

## References and Notes

1. L. Alonso, E. Fuchs, *J. Cell Sci.* **119**, 391 (2006).
2. E. Fuchs, V. Horsley, *Genes Dev.* **22**, 976 (2008).
3. V. Levy, C. Lindon, Y. Zheng, B. D. Harfe, B. A. Morgan, *FASEB J.* **21**, 1358 (2007).
4. S. Claudinot, M. Nicolas, H. Oshima, A. Rochat, Y. Barrandon, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14677 (2005).
5. S. Ghazizadeh, L. B. Taichman, *EMBO J.* **20**, 1215 (2001).
6. M. Ito *et al.*, *Nat. Med.* **11**, 1351 (2005).
7. V. Levy, C. Lindon, B. D. Harfe, B. A. Morgan, *Dev. Cell* **9**, 855 (2005).
8. E. Clayton *et al.*, *Nature* **446**, 185 (2007).
9. V. Horsley *et al.*, *Cell* **126**, 597 (2006).
10. C. S. Trempus *et al.*, *J. Invest. Dermatol.* **120**, 501 (2003).
11. C. Blanpain, W. E. Lowry, A. Geoghegan, L. Polak, E. Fuchs, *Cell* **118**, 635 (2004).
12. R. J. Morris *et al.*, *Nat. Biotechnol.* **22**, 411 (2004).
13. G. Cotsarelis, T. T. Sun, R. M. Lavker, *Cell* **61**, 1329 (1990).
14. K. M. Braun *et al.*, *Development* **130**, 5241 (2003).
15. T. Tumber *et al.*, *Science* **303**, 359 (2004); published online 11 December 2003 (10.1126/science.1092436).
16. M. Ito, K. Kizawa, K. Hamada, G. Cotsarelis, *Differentiation* **72**, 548 (2004).
17. J. G. Nijhof *et al.*, *Development* **133**, 3027 (2006).
18. U. B. Jensen *et al.*, *J. Cell Sci.* **121**, 609 (2008).
19. K. B. Jensen *et al.*, *Cell Stem Cell* **4**, 427 (2009).
20. N. Barker *et al.*, *Nature* **449**, 1003 (2007).
21. V. Jaks *et al.*, *Nat. Genet.* **40**, 1291 (2008).
22. T. Van Loy *et al.*, *Gen. Comp. Endocrinol.* **155**, 14 (2008).
23. Materials and methods are available as supporting material on Science Online.
24. S. Iseki *et al.*, *Biochem. Biophys. Res. Commun.* **218**, 688 (1996).
25. V. P. Vidal *et al.*, *Curr. Biol.* **15**, 1340 (2005).
26. G. Van Schoore, F. Mendive, R. Pochet, G. Vassart, *Histochem. Cell Biol.* **124**, 35 (2005).
27. H. Rhee, L. Polak, E. Fuchs, *Science* **312**, 1946 (2006).
28. L. G. van der Flier *et al.*, *Cell* **136**, 903 (2009).
29. J. Pispis, M. Pummila, P. A. Barker, I. Thesleff, M. L. Mikkola, *Hum. Mol. Genet.* **17**, 3380 (2008).
30. M. Kasper *et al.*, *Mol. Cell. Biol.* **26**, 6283 (2006).
31. M. Ito *et al.*, *Nature* **447**, 316 (2007).
32. We acknowledge M. Cozijnsen, J. Korving, Å. Bergström, Å. Dackland, and Hubrecht Imaging Centre for technical help and support. This work was supported by the Koningin Wilhelmina Fonds and a European Research Council grant to H.C., a European Molecular Biology Organization Long-Term Fellowship to A.H., and by grants from the Swedish Cancer Society and the Swedish Research Council to R.T. A patent is pending on work with Lgr5 and Lgr6 (A.H., N.B., and H.C.). GEO microarray accession number is GSE20269.

## Supporting Online Material

www.sciencemag.org/cgi/content/full/327/5971/1385/DC1

Materials and Methods

Figs. S1 to S9

References

16 November 2009; accepted 10 February 2010

10.1126/science.1184733

# Structural Sources of Robustness in Biochemical Reaction Networks

Guy Shinar<sup>1</sup> and Martin Feinberg<sup>2\*</sup>

In vivo variations in the concentrations of biomolecular species are inevitable. These variations in turn propagate along networks of chemical reactions and modify the concentrations of still other species, which influence biological activity. Because excessive variations in the amounts of certain active species might hamper cell function, regulation systems have evolved that act to maintain concentrations within tight bounds. We identify simple yet subtle structural attributes that impart concentration robustness to any mass-action network possessing them. We thereby describe a large class of robustness-inducing networks that already embraces two quite different biochemical modules for which concentration robustness has been observed experimentally: the *Escherichia coli* osmoregulation system EnvZ-OmpR and the glyoxylate bypass control system isocitrate dehydrogenase kinase-phosphatase–isocitrate dehydrogenase. The structural attributes identified here might confer robustness far more broadly.

Biological systems require robustness, that is, the capacity for sustained and precise function even in the presence of structural or environmental disruption (*I–II*). Examples of robustness exist over multiple scales of biological organization, from the biochemical circuit level [robust exact adaptation in bacterial chemotaxis (*2–4*)] to the cellular level [robustness of metabolic functions to changes caused by mutations (*12*)].

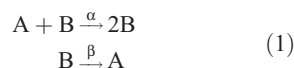
A biological system shows absolute concentration robustness (ACR) for an active molecular species if the concentration of that species is identical in every positive steady state the system might admit. The function of an ACR-possessing system is thereby protected even against large changes in the overall supply of the system's components.

<sup>1</sup>Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel. <sup>2</sup>William G. Lowrie Department of Chemical and Biomolecular Engineering and Department of Mathematics, Ohio State University, 125 Koffolt Laboratories, 140 West 19th Avenue, Columbus, OH 43210, USA.

\*To whom correspondence should be addressed. E-mail: feinberg@cblmeng.ohio-state.edu

We identify simple yet subtle structural attributes that will impart ACR to any mass-action network that includes them. We provide a mathematical theorem that precisely delineates a very large class of ACR-possessing systems, a class that embraces networks that differ in size, detail, and complexity. This class contains different ACR-possessing models (*9, 11*) of known examples for which approximate concentration robustness has been verified experimentally. We thus uncover an underlying mathematical unity found at the heart of robustness-producing mechanisms that are biochemically quite different.

To elucidate the concept of ACR, we first consider the toy two-species mass-action system



where A is the active form of a protein, B is the inactive form, and  $\alpha$  and  $\beta$  are rate constants. Suppose the protein is synthesized and degraded over long time scales, so that the total protein concentration can be regarded as constant over

the system's equilibration time scale. Under this assumption, the differential equations governing the time evolution of the molar concentrations of A and B, denoted  $c_A$  and  $c_B$ , are

$$\begin{aligned} \dot{c}_A &= -\alpha c_A c_B + \beta c_B \\ \dot{c}_B &= \alpha c_A c_B - \beta c_B \end{aligned} \quad (2)$$

The positive steady states of Eq. 2 are given by

$$\begin{aligned} c_A &= \frac{\beta}{\alpha} \\ c_B &= \Theta - \frac{\beta}{\alpha} \end{aligned} \quad (3)$$

where  $\Theta$  is the conserved total protein concentration:  $\Theta = c_A + c_B = c_A(0) + c_B(0)$ . Eq. 3 shows that system (1) has ACR: There is a positive steady state for each value of  $\Theta$  exceeding  $\beta/\alpha$ , and in each of these steady states  $c_A$  has precisely the same value.

In contrast, consider the simple module



Here, the positive steady states are given by

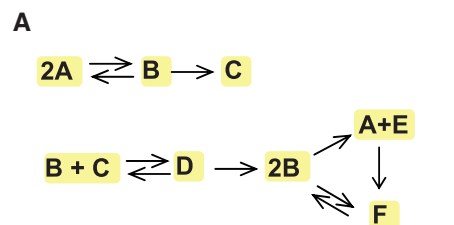
$$\begin{aligned} c_A &= \frac{\beta\Theta}{\alpha + \beta} \\ c_B &= \frac{\alpha\Theta}{\alpha + \beta} \end{aligned} \quad (5)$$

The steady state values of both  $c_A$  and  $c_B$  are proportional to the conserved total concentration  $\Theta = c_A + c_B$ . Thus, as  $\Theta$  varies, both  $c_A$  and  $c_B$  vary in step. The system does not have ACR.

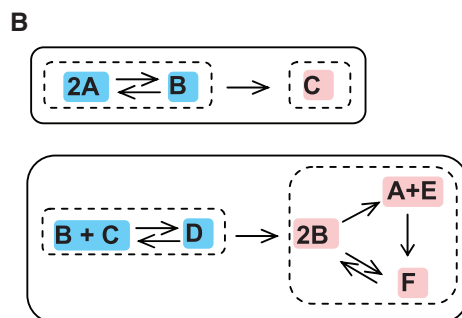
To state our main result, we require some terminology from chemical reaction network theory (*13–16*). The display in Fig. 1A is an example of a standard reaction diagram, that is, a directed graph whose nodes (*17*) are the distinct linear combinations of chemical species that sit at the heads and tails of the reaction arrows. In Fig. 1A, the chemical species are A, B, C, D, E, and F,

and the nodes are  $2A$ ,  $B$ ,  $C$ ,  $B + C$ ,  $D$ ,  $2B$ ,  $A + E$ , and  $F$ . Note that in a standard reaction diagram each node appears precisely once.

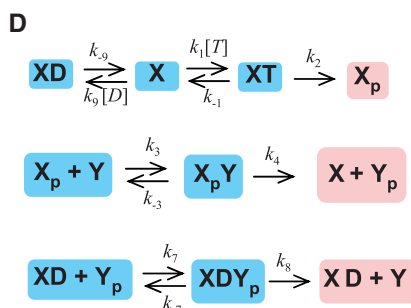
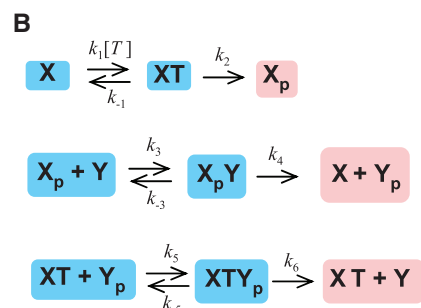
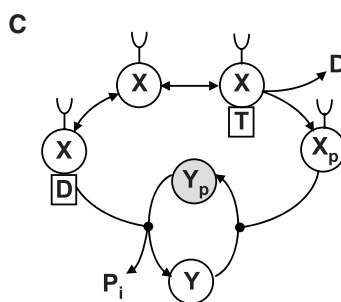
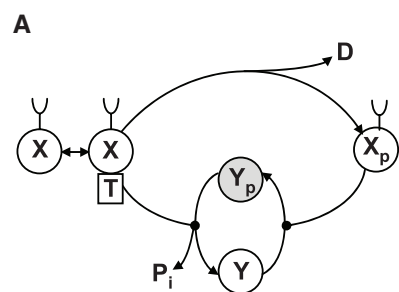
The standard reaction diagram in Fig. 1A is made of two disconnected pieces, one containing the mutually linked nodes  $2A$ ,  $B$ , and  $C$  and the other containing the mutually linked nodes  $B + C$ ,  $D$ ,  $2B$ ,  $A + E$ , and  $F$  (Fig. 1B). Such sets of mutually linked nodes are called the linkage classes of the network. The number of linkage classes is identical to the number of pieces of which the standard reaction diagram is composed.



**Fig. 1.** Some concepts from chemical reaction network theory. **(A)** A standard reaction diagram. The nodes of the diagram are shaded yellow. **(B)** Graph-theoretical properties of standard reaction diagrams. The connected pieces of the diagram, corresponding to linkage classes, are surrounded by solid outlines. Parts of the diagram corresponding to strong-linkage classes are surrounded by dashed outlines. Terminal nodes are colored pink, and nonterminal nodes are colored blue.



$$\begin{aligned} \#(\text{nodes}) &= 8 \\ \#(\text{linkage classes}) &= 2 \\ \text{rank} &= 5 \\ \text{deficiency} &= 8 - 2 - 5 = 1 \end{aligned}$$



**Fig. 2.** The EnvZ-OmpR system. **(A)** A schematic diagram of an EnvZ-OmpR model in which ATP is the cofactor in phospho-OmpR dephosphorylation.  $P_i$  denotes phosphate ion. **(B)** The mass-action model underlying (A).  $[T]$  denotes the ATP concentration, assumed fixed. Terminal nodes are colored pink, and nonterminal nodes are colored blue. **(C)** A schematic diagram of an EnvZ-OmpR model in which ADP is the cofactor in phospho-OmpR dephosphorylation. **(D)** The mass-action model underlying (C).  $[D]$  denotes the ADP concentration, assumed fixed.

We say that two nodes are strongly linked if there is a directed arrow path from one to the other and also a directed arrow path from the second back to the first. Thus, the nodes  $A + E$  and  $F$  in Fig. 1A are strongly linked because there is a directed arrow path from  $A + E$  to  $F$ , namely  $A + E \rightarrow F$ , and also a directed arrow path from  $F$  back to  $A + E$ , namely  $F \rightarrow 2B \rightarrow A + E$ . We adopt the convention that each node is strongly linked to itself.

A strong-linkage class of a reaction network is a maximal subset of its nodes that are strongly linked to each other. The parts of

Fig. 1B that are surrounded by dashed outlines correspond to the strong-linkage classes in our example network. A terminal strong-linkage class is a strong-linkage class in which no node reacts to a node in another strong-linkage class. (We say that one node reacts to another node whenever the first sits at the tail and the second sits at the head of the same reaction arrow.) Nodes belonging to terminal strong-linkage classes are called terminal (pink in Fig. 1B). Nodes that are not terminal are called nonterminal (blue in Fig. 1B).

By the rank of a reaction network we mean the maximum number of linearly independent reactions that the network contains. We can make this precise in the following way: With each reaction, we associate a reaction vector (18) obtained by subtracting the reactant node from the product node. Thus, we associate with reaction  $2A \rightarrow B$  the reaction vector  $B - 2A$ ; with  $B \rightarrow 2A$  we associate the reaction vector  $2A - B$ ; and so on. The five reaction vectors,  $B - 2A$ ,  $D - B - C$ ,  $2B - D$ ,  $A + E - 2B$ , and  $F - A - E$ , constitute a linearly independent set, and every other reaction vector for the network of Fig. 1A can be written as a linear combination of these. Thus, the rank of the network is five.

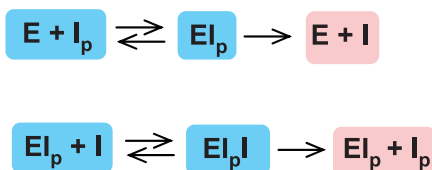
In formulating the reaction vector for a given reaction, we have subtracted the reactant node from the product node. In fact, we shall be interested in the difference of any two nodes, even those that belong to different reactions. For example, the difference of the nodes  $B + C$  and  $B$  is  $B + C - B = C$ . If, as in this example, the difference of two nodes is a nonzero multiple of a single species  $S$ , we say that the nodes “differ only in species  $S$ .”

Next, we introduce the important concept of deficiency. The deficiency of a network is an integer index obtained by subtracting both the number of linkage classes and the rank from the number of nodes. For the network of Fig. 1 the deficiency is one, because there are eight nodes, two linkage classes, and its rank is five. The deficiency of a network will invariably be a nonnegative integer (13). The relationship between the size and the deficiency of a network is at best weak; very large networks often have very low deficiency.

Lastly, a mass-action system is said to have “absolute concentration robustness in species  $S$ ” if the system admits a positive steady state and if in all positive steady states the concentration of  $S$  is the same.

We now have the vocabulary required for stating our main result, which is proved (section S3) in the supporting online material (SOM). Motivation for the proof (section S2) and generalizations (section S4) of the theorem are discussed there as well. A software implementation of the theorem is available for download (19).

**Theorem:** Consider a mass-action system that admits a positive steady state and suppose that the deficiency of the underlying reaction network is one. If, in the network, there are two



**Fig. 3.** A core ACR module in a model of the IDHKP-IDH system. Terminal nodes are colored pink, and nonterminal nodes are colored blue.

nonterminal nodes that differ only in species  $S$ , then the system has absolute concentration robustness in  $S$ .

The theorem becomes false if the requirement that the deficiency be one is dropped. In fact, there are networks with a deficiency of two that otherwise satisfy all the conditions of the theorem for which ACR obtains and still others for which ACR does not obtain. Moreover, no mass-action system with a deficiency of zero that is consistent with mass conservation can exhibit ACR relative to any species (20).

We now turn to applications of the theorem, beginning with a model of a prototypical two-component signaling system (21, 22), for which approximate concentration robustness has been observed experimentally (7). The *Escherichia coli* EnvZ-OmpR system consists of the sensor kinase EnvZ, denoted  $X$  in Fig. 2A, and the response-regulator OmpR, denoted  $Y$ . Both the sensor and the response-regulator have phosphorylated forms, denoted  $X_p$  and  $Y_p$ .  $X$  phosphorylates itself by binding and breaking down adenosine triphosphate (ATP) (T in Fig. 2) (21, 22),  $X_p$  catalyzes the transfer of a phosphoryl group to free  $Y$  (21, 22), and  $X$ , together with ATP (23, 24) or adenosine diphosphate (ADP) (D in Fig. 2) (25) as a cofactor, dephosphorylates  $Y_p$ . The crucial chemical species in the system is  $Y_p$ , a transcription factor that regulates the expression of various protein pores.

The EnvZ-OmpR module has been modeled (9) by using the differential equations derived from the mass-action kinetic system of Fig. 2B, where the effects of ADP as a dephosphorylation cofactor have been neglected. Note that the underlying network and its corresponding differential equations conserve the total concentrations  $\Theta_X$  and  $\Theta_Y$  of  $X$ -containing and  $Y$ -containing species. Analysis (9) of the mass-action equations indicates that, in all positive steady states (corresponding to various combinations of the parameters  $\Theta_X$  and  $\Theta_Y$ ), the concentration of  $Y_p$  depends solely on rate constants (9). Thus, the mass-action model of Fig. 2B exhibits ACR in  $Y_p$ .

To apply the theorem to Fig. 2B, we note that the network has nine nodes and three linkage classes. It is not difficult to verify that the rank of the network is five. Therefore, the deficiency of the network is  $9 - 3 - 5 = 1$ . Moreover, the nodes  $XT + Y_p$  and  $XT$  are nonterminal, and they differ only in the species  $Y_p$ . Thus, the theorem asserts that, if the mass-action model shown in Fig. 2B admits a positive steady state (as it does), then the

model has ACR in  $Y_p$ . This is consistent with the earlier ad hoc analysis (9). The theorem also ensures ACR for a network model, displayed in Fig. 2D (25, 26), in which ADP, rather than ATP, is the dominant cofactor (27, 28) for  $Y_p$  dephosphorylation. This and other model variants are treated in the SOM (Section S5).

ACR also obtains in the mass-action model of Fig. 3 (11), which lies at the core of more elaborate models for the *E. coli* IDHKP-IDH glyoxylate bypass regulation system. [Approximate concentration robustness has been observed experimentally (1) in this system.] Here,  $I$  denotes the active, unphosphorylated TCA cycle enzyme isocitrate dehydrogenase (IDH). The inactive, phosphorylated form is denoted  $I_p$ .  $E$  denotes the bifunctional enzyme IDH kinase-phosphatase (IDHKP). The deficiency is again one: The network has six nodes, two linkage classes, and a rank of three, and there are two nonterminal nodes,  $E I_p + I$  and  $E I_p$ , that differ only in the species  $I$ . Thus, the theorem asserts that, if a positive steady state exists (as it does), then the model has ACR in  $I$ .

The EnvZ-OmpR and IDHKP-IDH systems both use bifunctional enzymes that act simultaneously as a kinase and a phosphatase. This observation highlights the connection between the biological implementation of ACR-possessing modules and the theorem's requirement for two nonterminal nodes that differ only in a species. If the dephosphorylation reaction in Fig. 2B were catalyzed by some phosphatase  $Z$  and not by  $XT$ , we would replace the third piece in the network of Fig. 2B with  $Z + Y_p \rightleftharpoons ZY_p \rightarrow Z + Y$ . The resulting network would cease to have two nonterminal nodes that differ only in a species, and the theorem would not apply. In fact, ACR would be lost.

The mass-action models discussed here freely invoke irreversible reactions. The omitted reverse reactions generally have rate constants so small that, for practical purposes, the reactions themselves can be safely ignored. In such instances, the reduced model is deemed to be an approximate but reliable guide to the behavior of a fuller mass-action system with some or all of the reverse reactions included. In particular, it is reasonable to expect that the fuller model would exhibit strong but imperfect robustness. These considerations are discussed more fully in the SOM (section S6).

Lastly, we emphasize that the ideal of absolute concentration robustness is unlikely to be attained exactly for in vivo experimental systems, where biochemical modules do not exist by themselves and often interact with the intracellular environment. Therefore, complete reaction network models for experimental systems should not be expected to exhibit ACR exactly. Nevertheless, the theorem presented here describes a general class of core subnetworks, which, taken by themselves, do give rise to ACR and which, to the extent that they approximate their more completely articulated parent networks, may

confer on the fuller systems imperfect yet strong robustness.

## References and Notes

1. D. C. LaPorte, P. E. Thorsness, D. E. Koshland Jr., *J. Biol. Chem.* **260**, 10563 (1985).
2. N. Barkai, S. Leibler, *Nature* **387**, 913 (1997).
3. U. Alon, M. G. Surette, N. Barkai, S. Leibler, *Nature* **397**, 168 (1999).
4. T. M. Yi, Y. Huang, M. I. Simon, J. Doyle, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4649 (2000).
5. A. Levchenko, P. A. Iglesias, *Biophys. J.* **82**, 50 (2002).
6. A. Eldar et al., *Nature* **419**, 304 (2002).
7. E. Batchelor, M. Goulian, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 691 (2003).
8. M. Kollmann, L. Løvdok, K. Bartholomé, J. Timmer, V. Sourjik, *Nature* **438**, 504 (2005).
9. G. Shinar, R. Milo, M. R. Martínez, U. Alon, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19931 (2007).
10. A. Csikász-Nagy, O. S. Soyer, *J. R. Soc. Interface* **5** (suppl. 1), S41 (2008).
11. G. Shinar, J. D. Rabinowitz, U. Alon, *PLOS Comput. Biol.* **5**, e1000297 (2009).
12. J. A. de Visser et al., *Evolution* **57**, 1959 (2003).
13. M. Feinberg, *Lectures on Chemical Reaction Networks—Written Version of Lectures Delivered at the Mathematics Research Center, University of Wisconsin, Madison*, [www.che.eng.ohio-state.edu/~FEINBERG/LecturesOnReactionNetworks/](http://www.che.eng.ohio-state.edu/~FEINBERG/LecturesOnReactionNetworks/) (1979).
14. M. Feinberg, *Chem. Eng. Sci.* **42**, 2229 (1987).
15. M. Feinberg, *Arch. Ration. Mech. Anal.* **132**, 311 (1995).
16. G. Craciun, Y. Tang, M. Feinberg, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8697 (2006).
17. What we call “nodes” in the present paper are termed “complexes” in the chemical reaction network theory literature.
18. The reaction vectors reside in the vector space of all formal linear combinations of species.
19. E. Eden, G. Shinar, Reaction Network Analyzer V1.0, [www.weizmann.ac.il/mcb/UriAlon/people/GuyShinar/index.html](http://www.weizmann.ac.il/mcb/UriAlon/people/GuyShinar/index.html) (2009).
20. G. Shinar, U. Alon, M. Feinberg, *SIAM J. Appl. Math.* **69**, 977 (2009).
21. L. Pratt, T. J. Silhavy, in *Two-Component Signal Transduction*, J. A. Hoch, T. J. Silhavy, Eds. (American Society for Microbiology, Washington, DC, 1995), pp. 105–127.
22. A. M. Stock, V. L. Robinson, P. N. Goudreau, *Annu. Rev. Biochem.* **69**, 183 (2000).
23. M. M. Igo, A. J. Ninfa, J. B. Stock, T. J. Silhavy, *Genes Dev.* **3**, 1725 (1989).
24. W. Hsing, T. J. Silhavy, *J. Bacteriol.* **179**, 3729 (1997).
25. Y. Zhu, L. Qin, T. Yoshida, M. Inouye, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7808 (2000).
26. According to (25), ADP participates in phospho-OmpR dephosphorylation by stimulating phosphoryl hydrolysis, not by serving as a substrate in ATP reconstitution.
27. S. Sanowar, H. Le Moual, *Biochem. J.* **390**, 769 (2005).
28. Experiments on the PhoP-PhoQ two-component system suggest that phospho-PhoQ dephosphorylation by PhoP is stimulated by ADP, not by the ATP analog p[NH]ppA (27).
29. We are grateful to U. Alon for comments, insight, and support; E. Eden for writing the software implementation of the theorem; and A. Cohen, G. Enciso, R. Kishony, D. Koster, M. Laub, A. Mayo, R. Milo, and J. Rabinowitz for their help. Supported by NSF grant BES-0425459, NIH grant 1R01GM086881-01, and the Kahn Family Foundation.

## Supporting Online Material

[www.sciencemag.org/cgi/content/full/327/5971/1389/DC1](http://www.sciencemag.org/cgi/content/full/327/5971/1389/DC1)  
SOM Text  
Figs. S1 and S2  
References

15 October 2009; accepted 12 February 2010  
10.1126/science.1183372