COMPUTATIONAL BIOLOGY

Understanding Modularity in Molecular Networks Requires Dynamics

Roger P. Alexander,[1,2,3] Philip M. Kim,[4,5,6,7] Thierry Emonet,[1,2,8]
Mark B. Gerstein[1,3,9*]

The era of genome sequencing has produced long lists of the molecular parts from which cellular machines are constructed. A fundamental goal in systems biology is to understand how cellular behavior emerges from the interaction in time and space of genetically encoded molecular parts, as well as nongenetically encoded small molecules. Networks provide a natural framework for the organization and quantitative representation of all the available data about molecular interactions. The structural and dynamic properties of molecular networks have been the subject of intense research. Despite major advances, bridging network structure to dynamics—and therefore to behavior—remains challenging. A key concept of modern engineering that recurs in the functional analysis of biological networks is modularity. Most approaches to molecular network analysis rely to some extent on the assumption that molecular networks are modular—that is, they are separable and can be studied to some degree in isolation. We describe recent advances in the analysis of modularity in biological networks, focusing on the increasing realization that a dynamic perspective is essential to grouping molecules into modules and determining their collective function.

Frameworks for Organizing Molecular Interactions

Because of the information available to date, much work has been done to define modules based on properties of network structure. A structural module can be viewed as a static representation of all the interactions that are possible between the elements of a module. In this case, the analysis of aggregated data sources provides enough information to define modules. However, knowledge of network structure is often not sufficient to infer function, and dynamic modularity can exist in the absence of structural modularity. A clear understanding of dynamic modularity emerges only when the time-dependent activity of molecular networks is monitored. As more dynamical data become available, it will be possible to complement our understanding of the structure of molecular networks with an understanding of their dynamics, and thereby to understand more about biological function and how interactions between molecules generate cellular phenotypes.

Analysis of Network Structure

Generating data on molecular interactions at the genome scale is still a difficult enterprise. As of 2008, only about 20% of the predicted interactome in yeast has been experimentally characterized (1). Time-resolved molecular interaction data are even scarcer, so a substantial amount of research effort has been expended elucidating biological function from structural parameters of molecular interaction networks (2). The first step in finding modules from protein interaction data includes finding fully connected subnetworks or cliques, or defective cliques that have lower connectivity (3, 4). Proteins on the shortest path between many pairs of nodes in the network, called bottlenecks, often represent interfaces between modules (5), whereas highly connected nodes or hubs often lie in the center of modules (6) (Fig. 1).

Here, we highlight examples of structural features of molecular networks for which the importance of dynamics is evident even in the absence of dynamical information. First, by mapping known three-dimensional structures of interacting proteins onto the yeast protein interaction network, two classes of hubs with distinct dynamical behavior were identified (7). Multi-interface hubs have several interaction surfaces and interact with multiple partners simultaneously, anchoring them into complexes that are stable across multiple conditions and time points. Single-interface hubs have one or sometimes two interaction surfaces and interact with partners one or two at a time. Multi- and single-interface hubs, defined by these biophysical properties, resemble party and date hubs, respectively, which were defined on the basis of their different signatures in gene expression data (8). Distinguishing conclusively between these two classes of hubs requires dynamic information.

The analysis of network motifs, pioneered by Alon (9), provides information about structural modularity on the scale of a few interacting molecules. Network motifs can be recognized from network structure alone by finding groups of molecules connected in the network more often than would be expected at random. The feed-forward loop (FFL) and single-input module (SIM) are classic examples where just knowing network structure can provide insight into dynamics and function (Fig. 1). For example, FFLs often play a role in noise filtering, turning on gene transcription only in response to sustained input signals (10). SIMs are often present in terminal differentiation cascades, where they lead to expression of a battery of functionally related genes in response to an activation signal (11). However, for another common network motif, the bifan, it is difficult to infer anything about its function from structure alone: Depending on the strength and sign of the interactions between nodes in the bifan motif, completely different dynamical behavior can result (12) (Fig. 1). A systematic analysis of network motif dynamics showed that several classes of real biological networks are enriched with network motifs that are robust to changes in dynamical parameters, like the FFL and SIM. The same networks were found to be depleted in network motifs that exhibit widely varying behavior in response to perturbation of their dynamical parameters, like the bifan (13). This work suggests that natural selection acts on network dynamics and that network structure and dynamics are coupled.

Examination of condition-dependent expression of network motifs in the yeast regulatory network found some conditions enriched in FFLs and other conditions enriched in SIMs (11). In conditions that monitor the cell's internal state (endogenous conditions), FFLs buffer multiple layers of internal signal processing,

[1]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA. [2]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520, USA. [3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA. [4]Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada. [5]Banting and Best Department of Medical Research, University of Toronto, Toronto, ON M5S 3E1, Canada. [6]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3E1, Canada. [7]Department of Computer Science, University of Toronto, Toronto, ON M5S 3E1, Canada. [8]Department of Physics, Yale University, New Haven, CT 06520, USA. [9]Department of Computer Science, Yale University, New Haven, CT 06520, USA.

*Corresponding authors. E-mail, thierry.emonet@yale.edu (T.E.); mark.gerstein@yale.edu (M.B.G.)

whereas in conditions that involve external stimuli (exogenous conditions), SIMs generate rapid responses to external conditions by expressing batteries of output genes appropriate to the condition. This work took advantage of the functional information inherent in the structure of FFLs and SIMs to predict dynamical behavior in different situations. It also revealed that hubs in some conditions are not hubs in others, showing that important features of network structure are transient and condition specific (*11*). Another dynamic extension of the concept of network motifs is the activity motif. There, activity is defined using some quantitative functional genomics data set that is overlaid onto the network structure. The network is then searched for series of nodes with quantitative patterns that are present more often than would be expected at random (*14*). For example, a chain of genes with decreasing transcription factor binding affinity might represent a just-in-time transcription program: The first gene with strongest affinity gets activated first, and so on, in a linear cascade.

Dynamic Modularity

These glimpses of dynamic signatures embedded in structural modules have led researchers toward the realization that network dynamics must be studied more systematically. The dynamics of molecular interactions can be studied at the genome scale by integrating time series of high-throughput data sets or data sets generated under various conditions. Even in the data-rich environment of functional genomics, however, the problem of relating network structure to dynamics remains very challenging. The problem is largely underdetermined because there are always more network parameters than data points available to constrain them. Accordingly, the necessary first step in such methods is to group molecules into modules and study the dynamic interactions between modules (*15*). Great care in designing the experiments helps to maximize the amount of information provided by these methods. Modular response analysis (MRA) is a classic framework for designing experiments to determine molecular
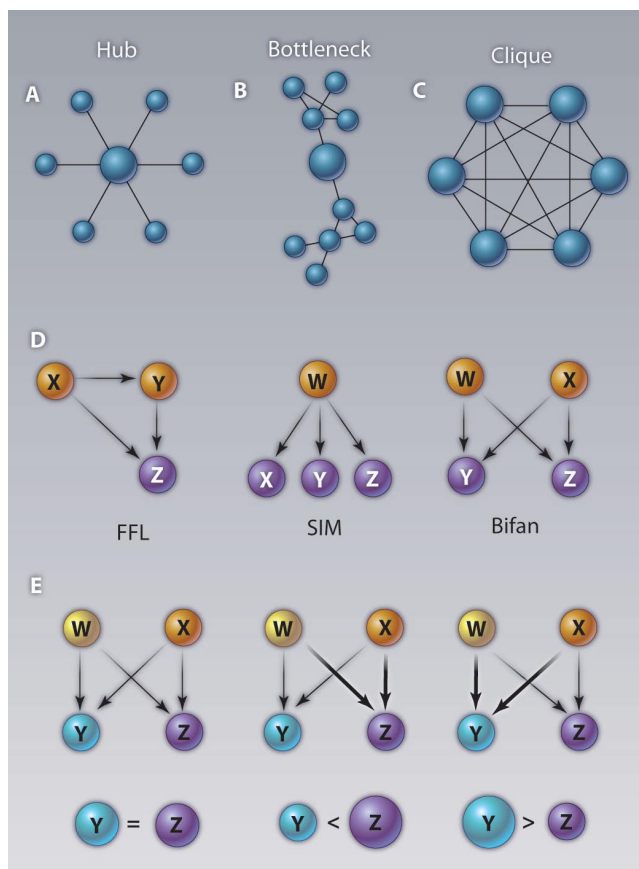

Fig. 1. Hubs, bottlenecks, cliques, and network motifs are features of the topological structure of molecular networks that have been well studied and can sometimes be used to infer function. (**A**) Hubs are nodes in the network with more interaction partners than average; they tend to be essential. (**B**) Bottlenecks are nodes that lie on the shortest path between many pairs of nodes in the network; they often lie at the interface between modules and also tend to be essential. (**C**) Cliques in protein interaction networks are indicative of protein complexes that function together. (**D**) The feed-forward loop (FFL) and single-input module (SIM) are network motifs with output behavior that varies little when dynamical parameters are changed, whereas (**E**) the bifan motif can generate widely different dynamics from different dynamical parameter combinations (*12, 13*).

network dynamics (*15*). On the basis of prior knowledge, a biological system is partitioned into functional modules such that each module can send and receive just one input from each other module. A series of experiments perturb each module separately and the framework of MRA is applied to the experimental results, thus enabling the strength of interaction between all the modules to be quantified, generating a dynamic modular network.

Although modularization reduces the size of the network and makes it less underdetermined, it presupposes that molecular networks can, in fact, be treated as separable modules (*16, 17*). An alternative view is that molecular interactions in the cell cannot be broken down

into structural modules and that there is a large amount of crosstalk between molecules with unrelated functions. These two possibilities might reflect different evolutionary conditions. Nonmodular networks are more likely to evolve in stable environments, whereas modular networks are likely to evolve in organisms challenged by multiple environments (*18, 19*). The concepts of retroactivity and kinetic insulation provide a framework for understanding dynamic modularity.

Retroactivity

The concept of retroactivity emerged in the field of electric engineering. When electronic modules are connected in series, the dynamical properties of a downstream element may retroactively affect the functioning of an upstream element—for example, by draining electrons too fast. An electrical module with low retroactivity can be mixed and matched with others to build more complex circuits in a modular way. Like electrical circuit elements, dynamic biological modules with low retroactivity may be mixed and matched in various combinations without altering the internal dynamics of each module. It is possible that molecular networks evolved to exhibit low retroactivity and improve their dynamic modularity. New methods to analyze retroactivity in molecular networks are emerging: A theoretical framework for such analysis has been proposed (*20*), and an algorithm was developed to partition molecular networks such that retroactivity between modules is minimized (*21*). Certainly, functional modules designed for synthetic biology must exhibit low retroactivity, but whether natural molecular networks have evolved the same characteristic is an open question.

Although MRA can aid in the generation of dynamic modular networks, the amount of retroactivity between modules cannot be determined with MRA, because it does not provide a mechanism to determine whether feedback from one module to another alters the internal dynamics of the receiving module. The only way to determine retroactivity between modules is to generate more detailed data about the dynamical parameters of the nodes inside the modules. As more such dynamical data become available, it will be possible to answer the question of whether nature evolves

dynamical modules that are dynamically isolated by minimizing retroactivity. If the answer is no, that some functional modules are not dynamically isolated from each other, then kinetic insulation provides a way to reduce crosstalk by separating the time scales of dynamic signals that pass through the module from different inputs (*22*).

Kinetic Insulation
Both theoretical and experimental analyses indicate that a single molecular network can generate distinct outputs in response to input signals with different dynamical profiles. This phenomenon has been termed "kinetic insulation" (*22*), emphasizing the importance of dynamics to signaling specificity. For example, a progenitor cell in a tissue differentiation cascade can turn on different batteries of genes in response to input signals with different dynamic signatures. A classic example is the case in which a mitogen-activated protein kinase (MAPK) cascade responds transiently to epidermal growth factor (EGF) but exhibits a sustained response to neural growth factor (NGF) (*23*). The PC12 cancer cell line is derived from adrenal gland tissue and can be thought of as an imperfect model of adrenal gland progenitor cells, which have the ability to differentiate into either sympathetic neurons or adrenal chromaffin cells that synthesize adrenaline. NGF generates sustained activation of MAPK in PC12 cells, leading to a specific gene expression program that causes the cells to differentiate into neuronal cells. In contrast, EGF activates the same MAPK only transiently, triggering a gene expression program that causes the cells to proliferate, presumably allowing them to differentiate into adrenaline-producing cells in response to a later cue. This tissue-specification program is only possible because of the expression in the progenitor cell of different batteries of genes in response to different input signals. The MAPK cascade uses kinetic insulation to process different dynamic signals through the same set of molecules to generate different responses (*23*). The nuclear factor κB (NF-κB) regulatory network that is important in the mammalian inflammation response also exhibits kinetic insulation, producing transient or sustained activation of the NF-κB transcription factor in response to tumor necrosis factor α (TNF-α) or lipopolysaccharide (LPS) input signals, respectively (*24*).

Determining Dynamical Parameters of Cellular Networks
Dynamical models of molecular networks rely on knowledge of biochemical parameters, such as enzymatic rate constants, binding affinities, and protein concentrations, determined over many decades of biochemical and molecular biological experiments. The assumption that a set of well-studied molecules can be treated as an isolated system is equivalent to the contention that they represent a functional module and that the module dynamics will not be strongly affected by crosstalk with other molecules that have not been studied in the same functional context. There are a growing number of dynamical models of well-characterized molecular networks (*25*), which may reveal biological insight, keeping in mind the caveat that the assumption of modularity and lack of crosstalk may not hold in the face of future data.
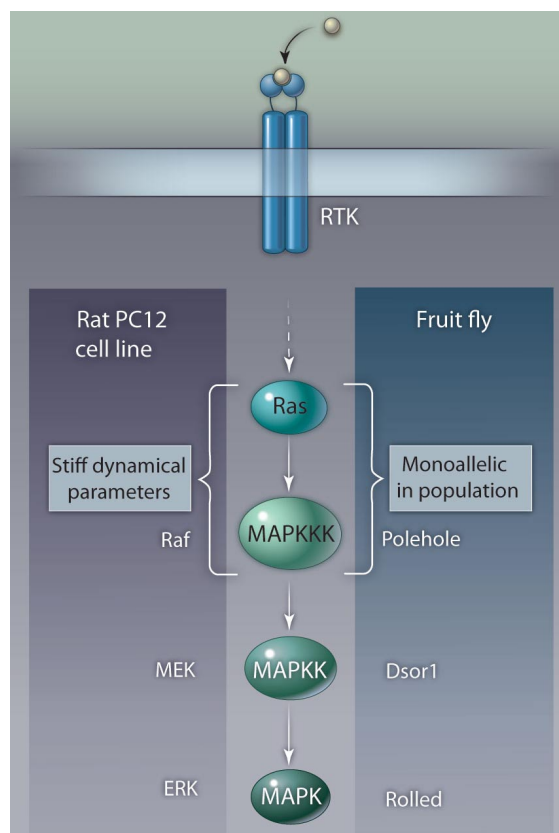

Fig. 2. Dynamical constraints on a signaling pathway involved in tissue differentiation in rat correlate with allelic variation at the population level in the same pathway in fruit flies, revealing a link between dynamics and evolution. The "stiffest" combination of parameters was associated with the Ras and Raf proteins at the top of the pathway, and in fruit flies Ras and Raf exhibit the least allelic variation of all of the components of the pathway, consistent with the concept that stiff parameter combinations are constrained during evolution. The dashed arrow indicates that some accessory components of the two cascades are excluded here for clarity of presentation. RTK, receptor tyrosine kinase.

The behavior of a dynamical model might be insensitive to changes in individual parameters but remain sensitive to changes in combinations of those parameters. For example, the ratio of activity of a kinase and phosphatase acting on the same target might be constrained while allowing the abundance and rates of the two enzymes to vary in concert over a wide range (*26*). Sensitivity analysis is a mathematical tool for identifying "stiff" dynamical parameters, which are combinations of biochemical parameters that when changed strongly affect the dynamical behavior of the model. Interestingly, most dynamical models in systems biology are "sloppy," meaning that only a few combinations of parameters are stiff and the other parameters can vary over an extremely wide range without altering output behavior (*25*). A dynamical model of the MAPK cascade that controls PC12 cell differentiation exhibits this characteristic (*23*). The stiffest dynamical parameter in the model is composed of biochemical parameters affecting the interaction between the Ras and Raf proteins, meaning that altering that interaction slightly has a strong effect on dynamical behavior (Fig. 2).

Currently, most systems biology models are parameterized by reference to the extensive biochemical literature that has developed over many decades from studies of one or a few gene products in any given experiment. Knowing many parameter values accurately provides enormous predictive power, as recently demonstrated by a model of cytokinesis in yeast that relied on characterization of the biochemical parameters of interactions between actin, myosin, and associated proteins (*27*). It might seem natural to extend this practice to the genome scale by developing high-throughput methods to measure individual biochemical parameters one at a time. However, the sloppy character of dynamical network models suggests that not all parameters, even if accurately measured, will provide the same amount of constraint on the model. In this case, using sensitivity analysis on the model and fitting all of the parameters at once to global time series data may iden-

tify which combinations of parameters are most important to measure.

Robustness and Evolvability

The uneven stiffness of dynamical parameters has interesting implications for the robustness and evolvability of molecular networks. The set of nonstiff dynamical parameters represents a "neutral space" where network phenotype is robust to parameter changes generated by molecular evolutionary processes (*28*). Changes in stiff dynamical parameters should be more strongly constrained by selection.

Reconstructing evolutionarily related families of molecular networks might reveal which combinations of parameters are sensitive and which are insensitive based on the extent of variation at the population level. An interesting example is again the network representing the response of rat adrenal gland progenitor cells to neural and epithelial growth factors. Analysis of a model of EGF versus NGF activation of the Ras-MAPK pathway (Ras-Raf-MEK-ERK) showed that the stiffest dynamical parameter was associated with Ras and Raf, the upstream elements of the pathway (*23*). A population genetic study of a homologous MAPK cascade in *Drosophila* showed that the fly Ras and Raf homologs exhibited the least polymorphism across the worldwide population of fruit flies, whereas other components of the pathway were more polymorphic, that is, less constrained (*29*). Thus, a model of a MAPK cascade from rat made predictions about which proteins were dynamically constrained, and these predictions were confirmed at the population level in the homologous pathway in fruit flies (Fig. 2).

The proliferation of data about variability in human populations will be a useful resource for further studies of this type (*30*). Most studies of human population variability focus on variations that affect phenotype, usually associations with disease, but even data about genotypic variability without phenotypic effect are useful, because they reveal which genes in a network are robust to variation in parameters (*31*, *32*).

References and Notes

1. H. Yu, P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.-F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A.-S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. P. Roth, A.-L. Barabasi, J. Tavernier, D. E. Hill, M. Vidal, High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
2. X. Zhu, M. Gerstein, M. Snyder, Getting connected: Analysis and principles of biological networks. *Genes Dev.* **21**, 1010–1024 (2007).
3. G. D. Bader, C. W. Hogue, An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003).
4. H. Yu, A. Paccanaro, V. Trifonov, M. Gerstein, Predicting interactions in protein networks by completing defective cliques. *Bioinformatics* **22**, 823–829 (2006).
5. H. Yu, P. M. Kim, E. Sprecher, V. Trifonov, M. Gerstein, The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS Comput. Biol.* **3**, e59 (2007).
6. H. Jeong, S. P. Mason, A. L. Barabasi, Z. N. Oltvai, Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
7. P. M. Kim, L. J. Lu, Y. Xia, M. B. Gerstein, Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* **314**, 1938–1941 (2006).
8. J. D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, M. Vidal, Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
9. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
10. R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, U. Alon, Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542 (2004).
11. N. M. Luscombe, M. M. Babu, H. Y. Yu, M. Snyder, S. A. Teichmann, M. Gerstein, Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308–312 (2004).
12. P. J. Ingram, M. P. H. Stumpf, J. Stark, Network motifs: Structure does not determine function. *BMC Genomics* **7**, 108 (2006).
13. R. J. Prill, P. A. Iglesias, A. Levchenko, Dynamic properties of network motifs contribute to biological network organization. *PLoS Biol.* **3**, e343 (2005).
14. G. Chechik, E. Oh, O. Rando, J. Weissman, A. Regev, D. Koller, Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat. Biotechnol.* **26**, 1251–1259 (2008).
15. B. N. Kholodenko, A. Kiyatkin, F. J. Bruggeman, E. Sontag, H. V. Westerhoff, J. B. Hoek, Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12841–12846 (2002).
16. R. Bonneau, M. T. Facciotti, D. J. Reiss, A. K. Schmid, M. Pan, A. Kaur, V. Thorsson, P. Shannon, M. H. Johnson, J. C. Bare, W. Longabaugh, M. Vuthoori, K. Whitehead, A. Madar, L. Suzuki, T. Mori, D.-E. Chang, J. DiRuggiero, C. H. Johnson, L. Hood, N. S. Baliga, A predictive model for transcriptional control of physiology in a free living cell. *Cell* **131**, 1354–1365 (2007).
17. D. J. Reiss, N. S. Baliga, R. Bonneau, Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. *BMC Bioinformatics* **7**, 280 (2006).
18. N. Kashtan, U. Alon, Spontaneous evolution of modularity and network motifs. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13773–13778 (2005).
19. M. Parter, N. Kashtan, U. Alon, Environmental variability and modularity of bacterial metabolic networks. *BMC Evol. Biol.* **7**, 169 (2007).
20. D. Del Vecchio, A. J. Ninfa, E. D. Sontag, Modular cell biology: Retroactivity and insulation. *Mol. Syst. Biol.* **4**, 161–176 (2008).
21. J. Saez-Rodriguez, S. Gayer, M. Ginkel, E. D. Gilles, Automatic decomposition of kinetic models of signaling networks minimizing the retroactivity among modules. *Bioinformatics* **24**, i213–i219 (2008).
22. M. Behar, H. G. Dohlman, T. C. Elston, Kinetic insulation as an effective mechanism for achieving pathway specificity in intracellular signaling networks. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16146–16151 (2007).
23. K. S. Brown, C. C. Hill, G. A. Calero, C. R. Myers, K. H. Lee, J. P. Sethna, R. A. Cerione, The statistical mechanics of complex signaling networks: Nerve growth factor signaling. *Phys. Biol.* **1**, 184–195 (2004).
24. R. Cheong, A. Hoffmann, A. Levchenko, Understanding NF-kB signaling via mathematical modeling. *Mol. Syst. Biol.* **4**, 192 (2008).
25. R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, J. P. Sethna, Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189 (2007).
26. A. Goldbeter, D. E. Koshland, An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6840–6844 (1981).
27. D. Vavylonis, J. Q. Wu, S. Hao, B. O'Shaughnessy, T. D. Pollard, Assembly mechanism of the contractile ring for cytokinesis by fission yeast. *Science* **319**, 97–100 (2008).
28. A. Wagner, Robustness and evolvability: A paradox resolved. *Proc. Biol. Sci.* **275**, 91–100 (2008).
29. R. M. Riley, W. Jin, G. Gibson, Contrasting selection pressures on components of the Ras-mediated signal transduction pathway in Drosophila. *Mol. Ecol.* **12**, 1315–1323 (2003).
30. E. Pennisi, Breakthrough of the year: Human genetic variation. *Science* **318**, 1842–1843 (2007).
31. Y. Chen, J. Zhu, P. Y. Lum, X. Yang, S. Pinto, D. J. MacNeil, C. S. Zhang, J. Lamb, S. Edwards, S. K. Sieberts, A. Leonardson, L. W. Castellini, S. Wang, M. F. Champy, B. Zhang, V. Emilsson, S. Doss, A. Ghazalpour, S. Horvath, T. A. Drake, A. J. Lusis, E. E. Schadt, Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
32. P. M. Kim, J. O. Korbel, M. B. Gerstein, Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20274–20279 (2007).

Science Signaling

Understanding Modularity in Molecular Networks Requires Dynamics

Roger P. Alexander, Philip M. Kim, Thierry Emonet and Mark B. Gerstein

ARTICLE TOOLS	http://stke.sciencemag.org/content/2/81/pe44
RELATED CONTENT	http://stke.sciencemag.org/content/sigtrans/2/81/pc14.full http://stke.sciencemag.org/content/sigtrans/3/113/ra20.full http://stke.sciencemag.org/content/sigtrans/3/146/ra79.full http://stke.sciencemag.org/content/sigtrans/4/173/pt3.full http://stke.sciencemag.org/content/sigtrans/2/81/eg10.full http://stke.sciencemag.org/content/sigtrans/2/64/eg3.full
REFERENCES	This article cites 32 articles, 14 of which you can access for free http://stke.sciencemag.org/content/2/81/pe44#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service