished in current models, using extensive literature surveys to categorize some 244 models drawn from two public databases: DOQCS[3] (http://www.doqcs.ncbs.res.in) and BIOCARTA (http://www.biocarta.com/genes/index.asp). We use the survey to identify possible factors responsible for the success or failures of quantitative modeling efforts in specific signaling systems. We then examine the time-evolution of modeling strategies by studying a family tree of models of the Mitogen Activated Protein Kinase (MAPK) signaling pathway. Finally, we discuss how different levels of modeling have helped to advance the field by deciphering biological complexity.

2. The Spectrum of Models

We define signaling models as any abstraction of signaling function that has both descriptive and predictive power. A word model, in its typical embodiment as a signaling block diagram, has both, and so does a reaction-level simulation. Such models differ both in terms of the amount and kind of data that are needed to specify them, and in terms of the precision of their predictions. For concreteness, we consider three prototypical kinds of model: the block-diagram model, the systems model, and the semiquantitative model. In each case we describe the scope of the model, the kind of data that it incorporates, what this information boils down to in the model, and the kinds of predictions it can make. We illustrate each with an example.

2.1 Block-diagram models

Block-diagram models outline the relationships between the components of a signaling network. Such models specify the topology of connections between signaling molecules or pathways. These models usually also indicate the sign of the connections; that is, whether the interaction is excitatory or inhibitory. Block level models are the most commonly encountered descriptions of signaling pathways and are the staple for textbooks and databases. Such models are equivalent to pictorial transcriptions of word models.

An enormous amount of data is embodied in block-diagram descriptions. The initial observation often comes from genetic experiments in which mutants lacking a given molecule exhibit an interesting phenotype. As the molecule is characterized, its interactors, upstream or downstream, are worked out. These

[a] S. J. Vayttaden, S. M. Ajay, Dr. U. S. Bhalla
National Centre for Biological Sciences
Tata Institute of Fundamental Research
GKVK Campus, Bangalore 560065 (India)
Fax (+91) 80-2363-6662
E-mail: bhalla@ncbs.res.in

 DOI: 10.1002/cbic.200400127

observations add new molecules to the block diagram, and often link it to known pathways. Experiments that knock out specific molecules by genetic, molecular, or pharmacological methods help in the working out of further details of the interactions, such as the sign of the effect. By using genomic databases it is now often possible to relate known pathways between species, and to fill in homologous pathway diagrams fairly quickly.

The model abstraction of this immense amount of data is essentially a listing of interacting molecules or pathways, together with the sign of each interaction. Frequently this representation is expanded to include modulatory interactions where a third molecule can modulate the interaction between two others. Such data abstractions are sufficient to specify pathway block diagrams. Clearly, models of the form of pathway block diagrams are very concise descriptions of highly complex interactions. As we discuss below, such models are the starting point for any detailed description.

As is obvious, block-diagram models provide a limited range of qualitative predictions. These are often simply a matter of following the interaction diagram and predicting that removal of an upstream molecule will eliminate an effect on a downstream one. These qualitative predictions can, however, be extremely important. For example, block-diagram models of signaling networks often form the basis for designing knockouts or pharmacological interventions that are predicted to have specific effects. The key role of the NMDA receptor in learning and memory, for example, was confirmed through knockouts of specific receptor subtypes in specific regions of the hippocampus.[4]

An example of a topological model and its results are shown in Figure 1a and b. This model deals with the *obese* gene and the prediction that its protein product leptin would have a target receptor in the hypothalamus. More than 50 years ago, Kennedy hypothesized that the hypothalamus senses some factor that provided it information about the body's fat reserves, as a result of which the hypothalamus would regulate food intake.[5] A critical clue about the factor that regulates body weight came from Coleman's finding in the 1970s that recessive mutations in the mouse *ob* and *db* genes resulted in obesity and diabetes.[6] Through his experiments Coleman concluded that the blood-borne factor was encoded in the *ob* gene and the receptor for this factor was encoded in the *db* gene. It was only in 1994 that Friedman's group identified and characterized the *ob* gene and its product, leptin.[7] Tartaglia's group followed this up with the discovery of the leptin receptor in 1995.[8] Here we see that Kennedy and Coleman's word/block diagrams led the way to identification of leptin and its receptor.

In summary, block-diagram models encapsulate immense amounts of information about the topology of signaling interactions. Such models enable fundamental predictions about causality in signaling events, and are the starting points for all modeling analyses in signaling.

2.2 Systems models

Systems models are models that incorporate partial mechanistic detail about signaling interactions. Such models mathematically represent the architecture of the system, and thus can predict the general range of behavior that the system can adopt. The information is at a level that facilitates mathematical and numerical analysis over a range of possible parameters. For example, many such models specify the reaction scheme of the system and may also include partial reaction rates.

Such information is typically represented in the model as a set of differential or algebraic equations in which the specific rates are free parameters over a certain plausible range.

Armed with this information, one can show, for example, that under certain conditions the system will be stable, but in other parameter ranges it might oscillate or flip into a different state of activity (Figure 1c and d). Thus, even though such models lack detailed kinetic and quantitative parameters they still possess significant predictive power. The cell cycle is an example of a system that has been extensively studied at the systems level.[9] Early models were essentially theoretical and had abstract structures in which specific chemical reactions were not modeled completely. For example, one version of this model required only two equations and four kinetic parameters to specify its behavior.[9] Although these models were made before experimental data on system structure and kinetics were available, they were capable of qualitatively describing many aspects of in vivo and in vitro behavior. Tyson's 1991 model had an oscillatory period from 10 to 50 min, whereas in vivo the period for Xenopus cleavage cycles was shown to be 30 min[10] and the in vitro period was shown to be around 60 min.[11]

As illustrated by the cell cycle example, systems models are able to capture the essential behavior of a system qualitatively and to suggest directions for experiments to refine understanding of the system.

2.3 Semiquantitative models

Semiquantitative models are mechanistically the most advanced models currently available. The scope of such models is typically at the level of biochemical representation of signaling. More modern modeling efforts frequently go well beyond this, and may incorporate cell biological, spatial, and even mechanical details of cellular function.

The experimental data incorporated into such models is extremely varied. It is, of course, a superset of the systems and block-diagram representations. In addition to the genetic and molecular biological specification of interactants and the identification of reaction mechanisms, there is an immense range of kinetic and pharmacological techniques needed to parameterize models quantitatively. Test-tube biochemistry has a prominent role in specifying rates, though this is increasingly being supplanted by less precise but more sensitive measurements based on specific antibodies, such as western blots. High-resolution imaging methods are invaluable in determining localization and traffic of molecules, and such methods are

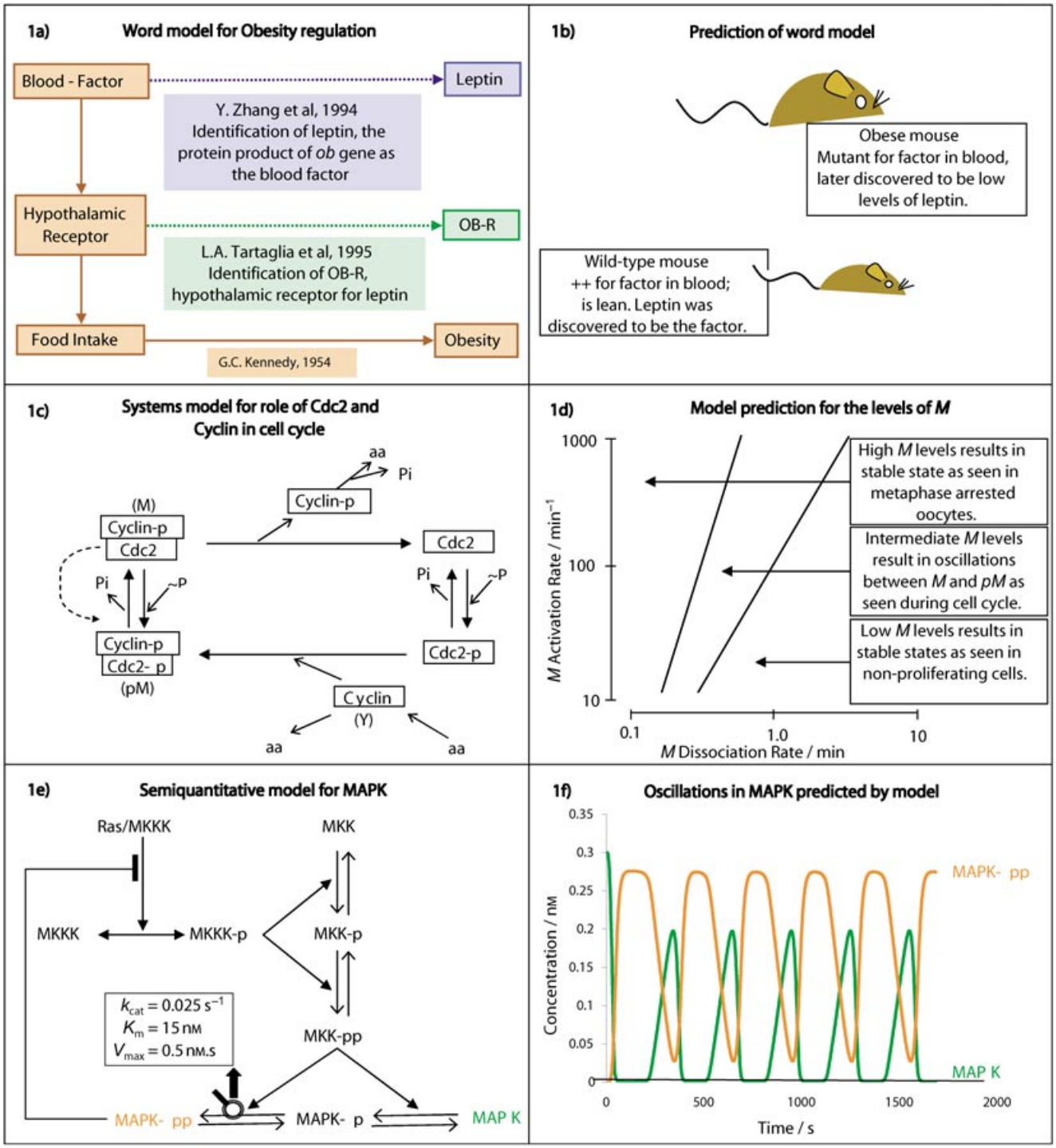

Figure 1. *Scope of a model. a) Example of block-diagram/word model for obesity. Different components of the word model prediction were discovered subsequently, as shown in purple and green. b) Prediction and experimental validation of obese phenotype. c) Example of a systems model for the cell cycle. The model represents the role of cdc2 and cyclin in cell cycle progression as described by Tyson in 1991. d) Qualitative prediction of cell cycle fate made by the model based on levels of* M *(Mitosis Promoting Factor or Complex of Cdc2 and Cyclin-p), the periodicity of which was later verified through experiments. e) Semiquantitative model for MAPK proposed by Kholodenko in 2000, describing a negative feedback from MAPK-PP to MKKK activating reaction resulting in oscillations of dually phosphorylated and non-phosphorylated forms of MAPK. f) Oscillations in MAPK predicted by the model.*

often amenable to kinetic measurements on the timescale of seconds. Surface plasmon resonance methods, though technically challenging, have great promise for providing high-quality kinetic parameters.

Once digested, this vast amount of data typically reduces to just three model quantities: reaction schemes, rate constants, and concentrations of molecules. In more advanced modeling efforts the localization of molecules and possibly their transport rates may be represented.

Good quantitative information opens the way to extremely detailed and specific predictions. Essentially, any experiment in the system with a readout expressed as a concentration term can be simulated. Time-series experiments are particularly challenging tests for simulation predictions, as these test both the steady-state properties and the dynamics of the model. Often the limiting point in making good predictions is precisely the point at which the model runs out of experimental input. Such situations are frequently the most productive interfaces between model and experiment.

The NF-κB system illustrates this point. NF-κB is a nuclear transcriptional factor present in the cytoplasm. It has a large number of activators and in response to specific activators it selectively up-regulates different subset of genes. Although some quantitative information on NF-κB's interactions in the cytoplasm exists, it is much more difficult to measure nuclear protein activity. The presence of some amount of quantitative information has allowed the construction of a model capable of predicting the cytoplasmic behavior and making semiquantitative predictions with respect to nuclear localization of NF-κB.[12]

The predictive value conferred by semiquantitative modeling often helps in refining theory and improving our understanding of a system. An example is Kholodenko's MAPK oscillatory model (Figure 1e and f).[13] This modeling effort makes specific predictions on the behavior of the MAPK biochemical pathway in mammalian cells. The study shows through computation modeling that the presence of a negative feedback loop in the MAPK cascade allows the system to undergo sustained oscillations. There are some recent examples of oscillations in the MAPK system: Akhthar et al.[14] and Duffield et al.[15] have shown that some genes encoding components of the Ras/MAPK signaling pathway do show oscillatory peaks, although these do not match the model in the details of time-course and amplitude.

Thus quantitative modeling efforts include extensive mechanistic and kinetic experimental data, and in turn provide specific and testable predictions as well as the basis for better understanding of signaling events.

3. Classification of Current Models of Signaling Pathways

The ideal model would, of course, be a superset of all the categories above. It would specify the identities of all the interacting pathways, and would provide full chemical and kinetic details about each of the chemical steps. In addition, it would include information about the cellular localization of each event, together with cell biological details such as trafficking and genetic interactions. In other words, it would embody all the relevant information about the system. Of course, such a model can only be constructed with hindsight, when the system is already well understood. The most influential models are often those that go out on a limb to predict biological phenomena on the basis of decidedly incomplete data. In this study we do not attempt to predict which models will eventually be seen in this light. Instead we construct a somewhat subjective classification based on the question of how currently known pathways would fare with regard to available experimental data. We have addressed this issue by analyzing a somewhat biased sample of 244 pathways for their prospects for being modeled. The systems for analysis have been selected from entries in BIOCARTA (http://www.biocarta.com/genes/index.asp) and from our database of quantitative cellular signaling DOQCS[3] (http://doqcs.ncbs.res.in). The use of these publicly accessible databases already applies a certain level of sampling bias towards fairly well known pathways. Clearly, by being included in such databases, the pathways are already at least at the block-diagram level. Nevertheless, the rather broad scope of BIOCARTA gives us a useful starting sample for the analysis.

Experience with modeling suggests that different systems have different kinds and amounts of experimental detail available. Therefore, we have used a flexible and simple set of criteria to classify them (Table 1). As described briefly in the table, the information required to model a system includes:

1) Component connectivity or topological information. This is the "bare bones" of the reaction connectivity between individual elements of the proposed pathway.

Table 1. *Different levels of interaction pathways based on availability of quantitative information.*

Model level	Topological detail	Cellular location	Temporal details	Structural details	In vivo kinetics	Test-tube kinetics
Level 0: Gold standard	available	available	available	available	available	available
This is essentially an unreachable standard for modeling and is likely to remain so as further studies add more parameters required for full system specification.						
Level 1: Blue standard	available	available	available/ not available	available/ not available	available/ not available	available
These are some of the best constrained models available, such as the *E. coli* chemotaxis model, where almost all the components and most of their interactions have been identified.						
Level 2: Green standard	available	available/ not available	available/ not available	available/ not available	available/ not available	available/ not available
Most of the semiquantitative models at this level and the level below are capable of making systems-level predictions. Models at this level are also capable of making good mechanistic predictions.						
Level 3: Yellow standard	available/ not available	available/ not available	not available	not available	not available	available/ not available
In these models most of the suspect interactions in the red model have been eliminated and some test tube level kinetic details is available for a few of the components.						
Level 4: Red standard	available/ not available	not available	not available	not available	not available	not available
Models at this level are essentially collections of "interactions" between various components of the system. These interactions themselves may be suspect, they have no quantitative details available, and in most cases lack cellular context. They do, however, provide a probable overview of the global structure or topology of the cell signaling networks.						

2) Cellular localization of components of the system. It is frequently the case that localization data is unavailable even when details of specifically interacting molecules are known and their kinetic effects have been observed. Many experimental results are of the form of blots or gene expression profiles from homogenized tissue samples. These experimental techniques monitor populations of cells rather than single cells, and so their ability to delineate intracellular signaling pathways will be limited. A more detailed model requires information on cellular localization of key molecules.
3) Temporal dynamics of expression and activation of the components. Current experimental approaches frequently sample individual time points rather than the continuous progression of signaling events. Temporal details are often monitored only in systems with slow time-courses, such as the circadian rhythm or the cell cycle; however, such data is valuable in most models.
4) Structural details of the interacting components. In many models of signaling pathways it is not known whether the reactions are between freely moving components or between tethered or scaffolded molecules. Further, reactions occurring between molecules in distinct compartments are often poorly characterized, especially with regard to the distribution of molecules between compartments. A more detailed mechanistic model may need to incorporate these features to represent the nuances of the reaction dynamics.
5) Identification of physiologically accurate quantitative values of each of the components. Old-fashioned "test-tube" biochemistry is a good source of reaction rates, which are very important to constrain a model. It is often valuable to compare such rates with in vivo measurements to obtain good parameters for models.

Using the availability of these parameters as a criterion for classification, we have classified our sample of models into one of five categories on a scale of biochemical exactitude. These range from 0 for excellent models (currently unattainable) to 4 for very sketchy ones. The availability of parameters is based on published literature associated with the system. Table 1 shows the color coding and the classification criteria. In the interests of space, we describe only our criteria of classification, but the pathways classified under this scheme are tabulated and provided at our website (http://www.ncbs.res.in/~bhalla/model_spectrum/index.html).

3.1. Model-classification results

We performed the above classification for 244 models, ranging from block-diagram descriptions to very carefully quantified reaction-level simulations. We stress that the classification combines an assessment of the potential of a model for quantitative modeling, based on available experimental data, as well as its current status, based on published models. The purpose of this classification is not to make any value judgment of the models but only to identify the level of chemical exactitude in these models. It is important to note that highly quantitative models may or may not have a correspondingly high impact in the future development in the field. Arguably, it is difficult to compare models with different features addressing different facets, so a predefined set of criteria are used for making the comparison. The models are evaluated as explained in Table 1 on a given set of parameters as well as on the validity of predictions within the system and the advancement made by the model. We first simply categorized models into five levels. We then analyzed subfields of study to see how different fields were represented. We considered some of the experimental constraints on different model systems, and assessed how this contributes to model development in that area.

3.1.1 Levels of models: As expected, there are more models at the block-diagram level than at the quantitative level. Although it may be an artifact of our sample population, it is interesting that the numbers of semiquantitative models (levels 2 and 3) is potentially rather large, rather than being a miniscule fraction of the level 4 models (Figure 2a). This is extremely encouraging from the systems biology viewpoint as it indicates that even with current techniques we can advance quantitative modeling in many subfields of biology. At the same time, it is sobering that there are a miniscule number of models that are "well quantified" by our classification scheme, and essentially no excellent ones of level 0. It is therefore clearly important to consider development of high-throughput methods that would be capable of generating the kind of data needed to bring models to this more quantitatively predictive level.

3.1.2 Experimental constraints on modeling in different subfields: We have loosely categorized model subfields into metabolic, neuronal, immunological, and cell signaling. Analysis of the classified data revealed several interesting trends in the types of data characteristic of each field. Experimentally difficult systems have limiting quantitative data, as expected, and models made in these cases would fall into the very sketchy or level 4 category. For example, brain-related disorders and neurodegenerative disorders rarely have quantitative information. Quite often the topological connectivity itself seems to be incomplete in many cases. In contrast, disorders of accessible systems, such as blood-clotting defects, have been studied to a high level of quantitative detail (level 2), and so a model of blood clotting defects with significant predictive capacity can be made. It is also interesting that though brain disorders are poorly quantified, neuronal signaling is rather well studied, especially at the synaptic level (Figure 2c).

A similar bias in data availability is seen in mammalian as opposed to non-mammalian systems. Simpler non-mammalian systems frequently have better quantitative data. This bias could be because of ease of accessibility and a longer history of using a non-mammalian experimental model. This is reflected in higher percentages of models of non-mammalian systems in levels 2 and 3, whereas mammalian models tend to be sketchy and largely populate level 4 (Figure 2b).

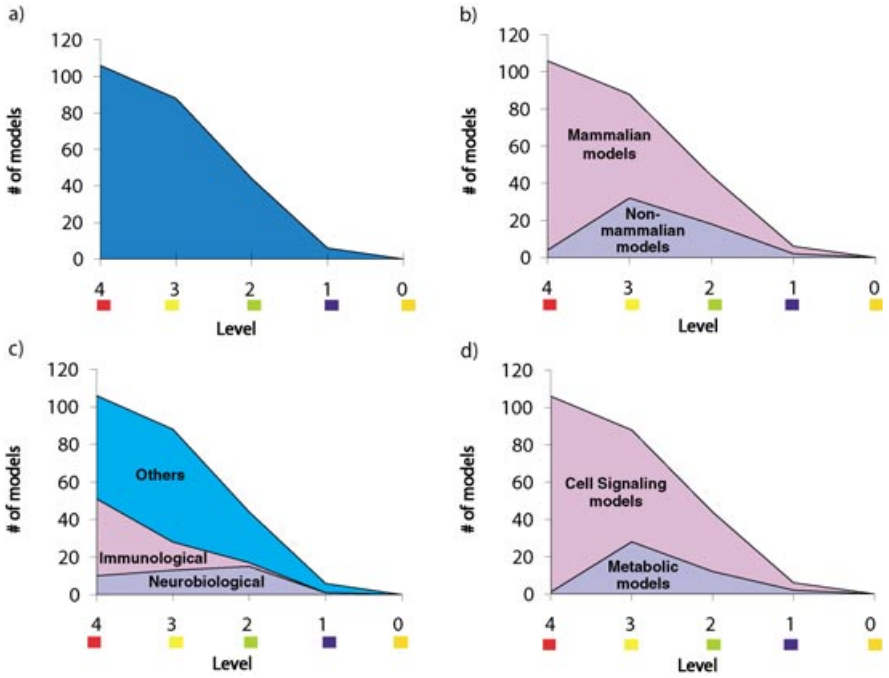

Figure 2. *Classification and analysis of models in different levels. a) Distribution of 244 models over the five levels of classification (see text and Table 1). b) Subdivision of models between mammalian and non-mammalian systems. Non-mammalian systems have a greater proportion of quantitative models. c) Distribution of models in different systems: immunological, neurobiological, and others. Neurobiological models tend to be better quantified. d) Subdivision of models between signaling and metabolic models. Metabolic models tend to be better quantified.*

An interesting finding is that experimental access does not necessarily result in quantitative data measurements. Though immunology is a very important and well represented area, the data available for model construction are mostly of level 3 or level 4 (Figure 2c). One possible historical reason for this is that the approach adapted in immunological experiments is to identify interactants and/or genetic relation between components. The problem is compounded by redundancy and large number of alternate interactions in signaling events in this field.

Metabolic pathways have had a longer history of investigation than cellular signaling pathways. Furthermore, many metabolic pathway investigations are associated with studies to probe for pharmacological intervention. This has apparently pushed the field towards adopting a quantitative approach. Consequently, most models of metabolic pathways are quite biochemically detailed, at level 2 or level 3. In comparison, the distribution of models of signaling pathways is broader (Figure 2d).

Overall, in our survey of models, we find that there are a substantial number of pathways "ripe" for more quantitative modeling. There is a surprisingly uneven distribution of model types over different fields, and this is only partially accounted for by experimental difficulties. There appears to be considerable scope for more experimentation that may bring many more pathways within reach of more quantitative models.

4. Evolution of Models

In view of the stated goals of systems biology in producing ever more accurate representations of biological systems, it is interesting to follow the evolution of models in a specific signaling pathway over time. Do models really become better over time? Can one trace the influence of models and experiments on each other? Do the models provide successive improvements in biological understanding? We chose the MAPK signaling pathway to address these issues, as it has been modeled at successively greater detail since the early 1990s, and now boasts a greater proliferation of models than almost any other signaling pathway (Figure 3).

The MAPK cascade as we know it today is a three-tiered cascade of kinases: Raf→MEK→MAPK (Figure 1e shows a generic MAPK cascade). It took several years to resolve even this basic topology of the cascade. As recently as 1990 it was not clear whether the kinases in the cascade were sequentially or simultaneously activated.[16] A two-tiered ordered arrangement of the kinases had been proposed by 1991,[17] and in the span of two years from 1991 to 1993 the three-tiered core structure of the MAPK cascade was identified.[18] Further refinements to the topology were made by the discovery that the MAPK cascade involved scaffold proteins.[19]

The first simulations of MAPK (level 3 models) were carried out in 1996[20] in a kinetic model of MAPK demonstrating ultrasensitivity. This model was constrained by experimental data, but contained approximations to the rates. In 1997, Ferrel and Bhatt showed that MAPKK phosphorylates p42 MAPK by a two-collision distributive mechanism rather than a single-collision processive mechanism.[21] This experimental model provided a mechanistic basis for understanding of how MAPK can convert graded inputs into switch-like outputs. In the same year, Burack and Sturgill showed through experiments and kinetic analysis of available data that the mechanism of ERK2 activation by MEK1 in vitro is actually nonprocessive.[22] In 1998, Ferrell and Machleder showed that the MAPK cascade is activated essentially in an all-or-none fashion during Xenopus oocyte maturation.[23] This behavior was proposed to arise from two known properties of the oocyte's MAPK cascade: positive feedback and the cascade's intrinsic ultrasensitivity, proposed in 1996 by Huang and Ferrell.

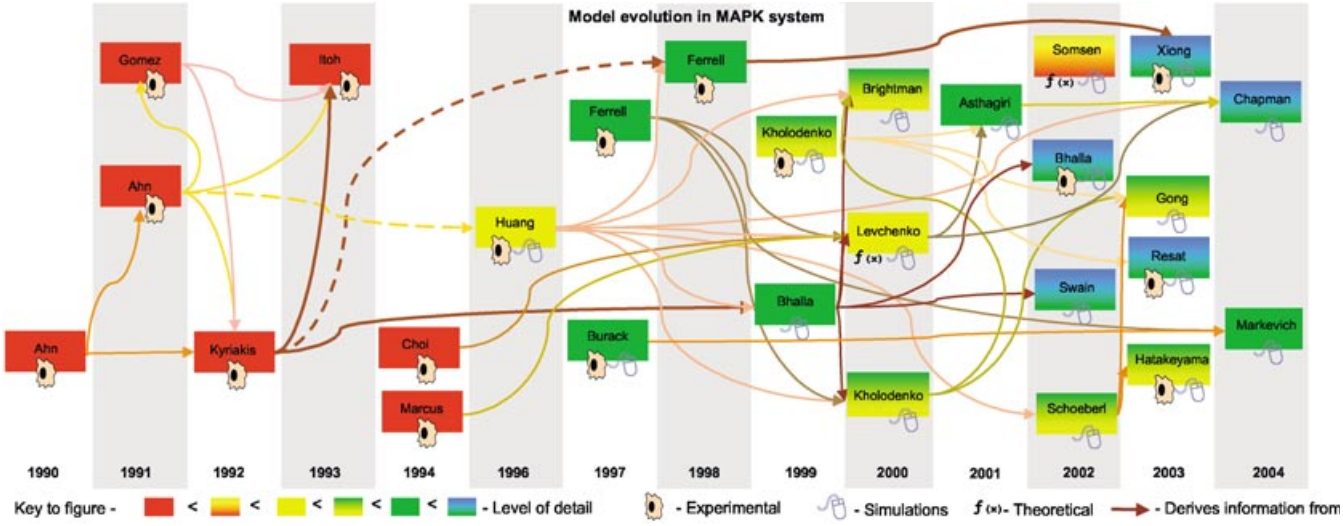

Figure 3. *Evolution of modeling, with modeling in MAPK taken as an example. Models are arranged chronologically from left to right by the year of their publication. Each model considered is represented by a box, and the color in the box is indicative of the level of detail incorporated in the model. The classification of the levels is in accordance with the criteria detailed in Table 1 (also see text). The boxes are color coded in increasing order of detail, red being the lowest and blue being the highest. The icon at the base of the box indicates the approach adopted by the study for modeling. The connections to each model indicates important source of starting information. All connections originating from a single model are represented by a single color.*

Following this initial set of models, the next phase began to analyze additional components of the MAPK cascade, and to explore its dynamics in detail.

In 1999, Bhalla and Iyengar used simulations based on kinetic data available at the time to propose that the MAPK cascade might participate in a bistable feedback loop.[24] This MAPK model was one of the earliest to incorporate EGF activation in its description. In the same year, Kholodenko modeled the EGF signal transduction from the receptor to the RasGTPase.[25] In 2000 Kholodenko proposed an oscillatory mechanism in a MAPK cascade model.[13] Levchenko explored effects of scaffolding on the MAPK cascade, and his model proposed that scaffold proteins may biphasically affect the levels of MAPK signaling and thereby reduce its threshold properties.[26] Brightman and Fell showed through quantitative modeling that feedback inhibition of the MAPK cascade determined the duration of cascade activation.[27] A further study on MAPK system dynamics by Asthagiri and Lauffenberger showed that negative feedback could enhance an upstream signal in the MAPK cascade.[28]

In the most recent phase of modeling, two trends are apparent. Firstly, recent models incorporate still further cell biological detail, including receptor traffic and transcriptional control. Secondly, recent studies have begun to utilize both experiments and simulations in an integrative fashion. We mention some of these studies in Figure 3. Models have incorporated various kinds of cell biological detail. Schoeberl et al. have modeled the effects on MAPK due to receptor internalization of EGF receptors.[29] In 2002, Bhalla et al. combined experiments and modeling to support MAPK involvement in a bistable feedback loop.[30] This study also considered transcriptional activation of MKP-1 as an important component of the history dependence of the cellular response. Swain and Siggia modeled the multisite phosphorylation of MAPK and suggested that it acts to improve signaling specificity.[31] In a theoretical study, Somsen et al. showed that scaffolding could induce selectivity in different MAPK modules even if they shared the same kinases at some levels in the cascade.[32] In one of the largest models of the MAPK cascade yet attempted, Resat et al. simulated the differential kinetics of EGFR activation by EGF and TGF-alpha using a large multi-compartment spatio-temporal model. In the same study they also provided experimental support for the model predictions about differential receptor activation.[33] In 2003, Hatakeyama et al. modeled dual regulation of heregulin-induced ErbB signaling.[34] This work was based on Schoeberl's 2002 model and also involved some experimental verification. Xiong et al. showed in 2003 that a positive feedback loop between activation of MAP kinase (MAPK) and the cell cycle regulator cdc2 is responsible for ensuring the self-supporting decision of the oocyte to mature.[35] This experimental work verified this proposal put forward by the 1998 Ferrell and Machleder model. In 2003, Gong et al. modeled redundancy and dominance of the Shc-dependent pathway during MAPK activation.[36] Gong based this simulation work on Schoeberl's 2002 model and also on the 1999 and 2000 Kholodenko model. In 2004 Markevich et al. showed by kinetic analysis that bistability and hysteresis are inherent properties of multi-step phosphorylation-dephosphorylation cycles.[37] They proposed that this might cause a MAPK cascade to exhibit bistable behavior even in the absence of feedback loops. In 2004, Chapman and Asthagiri showed through network component analysis that amplification, input potency, and dynamic range of output in the MAPK cascade may be tuned by manipulating module components.[38]

There are several interesting features that emerge from our analysis of the lineage of these models. Firstly, there is a very strong evolutionary dependence on older models. This is even more striking when one compares rate constants across

models (data not shown). It is clearly a field where cross-fertilization is important and effective. Secondly, models get bigger, not just better. The scope of some of the recent models extends from the cell surface to the nucleus, and includes spatial detail in-between. The kinds of questions addressed by such models are correspondingly more biologically complete. Another interesting observation that can be seen in Figure 3 is that although many of these models share common information sources, they are not all given the same rating. The following example illustrates this point. Two simulation models in 2000 explored distinct, novel aspects of the MAPK cascade. Levchenko's model was one of the first models to state a theoretical quantitative model of the MAPK cascade with a generic scaffold protein. Kholodenko's model proposed a new behavior for the MAPK cascade based on well defined existing properties of the MAPK cascade: namely, negative feedback and ultrasensitivity. Though they have similar sources of starting information, the Kholodenko model is rated as more quantitatively defined. The Levchenko model introduced a new generic component and is thus chemically less well defined than Kholodenko's model, which explored new behavior of the system. There is no new component introduced into Kholodenko's reaction scheme, and hence it is chemically better defined, which reflects as a higher rating. However, one should not confuse the chemical exactness used in this classification with the impact the model may have on future research in the field. Indeed, it could be argued that one of the key roles of such modeling is to explore the effects of novel or speculative biological interactions.

A more subtle observation is that the environment of previous models sets a ratchet for the quantitative precision of new studies. The general color in the figure shifts from red (low quantization) toward blue (very good quantization). This may seem like an inevitable outcome of refinement of models, until one recognizes that new models also incorporate additional interactions, spatial detail, and stochasticity. Given that each of these additional attributes needs characterization, it is by no means a given that larger models should also have better parameters.

Finally, this chart focuses on modeling studies rather than the larger experimentally driven context of the field. Nevertheless, even within this limited set of modeling studies, a clear intertwining of experiments with the models is evident. This has become explicit in recent years with several combined model/experiment studies.

Thus, over a span of some 14 years, modeling efforts in the MAPK system have evolved from topological models to extremely detailed studies of specific hypotheses with close reference to experiments. The questions involved have changed from broad issues of signal flow to specific tests of the nature of feedback and other key systems biology attributes of the cascade. While biological experiments remain the arbiter of all these studies, it seems likely that the conceptual level of the questions being posed owes a great deal to modeling.

5. Biological Complexity and Modeling

Models help to handle complexity in at least three major ways: by organizing masses of data, by predictions in complex situations where intuition fails, and ultimately, as a tool to improve understanding. In this survey we have considered 244 models, most of which do all of these things. We conclude with some illustrations of the use of models to handle complexity in biological signaling pathways.

Data organization is obviously the territory of databases, and the management and analysis of biological data is now an entire field in itself. This is variously referred to as bioinformatics, or one or other form of the much-abused suffix "omics". Models play a very important complementary role to the conventional database representations of signaling.[39] In particular, with increasing model detail, models can provide not just a parts list, but a data representation with functional capabilities built in. Even at the block-diagram level, the flow of information in a signaling cascade is evident. At the systems level, models can be used to identify gaps in the original data that are needed to explain system behavior.[9,40] At the quantitative level, the model embodies the state of knowledge not just about the structure of the signaling pathway, but also about its dynamics and behavior under many sets of conditions.[41] These capabilities clearly go far beyond the basic search-and-link features of most databases in representing the function of complex signaling. Modeling also serves as a means for checking the completeness and consistency of the different sets of information available for a system.

Prediction of system function is the next step for models. For simple signaling it is probably not critical to have a model representation. For example, the classical cyclic AMP signaling cascade is a linear pathway with some amplification. As one incorporates additional data, such as receptor down-regulation, turnover, alternative isoforms, and signal cross-talk, even this simple pathway needs at least a block diagram to predict what might happen in response to an input or modulator. A systems level description improves on this by predicting parameter ranges in which different kinds of behavior may occur. With quantitative models one can close the loop with experiment, and specifically predict outcomes of manipulations. This is especially valuable when such outcomes are non-intuitive.[23,42] A particularly powerful illustration of this is when one can distinguish between physiological outcomes of competing hypotheses embodied as models.[26] Models of complex signaling systems are therefore extremely important tools to supplement human intuition in trying to predict how such systems will behave.

Understanding of complex signaling is where we propose that models will have their most fundamental impact. Quite often such models propose concepts far ahead of any experimental validation.[13,26] Notwithstanding the accuracy of such proposals, they make significant contributions to conceptual advancement. Even at the block-diagram level, systems biologists can identify functional modules that allow one to replace a jumble of pathways with an abstracted black box[43] With the systems level model to hand, the mathematical form of the

model can reveal similarity with other well understood systems from engineering or physics. For example, the form of the feedback circuit in bacterial chemotaxis adaptation matches that of a perfectly adapting integral feedback control circuit in engineering.[44] Finally, with quantitative models, one can not only recognize such functional modules and test their operation,[30] but even design signaling systems with desired properties.[45]

We propose that this enumeration of the roles of different levels of modeling is not merely a classification, but instead a framework for the emerging field of systems biology. How does one understand these enormously complex biological systems? Going by the spectrum of models described above, we would propose that a first pass is to define the interactions (block-diagram descriptions), the second pass is to define mechanisms (systems descriptions), and the third pass is to quantify each step (quantitative specifications). As the level of description is refined, models go from being completeness checks, to doing quality control, to providing predictions and a deeper understanding. Our survey, even though biased towards already developed descriptions, suggests that much of the field is currently at a "first pass" level. From this perspective, there are clear further steps to pursue for the development of experimental and computational tools to tackle the second and third levels.

6. Summary

In this review we have conducted a survey of approximately 250 models of cellular signaling from the block-diagram level up to quantitative models. We assess the current level of representation of each model and find that a significant fraction of systems are sufficiently characterized to be amenable to quantitative modeling. We find that certain fields, such as metabolic analysis, have been very successful in bringing quantitative modeling techniques to their systems. Other areas, such as immunology, appear to be underrepresented among quantitative models, especially given the experimental strength in the area. It is tempting to speculate that the recent interest in systems biology may change this mismatch.

As a specific example of model evolution, we have traced the family tree of models of the MAPK signaling cascade. We observed several phases in the evolution of such models, from sketchy qualitative block diagrams, through a range of early models outlining the basic properties of the cascade, to an explosion of models using the cascade in the context of larger circuits. A remarkable synergy between models and experiments is evident in this family tree. If this history anticipates model development in other systems, it seems likely that combinations of experiments and models will be a strong driving force in the area. We propose that this example suggests a framework for the evolution of the field of systems biology as a whole, through staged refinement of biological understanding. Our survey suggests that the field is still finding its feet in defining the most basic block diagrams of complex biological systems.

Overall, our study is a snapshot of the development of a young and vibrant field. The models we describe are still playing catch-up with the flood of high-throughput biological data,[46] but they are already able to set a bar for the quality of such data that will be needed for useful quantification of signaling.[47] Given the rapid pace of development in the area, it will be very interesting to compare this snapshot with the state of the field in coming years.

Acknowledgements

We thank Apoorva Bhandari for comments on the manuscript. U.S.B. is supported by NCBS and the Wellcome Trust. S.J.V. is supported in part by a NIGMS grant GM-54508 to Ravi Iyengar, Mount Sinai School of Medicine. S.M.A. is supported by NCBS/TIFR and a career development award from the TIFR Alumni Association.

Keywords: bioinformatics · MAPK cascade · signal transduction · simulations · systems biology

[1] D. Butler, *Nature* **1999**,*402*, 6761; *Suppl*: C67–70.
[2] a) T. Ideker, T. Galitski, L. Hood, *Annu. Rev. Genomics Hum. Genet.* **2001**, *2*, 343–372; b) S. R. Neves, R. Iyengar, *Bioessays* **2002**, *24*(12), 1110–1117; c) D. Noble, *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 459–463; d) E. Alm, A. P. Arkin, *Curr. Opin. Struct. Biol.* **2003**, *13*, 193–202; e) U. S. Bhalla, *Prog. Biophys. Mol. Biol.* **2003**, *81*, 45–65.
[3] S. Sivakumaran, S. Hariharaputran, J. Mishra, U. S. Bhalla, *Bioinformatics* **2003**, *19*, 408–415.
[4] J. Z. Tsien, P. T. Huerta, S. Tonegawa, *Cell* **1996**, *87*, 1327–1338.
[5] G. C. Kennedy, *Proc. R. Soc. London Ser. B* **1950**, *137*, 535–549.
[6] D. L. Coleman *Diabetologia* **1978**, *14*, 141–148.
[7] Y. Zhang, R. Proenca, M. Maffei, M. Barone, L. Leopold, J. M. Friedman, *Nature* **1994**, *372*, 425–432.
[8] L. A. Tartaglia, M. Dembski, X. Weng, N. Deng, J. Culpepper, R. Devos, G. J. Richards, L. A. Campfield, F. T. Clark, J. Deeds, C. Muir, S. Sanker, A. Moriarty, K. J. Moore, J. S. Smutko, G. G. Mays, E. A. Woolf, C. A. Monroe, R. I. Tepper, *Cell* **1995**, *83*, 1263–1271.
[9] J. J. Tyson, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 7328–7332.
[10] Y. Masui, P. Wang, *Biol. Cell* **1998**, *90*, 537–548.
[11] M. A. Hunt, *The Cell Cycle: An Introduction*, Oxford University Press, **1993**.
[12] A. Hoffmann, A. Levchenko, M. L. Scott, D. Baltimore, *Science* **2002**, *298*, 1241–1245.
[13] B. N. Kholodenko, *Eur. J. Biochem*. **2000**, *267*, 1583–1588.
[14] R. A. Akhtar, A. B. Reddy, E. S. Maywood, J. D. Clayton, V. M. King, A. G. Smith, T. W. Gant, M. H. Hastings, C. P. Kyriacou, *Curr. Biol.* **2002**, *12*, 540–550.
[15] G. E. Duffield, J. D. Best, B. H. Meurers, A. Bittner, J. J. Loros, J. C. Dunlap, *Curr. Biol.* **2002**, *12*, 551–557.
[16] N. G. Ahn, E. G. Krebs, *J. Biol. Chem*. **1990**, *265*, 11495–11501.
[17] a) N. G. Ahn, R. Seger, R. L. Bratlien, C. D. Diltz, N. K. Tonks, E. G. Krebs, *J. Biol. Chem*. **1991**, *266*, 4220–4227; b) N. Gomez, P. Cohen, *Nature* **1991**, *353*, 170–173.
[18] a) J. M. Kyriakis, H. App, X. F. Zhang, P. Banerjee, D. L. Brautigan, U. R. Rapp, J. Avruch, *Nature* **1992**, *358*, 417–421; b) T. Itoh, K. Kaibuchi, T. Masuda, T. Yamamoto, Y. Matsuura, A. Maeda, K. Shimizu, Y. Takai, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 975–979.
[19] a) K. Y. Choi, B. Satterberg, D. M. Lyons, E. A. Elion, *Cell* **1994**, *78*, 499–512; b) S. Marcus, A. Polverino, M. Barr, M. Wigler, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 7762–7766.
[20] C. Y. Huang, J. E. Ferrell, Jr., *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10078–10083.
[21] J. E. Ferrell, Jr., R. R. Bhatt, *J. Biol. Chem.* **1997**, *272*, 19008–19016.
[22] W. R. Burack, T. W. Sturgill, *Biochemistry* **1997**, *36*, 5929–5933.

[23] J. E. Ferrell, Jr., E. M. Machleder, *Science* **1998**, *280*, 895 – 898.
[24] U. S. Bhalla, R. Iyengar, *Science* **1999**, *283*, 381 – 387.
[25] B. N. Kholodenko, O. V. Demin, G. Moehren, J. B. Hoek, *J. Biol. Chem.* **1999**, *274*, 30 169 – 30 181.
[26] A. Levchenko, J. Bruck, P. W. Sternberg, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5818 – 5823.
[27] F. A. Brightman, D. A. Fell, *FEBS Lett.* **2000**, *482*, 169 – 174.
[28] A. R. Asthagiri, D. A. Lauffenburger, *Biotechnol Prog* **2001**, *17*, 227 – 239.
[29] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, G. Muller, *Nat. Biotechnol.* **2002**, *20*, 370 – 375.
[30] U. S. Bhalla, P. T. Ram, R. Iyengar, *Science* **2002**, *297*, 1018 – 1023.
[31] P. S. Swain, E. D. Siggia, *Biophys. J.* **2002**, *82*, 2928 – 2933.
[32] O. J. Somsen, M. Siderius, F. F. Bauer, J. L. Snoep, H. V. Westerhoff, *J. Theor. Biol.* **2002**, *218*, 343 – 354.
[33] H. Resat, J. A. Ewald, D. A. Dixon, H. S. Wiley, *Biophys. J.* **2003**, *85*, 730 – 743.
[34] 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, *Biochem. J.* **2003**, *373*, 451 – 463.
[35] W. Xiong, J. E. Ferrell, Jr., *Nature* **2003**, *426*, 460 – 465.
[36] Y. Gong, X. Zhao, *FEBS Lett.* **2003**, *554*, 467 – 472.
[37] N. I. Markevich, J. B. Hoek, B. N. Kholodenko, *J Cell. Biol.* **2004**, *164*, 353 – 359.
[38] S. Chapman, A. R. Asthagiri, *Biotechnol. Bioeng.* **2004**, *85*, 311 – 322.
[39] J. P. A. S. Subramaniam, *Curr. Opin. Biotechnol.* **2004**, *15*, 78 – 81.
[40] G. Von Dassow, E. Meir, E. M. Munro, G. M. Odell, *Nature* **2000**, *406*, *6792*: 188 – 192.
[41] M. D. Levin, C. J. Morton-Firth, W. N. Abouhamad, R. B. Bourret, D. Bray, *Biophys. J.* **1998**, *74*, 175 – 181.
[42] a) S. Kalir, J. Mcclure, K. Pabbaraju, C. Southward, M. Ronen, S. Leibler, M. G. Surette, U. Alon, *Science* **2001**, *292*, 2080 – 2083; b) A. C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, G. Superti-Furga, *Nature* **2002**, *415*, 141 – 147; c) E. Lee, A. Salic, R. Kruger, R. Heinrich, M. W. Kirschner, *PloS. Biol.* **2003**, *1*, E10; d) T. Bouwmeester, A. Bauch, H. Ruffner, P. O. Angrand, G. Bergamini, K. Croughton, C. Cruciat, D. Eberhard, J. Gagneur, S. Ghidelli, C. Hopf, B. Huhse, R. Mangano, A. M. Michon, M. Schirle, J. Schlegl, M. Schwab, M. A. Stein, A. Bauer, G. Casari, G. Drewes, A. C. Gavin, D. B. Jackson, G. Joberty, G. Neubauer, J. Rick, B. Kuster, G. Superti-Furga, *Nat. Cell. Biol.* **2004**, *6*, 97 – 105.
[43] a) L. H. Hartwell, J. J. Hopfield, S. Leibler, A. W. Murray, *Nature* **1999**, *402*, 6761, *Suppl*: C47–52; b) J. Hasty, D. Mcmillen, F. Isaacs, J. J. Collins, *Nat. Rev. Genet.* **2001**, *2*, 268 – 279.
[44] T. M. Yi, Y. Huang, M. I. Simon, J. Doyle, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4649 – 4653.
[45] a) A. Becskei, L. Serrano, *Nature* **2000**, *405*, 590 – 593; b) M. B. Elowitz, S. Leibler, *Nature* **2000**, *403*, 335 – 338; c) T. S. Gardner, C. R. Cantor, J. J. Collins, *Nature* **2000**, 403, 339 – 342; d) M. R. Atkinson, M. A. Savageau, J. T. Myers, A. J. Ninfa, *Cell* **2003**, *113*, 597 – 607.
[46] a) J. J. Tyson, K. Chen, B. Novak, *Nat. Rev. Mol. Cell. Biol.* **2001**, *2*, 908 – 916; b) H. Kitano, *Science* **2002**, *295*, 1662 – 1664.
[47] H. Husi, M. A. Ward, J. S. Choudhary, W. P. Blackstock, S. G. Grant, *Nat. Neurosci.* **2000**, *3*, 661 – 669.

Received: April 30, 2004

