

New Connections, New Components, Real Dynamics

Joel S. Bader*

Department of Biomedical Engineering and High-Throughput Biology Center, Johns Hopkins University, Baltimore, MD 21218, USA

*Corresponding author. E-mail, joel.bader@jhu.edu

Abstract

The best-studied signaling pathways still hold secrets. Recent studies have now applied a new wave of technologies encompassing computational approaches and experimental techniques to the mitogen-activated protein kinase pathway in yeast and have provided new knowledge of pathway connections, components, and dynamics. The computational algorithms build on advances in network science motivated by studies of large-scale social and WWW networks. Experimental techniques permit exploration of the frequency-space response, describing biological signaling networks in the language of control theory. Together, these technologies are revealing the design choices made by evolution, and they provide a framework for building new biological circuits to order.

New technologies, represented by computational work by Huang and Fraenkel (1) and experimental studies by Macia *et al.* (2) provide new insight into the network structure and dynamics of the well-studied yeast mitogen-activated protein kinase (MAPK) pathways. The computational challenge undertaken by Huang and Fraenkel is to infer hidden components and connections in a signaling pathway based on incomplete

data from high-throughput screens (1). Beyond standard problems with interpreting noisy data, crucial pathway components can be invisible to specific technologies. Transcript profiling identifies the genes targeted by a signaling pathway, but the signaling proteins themselves often show no transcriptional response. Genetic screens can identify essential components of a pathway, but partially redundant pathway branches can mask the effects of individual genes. Proteomic technologies remain limited in their ability to comprehensively scan all protein abundances, modifications, and activities.

All is not lost, however, because databases of protein-protein interactions and transcription factor–DNA binding sites can suggest hidden components on the basis of their connections to observed components. Extracting the hidden components based on the observed components has been termed the “active subnetwork” problem (3). Huang and Fraenkel’s approach uses prize-collecting Steiner trees (PCSTs), which are optimal solutions for connecting vertices in a graph (Fig. 1). When observed components are not directly connected to each other in the interaction network, prize-collecting trees connect them indirectly by adding hidden components as intermediates. Each interaction inflicts a cost that reflects its experimental uncertainty (4), and each observed component left out of the network also introduces a penalty. Although this problem is NP-hard—the same complexity class as the traveling salesman problem, finding the shortest tour through a list of cities that visits each city exactly once and returns to the start—an algorithmic advance provides optimal solutions for problems scaling to

thousands of genes (5). The beauty of this contribution is the ability to identify the exact solution to the optimization problem, as opposed to approximate solutions generated by less-sophisticated methods. Indeed, approximate solutions give inferior results for identifying functional modules in protein-interaction networks (6).

This PCST algorithm adds to a growing arsenal of network-inspired methods, such as those using “flow” to assess connectivity. The flow between two vertices in a network is roughly equivalent to the number of paths connecting them, with a bias toward short paths of high-confidence interactions. Flow does not necessarily represent actual flow of materials, although flow-based stoichiometric models, in which fluxes of signaling molecules are explicitly represented, have been introduced for signaling (7). Flow is also distinct from probabilistic models that represent the physical state of network components, such as Bayesian networks for MAPK signaling (8).

Network flow came to systems biology by way of original applications to the link structure of hypertext documents (9), which in turn motivated the PageRank algorithm in the Google search engine (10). An early biological application predicted protein function using flow over a protein-protein interaction network (11). Extensions of flow algorithms to positive and negative epistatic interactions (12) may be relevant for future applications to interactions representing likes and dislikes in social networks.

New applications of flow to signaling networks consider paths from “sources”, upstream genes in a

regulatory network, to “sinks” representing the downstream transcriptional output of the network. Flow algorithms have improved the analysis of expression quantitative trait locus (eQTL) data (13), which identify genetic variants that control expression levels of specific transcripts, and have been applied to MAPK signaling (14) to identify likely new components of the high-osmolarity glycerol (HOG) signaling pathway involved in yeast response to osmotic stress.

Macia *et al.* exploit the knowledge of pathway structure to probe its dynamics (2). One of the puzzles of the HOG MAPK pathway in yeast is the existence of two independent routes to pathway activation, one fast-acting and the other time-delayed. Macia *et al.* measure the response of each pathway branch to step-function inputs. They discover that the fast-acting branch has a remarkable amount of basal activity, causing weak activation of downstream components even in the absence of stimulation (Fig. 2). Leaky activation is usually dismissed as a bug, but here it is a feature that combines with negative feedback to generate a more rapid response to environmental cues. High basal activity controlled by negative feedback is also present in the mating pheromone and filamentous growth MAPK pathways, which share components with the HOG signaling pathway (15).

The response of MAPK pathways to step inputs has revealed additional negative-feedback loops in the time-delay branch of HOG signaling (16). Negative feedback is also present in the mating-response MAPK pathway (17). Inhibitory signals are sent between different MAPK pathways, with the cross talk creating bistability and

potentially stochastic cell-fate decisions (18–20). High basal activity regulated by negative feedback may be an important, generic, and hitherto unrecognized property of MAPK signaling.

Step-function perturbations, as used by Macia *et al.*, provide constraints on pathway activation and deactivation time scales, but much richer information comes from the response to driving the system at a range of frequencies. In the ideal case, a system's response function can be used as a black-box predictor of its dynamic response to any possible input. Calculating the response or transfer function, and then mapping elements of the response to network components, is called systems identification. Systems identification can be powerful even if only steady-state data are available for modeling, with perturbations generated by drug treatment or genetic variants (21).

Two groups have now reported the frequency response of HOG signaling by measuring the response to trains of square waves generated in microfluidic devices (22, 23). These studies, couched in the language of engineering control theory, reveal new properties of the signaling branches: The fast-acting branch behaves like a low-pass filter and integrates signals above its bandwidth; the slow-acting branch rejects high-frequency input entirely (Fig. 2). Individual terms in the response can be mapped to protein components and reactions in the pathway. As promised, the response function quantitatively predicts the pathway response to time-domain input.

These studies begin to achieve the vision of genome biology: using our knowledge of the gene and protein components of life to predict and shape

the behavior of living systems for medical and engineering applications. Detailed models of cellular networks, including the yeast MAPK networks highlighted here (24), are constantly improving as platforms for rapid *in silico* hypothesis generation and testing. Accurate models will be useful to predict therapeutic targets or to assess the systems-level consequences of individual variation in gene and protein activity. These models must grapple with bistable or multi-stable responses, strategies often selected by evolution to respond to a chaotic environment. Tinkering and hacking with biological networks are now possible through synthetic biology, with discrete signaling elements cut from existing networks and grafted into new contexts. Biological modules providing discrete functions, most recently counters (25), have been developed, and attempts to compose modules into larger synthetic networks are under way. The open question in this field is when we will be able to design a desired network, formally specified by a response function, from scratch.

Fig. 1. A PCST represents genes and proteins as vertices connected by edges representing different types of interactions. The vertices have prizes, indicated by dollar signs, representing the strength of evidence relating them to a specific phenotype. Evidence may come from different sources. Edges represent known interactions between genes and proteins, and the question marks indicate uncertainty in the evidence supporting an edge. Filled-in vertices represent genes or proteins that lack direct experimental evidence for association with a phenotype. Some of these may, nonetheless, be part of a valid network model but invisible to the technologies used. The goal of the

Steiner tree algorithm is to identify the subnetwork that maximizes profit, here defined as the dollars gained by including detected vertices minus the cost represented by the question marks. The network can include filled-in vertices lacking evidence, and it can also exclude detected vertices that are not connected by strong enough evidence to the rest of the network. The heavy lines indicate edges that are part of the optimal network. Three vertices with no direct evidence (in boxes) are part of the optimal network. Two detected transcripts (grayed out) are excluded from the model.

Fig. 2. The HOG pathway has two upstream sensors, Sln1 and Sho1, whose signals are integrated by Pbs2 to activate Hog1. The Sln1 branch has weak activity, even in the absence of signaling, demonstrated by Macia *et al.* (2). The leaky activity is a feature conferring fast response, rather than a design flaw, and is kept in check by feedback inhibition involving Hog1 through an as-yet-unknown mechanism. The Sho1 branch is also inhibited by negative feedback, discovered by Hao *et al.* with a hypothesis generated by Bayesian network modeling (16). Other researchers have treated the HOG pathway as a black box, characterizing the response of each branch to step function and periodic inputs. The fast-responding Sln1 branch responds proportionally to dose and integrates fast signals. The slow-responding Sho1 branch responds to high doses and slow inputs, but rejects low doses and fast inputs. Blue lines represent data from Macia *et al.* (2), and orange lines represent data from Hao *et al.* (16).

References

1. S. C. Huang, E. Fraenkel, Hidden components of signaling and regulatory networks are revealed by integrating proteomic, transcriptional, and interactome data. *Sci. Signal.* **2**, ra40 (2009).
2. J. Macia, S. Regot, T. Peeters, N. Conde, R. Solé, F. Posas, Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction. *Sci. Signal.* **2**, ra13 (2009).
3. T. Ideker, O. Ozier, B. Schwikowski, A. F. Siegel, Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18**, S233–S240 (2002).
4. J. S. Bader, A. Chaudhuri, J. Rothberg, J. Chant, Gaining confidence in high-throughput protein interaction networks. *Nat. Biotechnol.* **22**, 78–85 (2004).
5. I. Ljubic, R. Weiskircher, U. Pferschy, G. W. Klau, P. Mutzel, M. Fischetti, An algorithmic framework for the exact solution of the prize-collecting Steiner tree problem. *Math. Program.* **105**, 427–449 (2006).
6. M. T. Dittrich, G. W. Klau, A. Rosenwald, T. Dandekar, T. Müller, Identifying functional modules in protein-protein interaction networks: An integrated exact approach. *Bioinformatics* **24**, i223–i231 (2008).
7. M. S. Dasika, A. Burgard, C. D. Maranas, A computational framework for the topological analysis and targeted disruption of signal transduction networks. *Biophys. J.* **91**, 382–398 (2006).

8. I. Gat-Viks, R. Shamir, Refinement and expansion of signaling pathways: The osmotic response network in yeast. *Genome Res.* **17**, 358–367 (2007).
9. J. M. Kleinberg, Authoritative sources in a hyperlinked environment. *JACM* **46**, 604–632 (1999).
10. S. Brin, L. Page, The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* **30**, 107–117 (1998).
11. E. Nabieva, K. Jim, A. Agarwal, B. Chazelle, M. Singh, Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. *Bioinformatics* **21**, i302–i310 (2005).
12. Y. Qi, Y. Suhail, Y. Lin, J. Boeke, J. Bader, Finding friends and enemies in an enemies-only network: A graph diffusion kernel for predicting novel genetic interactions and co-complex membership from yeast genetic interactions. *Genome Res.* **18**, 1991–2004 (2008).
13. S. Suthram, A. Beyer, R. M. Karp, Y. Eldar, T. Ideker, eQED: An efficient method for interpreting eQTL associations using protein networks. *Mol. Syst. Biol.* **4**, 162 (2008).
14. E. Yeger-Lotem, L. Riva, L. J. Su, A. D. Gitler, A. G. Cashikar, O. D. King, P. K. Auluck, M. L. Geddie, J. S. Valastyan, D. R. Karger, S. Lindquist, E. Fraenkel, Bridging high-throughput genetic and transcriptional data reveals cellular responses to α -synuclein toxicity. *Nat. Genet.* **41**, 316–323 (2009).
15. M. A. Schwartz, H. D. Madhani, Principles of MAP kinase signaling specificity in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* **38**, 725–748 (2004).
16. N. Hao, M. Behar, S. C. Parnell, M. P. Torres, C. H. Borchers, T. C. Elston, H. G. Dohlman, A systems-biology analysis of feedback inhibition in the Sho1 osmotic-stress-response pathway. *Curr. Biol.* **17**, 659–667 (2007).
17. R. P. Bhattacharyya, A. Reményi, M. C. Good, C. J. Bashor, A. M. Falick, W. A. Lim, The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science* **311**, 822–826 (2006).
18. M. N. McClean, A. Mody, J. R. Broach, S. Ramanathan, Cross-talk and decision making in MAP kinase pathways. *Nat. Genet.* **39**, 409–414 (2007).
19. S. Paliwal, P. A. Iglesias, K. Campbell, Z. Hilioti, A. Groisman, A. Levchenko, MAPK-mediated bimodal gene expression and adaptive gradient sensing in yeast. *Nature* **446**, 46–51 (2007).
20. R. J. Taylor, D. Falconnet, A. Niemisto, S. A. Ramsey, S. Prinz, I. Shmulevich, T. Galitski, C. L. Hansen, Dynamic analysis of MAPK signaling using a high-throughput microfluidic single-cell imaging platform. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3758–3763 (2009).

21. T. S. Gardner, D. di Bernardo, D. Lorenz, J. J. Collins, Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
22. P. Hersen, M. N. McClean, L. Mahadevan, S. Ramanathan, Signal processing by the HOG MAP kinase pathway. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7165–7170 (2008).
23. J. T. Mettetal, D. Muzzey, C. Gómez-Uribe, A. Van Oudenaarden, The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*. *Science* **319**, 482–484 (2008).
24. E. Klipp, B. Nordlander, R. Krüger, P. Gennemark, S. Hohmann, Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.* **23**, 975–982 (2005).
25. A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church, J. J. Collins, Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).



