from the Alfred P. Sloan Foundation (to M.K.C.). Work was done partially at Stanford Synchrotron Radiation Laboratory (SSRL), which is operated by the Department of Energy, Office of Basic Energy Sciences. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR07707. The beamline X4A at the National Synchrotron Light Source, a Department of Energy facility, is supported by the Howard Hughes Medical Institute. The authors thank K. Greenchurch at the Ohio State University Chemical Instrumentation Center for her assistance with mass spectrometric analysis. The coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 1L2Q (NaCl form) and 1L2R [$(NH_4)_2SO_4$ form].

Supporting Online Material
www.sciencemag.org/cgi/content/full/296/5572/1462/DC1
Materials and Methods
Figs. S1 to S3

4 January 2002; accepted 27 March 2002

Combinatorial Synthesis of Genetic Networks

Călin C. Guet,[1,3] Michael B. Elowitz,[3] Weihong Hsing,[1] Stanislas Leibler[1,2,3*]

A central problem in biology is determining how genes interact as parts of functional networks. Creation and analysis of synthetic networks, composed of well-characterized genetic elements, provide a framework for theoretical modeling. Here, with the use of a combinatorial method, a library of networks with varying connectivity was generated in *Escherichia coli*. These networks were composed of genes encoding the transcriptional regulators LacI, TetR, and lambda CI, as well as the corresponding promoters. They displayed phenotypic behaviors resembling binary logical circuits, with two chemical "inputs" and a fluorescent protein "output." Within this simple system, diverse computational functions arose through changes in network connectivity. Combinatorial synthesis provides an alternative approach for studying biological networks, as well as an efficient method for producing diverse phenotypes in vivo.

Living cells respond to information from their environment on the basis of the interactions of a large yet limited number of molecular species that are arranged in complex cellular networks (*1*, *2*). A classic example of such biochemical computation is the chemotaxis behavior of *Escherichia coli*, which is mediated by a well-characterized signal transduction network (*3*). However, despite growing knowledge about the molecular components of the cell, the dynamics of even simple cellular networks are not well understood. For instance, a quantitative explanation of the high sensitivity and exact adaptation observed in bacterial chemotaxis is still lacking (*3*). Similarly, many other cellular networks, such as the ones responsible for signal transduction, regulation of gene expression, or metabolism, are poorly understood from a quantitative point of view. Thus, simple and modular experimental systems are needed to study how the genetic structure and connectivity of cellular networks are related to their function. To this end, we devised an in vivo synthetic system that enables the generation of combinatorial libraries of genetic networks.

We have generated a combinatorial library composed of a small set of transcriptional regulatory genes and their corresponding promoters with varying connectivity (Fig. 1). We chose genes of three well-characterized prokaryotic transcriptional regulators: LacI, TetR, and lambda cI (*4*). The binding state of LacI and TetR can be changed with the small molecule inducers, isopropyl β-D-thiogalactopyranoside (IPTG) and anhydrotetracycline (aTc), respectively. We also chose five promoters regulated by these proteins, which cover a broad range of regulatory characteristics such as repression, activation, leakiness, and strength. Two of the promoters are

[1]Howard Hughes Medical Institute, Department of Molecular Biology, [2]Department of Physics, Princeton University, Princeton, NJ 08544, USA. [3]The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

*To whom correspondence should be addressed. Laboratory for Living Matter, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

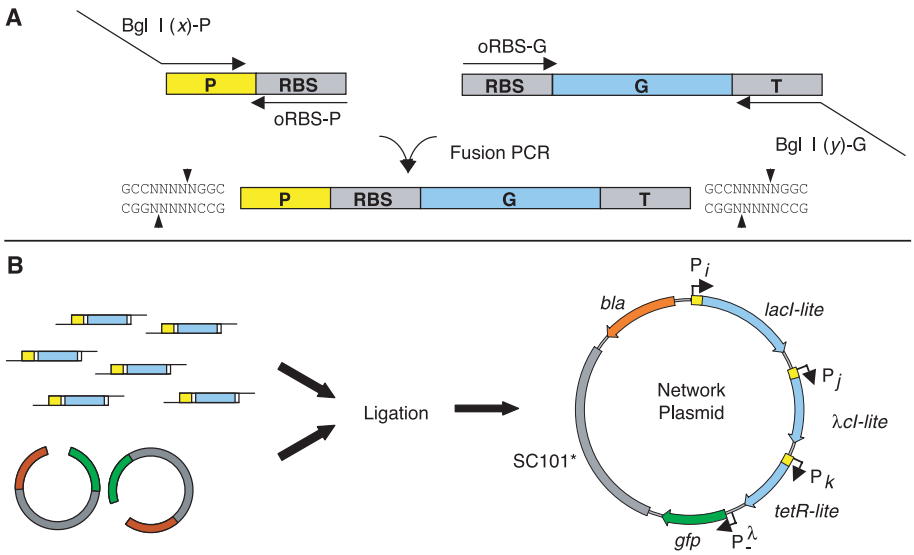

Fig. 1. The modular genetic cloning strategy used to generate combinatorial libraries of logical circuits. Construction of the library proceeded in two steps. In the first step (**A**), we built all 15 possible promoter-gene units. Individual promoters and genes were first amplified by PCR. The genes [denoted "*–lite*" in (B)] have an *ssrA* tag that reduces the half-life of the proteins encoded by the modified gene (*21*). The five promoters used were P^L_1 and P^L_2 (repressed by LacI), P^T (repressed by TetR), and P^{λ}_- and P^{λ}_+ (repressed and activated, respectively, by λ cI) (*5*). The transcriptional terminator T1 was present at the end of each gene. Identical RBS were used as internal primers for the subsequent fusion PCR step to form promoter-gene units (*27*). In order to control the number of promoter-gene units and the position of a given gene in the network, BglI sites were incorporated in PCR primers, as shown. The special recognition and restriction properties of BglI (*28*) allow various sticky ends to be produced by BglI cleavage. Here we designed the BglI sites such that specific cohesive ends *x* and *y* were associated with each regulatory gene (for *lacI*, x_{lac} = GCC, y_{lac} = TTC; for λ *cI*, x_{cI} = AAG, y_{cI} = GTG; and for *tetR*, x_{tet} = CAC, y_{tet} = TCG). Note that y_{lac} is compatible with x_{cI}, y_{cI} is compatible with x_{tetR}, and so on. Thus, in step (**B**) when all 15 possible fusion PCR products were mixed together and ligated, the resulting products contained exactly three promoter-gene units in one particular order (*lacI*, λ *cI*, *tetR*). These products were cloned into a low copy number plasmid (three to four copies per cell) (*23*), carrying the reporter gene *gfpmut3* under the control of P^{λ}_- (*29*).

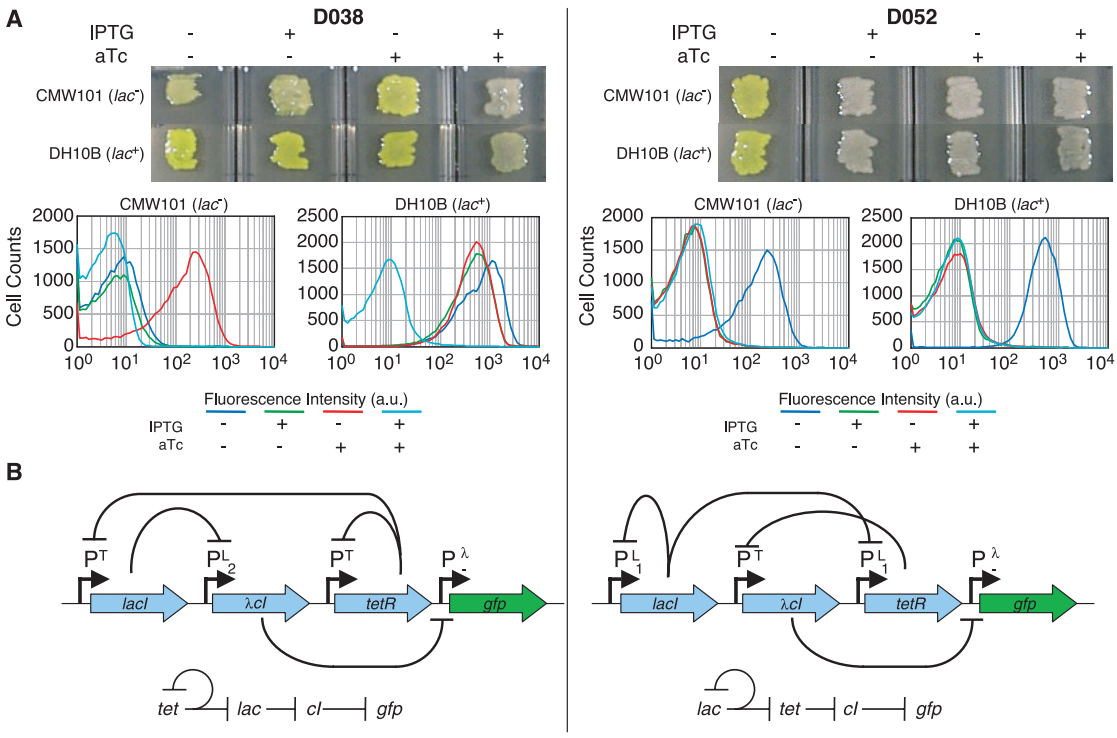

Fig. 2. Detailed analysis of two binary logical circuits (D038 and D052). (**A**) Cells of both *E. coli* host strains, *lac*$^-$ strain CMW101 and *lac*$^+$ strain DH10B, transformed with each of two networks. "Logical circuit" behavior can be observed directly on agar plates (top), where fluorescence of colonies defines "on" and "off" states. Cells containing the indicated network were patched onto minimal agar media containing all four combinations of the two inducers in separate wells, as indicated. To increase fluorescent signal and to show that reporter expression is not cis-dependent, cells contained plasmids deleted for *gfp* and were cotransformed with compatible plasmids (~15 copies/cell) containing an equivalent P^{λ}*-yfp* transcriptional unit. Cells were grown also in liquid culture and populations were analyzed with FACS for distributions of GFP expression (bottom). In each set of histograms, the blue curve shows the fluorescence distribution without inducers, the green curve shows when IPTG alone was present, the red curve indicates aTc alone, and the cyan curve shows the distribution when both inducers were present simultaneously. Single-peaked distributions are observed, but, in some cases, peaks contain long tails that overlap, corresponding to "leaky" or "fuzzy" logical circuits. (**B**) Sequencing was used to determine the connectivity of each of the two networks. Despite the fact that their logical behavior is different, both networks have the same connectivity ("topology"), as can be inferred from the corresponding diagrams of the interactions between the repressors and promoters. The schematic connectivity or topology diagrams shown at the bottom are identical for the two networks.

repressed by LacI, one is repressed by TetR, and the remaining two are regulated by λ cI, one positively and one negatively (*5*). The genetic assembly scheme we used ensures that each network in the library has the following structure: P_i-*lacI*—P_j-λcI—P_k-*tetR*, wherein each P_i, P_j, P_k represents any of the five promoters (Fig. 1). The regulatory genes on each plasmid activate or repress one another in various ways in vivo, generating networks with diverse connectivities. Altogether, $5^3 = 125$ different networks are possible. In principle, this method allows the construction of a wide range of regulatory motifs within the combinatorial library, including negative and positive feedback loops, oscillators, and toggle switches.

The processing of multiple chemical signals is a fundamental task performed by many natural biochemical networks. In order to screen for such behaviors, we incorporated in each plasmid a fourth transcriptional unit, in which green fluorescent protein (GFP) expression was controlled by the λ cI repressible promoter. The fluorescent signal acts as the network "output," whereas the levels of the two chemical inducers were used as "inputs." The plasmid library was transformed into two different host strains of *E. coli* (*6*), which differed most significantly by the presence of a wild-type copy of *lacI* at a chromosomal locus. Each clone was grown under four conditions, with and without IPTG and with or without aTc. GFP fluorescence was monitored simultaneously during cell growth (*7*). In this way, we searched the library for circuits in which the output is a binary logical function of both inducers. Examples of such "logical circuits" are NAND, NOR, or NOT IF (Figs. 2A and 3).

We found that, in many cases, the output fluorescence levels of an individual culture in our library were sufficiently distinct for different inputs that an unambiguous binary output value for each input state could be assigned at a glance. In other cases, the designation was somewhat arbitrary. However, as shown in Fig. 3, it is quite remarkable that one can impose a single universal output-fluorescence threshold on the entire library and still obtain a large number of logical circuits for which "on" and "off" states differ significantly (*8*). Note that the

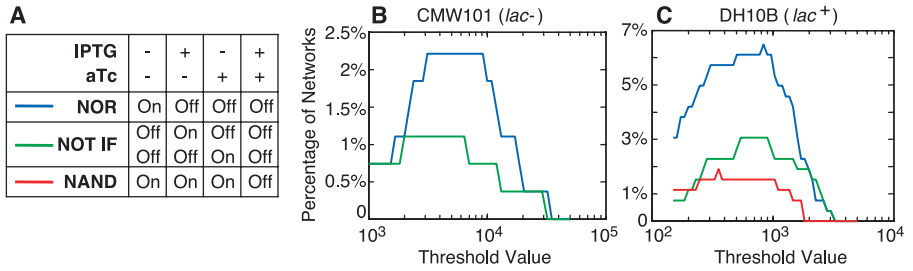

Fig. 3. Distribution of logical phenotypes in the two strains. (**A**) Definition of the logic operations performed by the circuits. In the top row, + and – indicate the presence or absence of each inducer input. The output (fluorescence) is indicated in the lower rows by "On" or "Off." We do not distinguish here between the two inputs and, thus, between two different types of NOT IF logic functions. Colored bars act as legends for (**B**) and (**C**). These histograms (one for each host strain) show the fraction of networks qualifying as logical circuits of each type for varying values of a threshold parameter. A single universal threshold value was applied simultaneously to all networks. Besides being greater than this threshold, the minimal "on" value in each particular network was also required to be at least fourfold greater than the maximal "off" value. (The difference in magnitudes on the x-axis for the two strains is due to different instrument settings.)

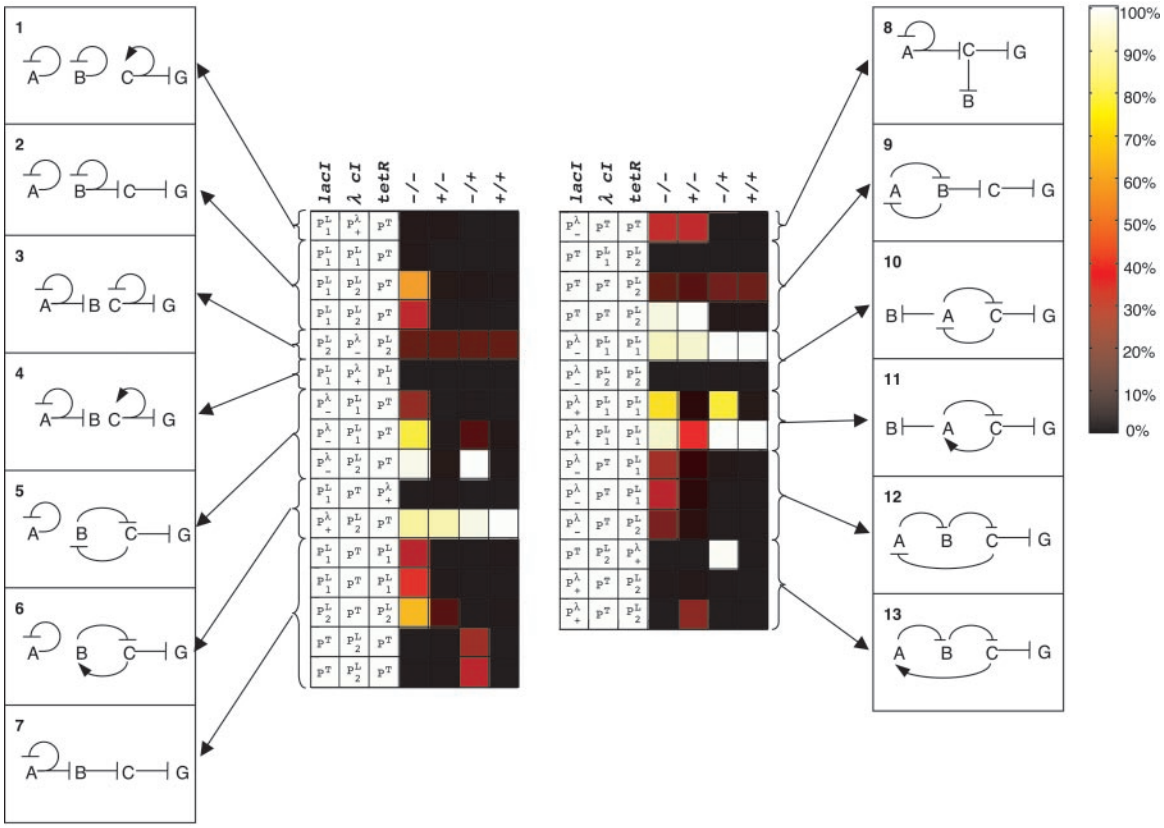

Fig. 4. Genetic structure and behavior of selected networks. A subset of 30 plasmids was characterized in the *lac*$^-$ host strain (CMW101), and their genetic composition was determined by sequencing. Levels of GFP expression in each of the four conditions are indicated by the color or intensity of the corresponding box on a linear intensity scale (see color bar). In many cases, logical behavior is strain dependent (i.e., is different in *lac*$^+$ strain DH10B; data not shown). The promoters incorporated in the network, shown in the left-hand columns, determine its connectivity diagram, representing known interactions according to which the networks are ordered. The 13 connectivity diagrams corresponding to different networks are drawn. A and B refer to either of the inducible repressors (LacI and TetR), C always denotes λ CI, and G denotes GFP (the output of the network). Activation is denoted by sharp arrows (↓), while repression is denoted by blunt arrows (⊥).

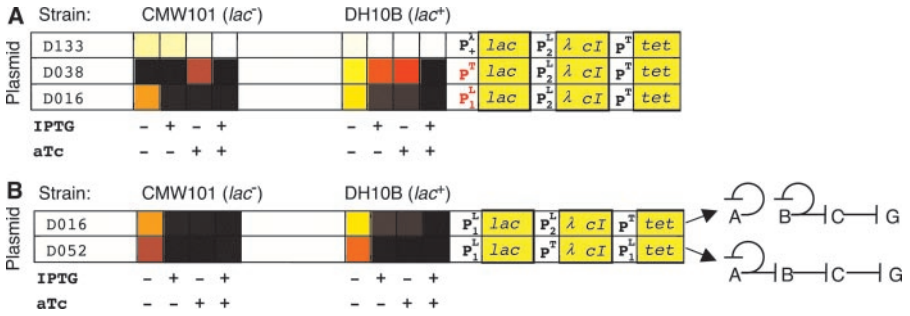

Fig. 5. Dependence of phenotypic behavior on network connectivity. (**A**) A single change of the promoter can completely modify the behavior of the logical circuit. The network on plasmid D133, which is always in the "on" state, differs from the networks encoded in plasmids D038 and D016 by the single promoter placed in front of *lac* gene (shown in red). Both circuits show diverse logical behaviors, which differ in the two bacterial host strains. (**B**) Networks can differ by their connectivity but have qualitatively the same logical function. Both, D016 and D052, behave as NOR logical circuits (in both *lac*$^+$ and *lac*$^-$ strains), but they have different connectivity, as shown schematically on the right side of the table. The linear color or intensity scale indicating the levels of GFP expression is the same as in Fig. 4.

spectrum of logical behaviors differed in the two hosts (Fig. 3). This is not surprising, because the chromosomal *lacI* gene present in one strain (DH10B) acts as a network component and thus may drastically change the resulting phenotype (Fig. 2A).

Phenotypic variation in organisms often arises through mutations in the protein coding regions of the DNA. Another important contribution to phenotypic variability comes from changes in the cis-regulatory connections of existing genetic elements (*9*–*10*). To determine the origin of the phenotypic variability observed in the library, 30 clones with a variety of different behaviors were retransformed into both hosts, rescreened (to verify their phenotypic behavior), and sequenced (Fig. 4). We found a low level of point mutations, which, in some cases, modify the logical behavior of the networks (*11*). More interestingly, however, a large variety of behaviors remains among the many networks that do not have mutations in their regulatory regions. Sequencing allowed us to identify the three promoters incorporated in each plasmid; connectivity between different genetic elements varies from network to network so that 13 different "topologies" can be distinguished among the sequenced networks (Fig. 4). This variety of network connectivity is evidently the major source of phenotypic diversity in the library. In fact, the sequence data show that single step changes to the network connections, in which one promoter replaces another, frequently converted network operation from one logical function to another. For example, by replacing a single promoter in a network (D133 in Fig. 5A), which is always in the "on" state, one obtains a network (D038), which acts as a NOT IF in one strain and as a NAND in the other. Alternatively, by performing a different promoter replacement, one obtains network D016, which acts as a NOR circuit in both strains.

The fact that our restricted and conserved set of genes and promoters has the potential to switch among a variety of different computational functions by one-step changes in connectivity is noteworthy. From an evolutionary point of view, this observation suggests that once a simple set of genes and cis-regulatory elements is in place, it should be possible to jump from one functional phenotype to another using the same "toolkit" of genes (*10*) by modifying the regulatory connections. Such discontinuous changes, differ-

ent from the more gradual effects driven by successive point mutations, may be achieved in evolution by natural combinatorial mechanisms like transposition, recombination, or gene duplication.

Connectivity of a network does not uniquely determine its behavior. We found examples of networks that share the same connectivity but perform different logical operations. For instance, two networks depicted in Fig. 2 have the same connectivity ("topology," see Fig. 2B) but show very different phenotypic behavior (Fig. 2A). We also found examples of networks with different connectivity that exhibited qualitatively similar behaviors. For instance, two networks shown in Fig. 5B (D016 and D052) both perform NOR operations despite their different connectivity. Thus, as shown by these two examples, the behavior of even simple networks built out of a few, well-characterized components cannot always be inferred from connectivity diagrams alone.

Can one predict the behavior of the logical circuits obtained here? "Boolean-type" models of gene regulation are often used to intuitively understand the operation of genetic networks. In this simplified description, one considers only discrete values of the biochemical variables and parameters (leading to commonly used reasoning, such as: "gene product A is produced, it inhibits the expression of gene product B, which is thus absent, etc . . . ") (*12*, *13*). This description, though adequate for some of the present networks, seems not to apply to others (*14*). For instance, for the circuits in the *lac*$^-$ strain, shown in Fig. 2, the Boolean description is consistent with the NOT IF behavior of network D038 but not the NOR behavior of network D052 (*15*). It is possible that even for these well-studied transcriptional regulators subtle additional regulation may be at work among the plasmid-encoded elements (*16*). Genetic networks are nonlinear, stochastic systems in which the unknown details of interactions between components might be of crucial importance. Combinatorial libraries of simple networks should be useful in the future to uncover the existence of such additional regulation mechanisms and to explore the limits of quantitative modeling of cellular systems (*17*). For instance, it would be interesting to see whether the behavior of all the networks in the library could be described within a single theoretical model, a model defined by a unique set of parameters characterizing the interactions between the genetic components.

Combinatorial techniques inspired by recombination, such as DNA shuffling, have often proven successful in enhancing or changing the enzymatic activities of proteins and pathways (*18*, *19*) without requiring an understanding of the mechanisms by which they work. However, combinatorial methods in simple and well-controlled systems, such as the one presented here, can and should also be used to gain better understanding of system-level properties of cellular networks. This is particularly important before using these powerful techniques more widely, e.g., in any practical applications.

The present results show that a handful of interacting genetic elements can generate a surprisingly large diversity of complex behaviors. Although the current system uses a small number of building blocks restricted to a single type of interaction (transcriptional regulation), both the number of elements and the range of biochemical interactions can be extended by including other modular genetic elements. The approach can be taken beyond the intracellular level by linking input and output through cell-cell signaling molecules, such as those involved in quorum sensing. Lastly, this combinatorial strategy can be used to search for other dynamic behaviors such as switches, sensors, oscillators, and amplifiers, as well as for high-level structural properties, such as robustness or noise-resistance (*20*).

References and Notes

1. D. Bray, *Nature* **376**, 307 (1995).
2. E. Szathmáry, F. Jordán, C. Pál, *Science* **292**, 1315 (2001).
3. D. Bray, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7 (2002).
4. In order to avoid toxicity effects due to overexpression and reductions in the dynamic regulatory range, we used the reduced-stability variants of these transcriptional regulators described previously (*21*).
5. The following promoters were used. Poid70.5 (designated P^L_2) from plasmid pOd O_{id} 70.5 O_1 (*22*); P_R (P^λ) and P_{RM} (P^λ_+) from lambda phage DNA, and $P_L$lacO1 (P^L_1) and $P_L$tetO1 (P^T) from their original plasmids (*23*). Each of these promoters was amplified by polymerase chain reaction (PCR). In the case of P^λ_+ (P_{RM}), the or3-r3 mutation (*24*), which eliminates repression of P_{RM} at high λ CI concentrations, was introduced by PCR.
6. Host strains were CMW101 (*lacI*$^-$, *tetR*$^-$) and DH10B (*lacI*$^+$, *tetR*$^-$). Colonies were screened: 271 in CMW101 and 262 in DH10B. CMW101 was constructed by P1 transduction of a *recA*::*cam* marker from CLC90 (a gift from T. J. Silhavy) to MC1061.
7. Individual colonies were grown overnight in 96-well round-bottom plates (Sarstaedt 82.1582.001, Newton, NC) on a microplate shaker at 30°C (180 μl/well). Defined medium was used: 0.5 g $(NH_4)_2SO_4$, 5.25 g K_2HPO_4, 0.225g $MgSO_4{\cdot}7H_2O$, 19 mg EDTA, 2.5 mg $FeSO_4$ per 500 ml of H_2O, pH 6.8, with 0.5% glycerol and 0.5% Casamino acids. IPTG was used at 1 mM, and aTc at 100 ng/ml. When required, ampicillin (100 μg/ml) and chloramphenicol (30 μg/ml) were added. All experiments were conducted in triplicate and in the dark. Overnight cultures were diluted 1:360 into fresh media with each of four inducer combinations. We measured GFP fluorescence and optical density (OD) four to five times along the growth curve, using a Wallac Victor2 multi-well fluorimeter (Turku, Finland). Absorbance measurements were made at 600 nm (10 nm bandpass, 0.1 s integration time). Fluorescence readings for strain CMW101 used the settings as follows: CW-lamp filter F485 and emission filter F535; integration time, 0.1 s; emission aperture, damp; CW-lamp energy, 7000; emission side, above. In the screen of the DH10B library, the settings were changed as follows: emission aperture, band; CW-lamp energy, 21,673; emission side, below.
8. Furthermore, the value of this threshold can be changed many times without much affecting the distribution of logic functions obtained. These results indicate that in general the "on" and "off" output states of the networks are well separated. The detailed distributions of logical circuits behaviors obtained from the library may depend on aspects of the construction process, including mutagenesis during PCR (*11*).
9. E. H. Davidson, *Genomic Regulatory Systems: Development and Evolution* (Academic Press, San Diego, CA, 2001).
10. S. B. Carroll, *Cell* **101**, 577 (2000).
11. We used sequencing primers located about 60 base pairs into the coding regions of the genes to check the integrity of the upstream regions that contain the important genetic control elements: ribosome-binding sites (RBS), promoters, terminators, and translational stop codons. Sequencing revealed that the construction method had introduced a number of mutations (we detected up to three) in some of the final plasmids. In at least three cases, this altered the behavior of the networks. On the basis of phenotypic behavior, 30 networks were chosen for sequencing. From this set, eight pairs were identified as having identical sets of promoters. Though five of these pairs exhibited matching behaviors, three differed. Detailed sequence analysis revealed several mutations that may explain the difference in behavior: one insertion event in the proximity of a promoter, point mutations within terminator sequences, and point mutations that affected amino acid coding. Among the networks described in Fig. 5, only D016 contained a point mutation (in the *tetR* terminator); none were found in the others. Nevertheless, in order to check whether such mutations have a significant impact on network function, we deliberately reconstructed (by ligation) independent clones of selected networks. In each set, most of the clones shared the same behavior in strain DH10B. For example, the results for networks shown in Fig. 5 are as follows: five out of five reconstructed D016 clones exhibited a NOR phenotype, five out of seven D038 clones behaved as NAND, and six out of seven D052 clones behaved as NOR. These values are typical. Thus the observed phenotypic variability is due to both permutation of promoters and a low level of mutation in some of the networks.
12. L. Glass, S. A. Kauffman, *J. Theor. Biol.* **39**, 103 (1973).
13. R. Thomas, R. D'Ari, *Biological Feedback* (CRC Press, Boca Raton, FL, 1990).
14. Boolean-type models neglect many potentially important intracellular phenomena, including stochastic fluctuations in the levels of components and the detailed biochemistry of protein-DNA interactions. Nevertheless within this framework, we enumerated all Boolean network structures theoretically possible in the library (for the *lac*$^-$ strain CMW101). Only three network structures were expected to depend on both inputs, and they all behaved as NOT IF. However, experimentally we found in addition NORs.
15. We can easily make sense of the NOT IF behavior of the first network (D038) in Fig. 2B, expressed in the *lac*$^-$ strain. Negative autoregulation of TetR maintains LacI levels below the activity threshold of the P^L_2 promoter. Therefore, in order to turn off λ cI, and hence turn on the expression of *gfp*, expression of *lacI* must be induced by aTc and not blocked by IPTG. The use of reporter plasmids for each of the promoters in the network showed that their activities are all consistent with this interpretation. (Data not shown; our set of reporter plasmids was based on the p15A origin of replication and contained a chloramphenicol resistance gene, along with a promoter from the set driving *yfp* expression.) When this plasmid was transformed into the *lac*$^+$ strain, NAND-type behavior, which could not be explained by similar arguments, was observed. The behavior of the second network (D052) in Fig. 2B, expressed in the *lac*$^-$ strain, is not deducible from simple Boolean analysis despite its structural similarity to the first network. A crucial difference is that the threshold for repression of P^T is known to be lower than that for P^L_1. Therefore, the feedback loop leads to sufficient expression of TetR to shut off λ cI production. However, when IPTG is added to the system, λ cI levels evidently increase.
16. RNA polymerase read-through, DNA looping, little-studied effects of low levels of antibiotics on repressor levels (*25*), and interactions between wild-type and destabilized alleles of *lacI* could have significant effects in this system.
17. Similarly, a recent quantitative study of gene expres-

sion in *Drosophila* embryos has revealed the existence of unknown molecular mechanisms in this well-studied prototype system for embryonic development (*26*).
18. W. P. Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10747 (1994).
19. Y. X. Zhang *et al.*, *Nature* **415**, 644 (2002).
20. L. H. Hartwell, J. J. Hopfield, S. Leibler, A. W. Murray, *Nature* **402**, C47 (1999).
21. M. B. Elowitz, S. Leibler, *Nature* **403**, 335 (2000).
22. J. Muller, S. Oehler, B. Muller-Hill, *J. Mol. Biol.* **257**, 21 (1996).
23. R. Lutz, H. Bujard, *Nucleic Acids Res.* **25**, 1203 (1997).
24. B. J. Meyer, R. Maurer, M. Ptashne, *J. Mol. Biol.* **139**, 163 (1980).
25. E. D'Haese, H. J. Nelis, W. Reybroeck, *Appl. Environ. Microbiol.* **63**, 4116 (1997).
26. B. Houchmandzadeh, E. Wieschaus, S. Leibler, *Nature* **415**, 798 (2002).
27. R.L. Mullinax *et al.*, *Biotechniques* **12**, 864 (1992).
28. S. L. Berger, *Anal. Biochem.* **222**, 1 (1994).
29. The backbone plasmid into which the library was cloned contains elements (in this order) as follows: the *kan*^{R} gene surrounded by Dra III sites, the promoter P$^{\lambda}_{-}$ followed by the *gfpmut3* (*21*) allele of GFP, the transcriptional terminator T1, the SC101* origin of replication (*23*), and the ampicillin resistance gene *bla*. This plasmid was digested with Dra III, and the larger fragment containing P$^{\lambda}_{-}$-*gfp*-T1-SC101*-*bla* was gel-purified and used in the ligation reaction for constructing the library. The Dra III sites were designed such that Dra III digestion exposed cohesive ends compatible with the 5′ and 3′ ends of the assembled networks. The library DNA was transformed into *E. coli* strains by electroporation with the use of standard protocols.
30. Plasmid pOd O_{id} 70.5 O_1 was a gift of B. Müller-Hill; source plasmids for $P_L$lacO1, $P_L$tetO1, and SC101* origin of replication were kindly provided by H. Bujard. We would like to thank the Silhavy lab for strain CLC90, A. Beavis for great help and expertise with FACS, and people for helpful discussions as follows: U. Alon, P. Cluzel, S. Da Re, B. Houchmandzadeh, R. Kishony, T. Lecuit, A. J. Levine, A. C. Maggs, J. Merrin, I. Mihalcescu, A. W. Murray, N. Questembert-Balaban, B. I. Shraiman, T. J. Silhavy, M. G. Surette, S. Tavazoie, J. M. G. Vilar. M.B.E. is supported by the Burroughs-Wellcome Fund and the Seaver Institute. This work was partially supported by a grant from the National Institutes of Health.

24 October 2001; accepted 20 March 2002

An LRR Receptor Kinase Involved in Perception of a Peptide Plant Hormone, Phytosulfokine

Yoshikatsu Matsubayashi,* Mari Ogawa, Akiko Morita, Youji Sakagami

The sulfated peptide phytosulfokine (PSK) is an intercellular signal that plays a key role in cellular dedifferentiation and proliferation in plants. Using ligand-based affinity chromatography, we purified a 120-kilodalton membrane protein, specifically interacting with PSK, from carrot microsomal fractions. The corresponding complementary DNA encodes a 1021–amino acid receptor kinase that contains extracellular leucine-rich repeats, a single transmembrane domain, and a cytoplasmic kinase domain. Overexpression of this receptor kinase in carrot cells caused enhanced callus growth in response to PSK and a substantial increase in the number of tritium-labeled PSK binding sites, suggesting that PSK and this receptor kinase act as a ligand-receptor pair.

The relative growth rate of plant cells in culture strictly depends on the initial cell density, even if sufficient amounts of auxin and cytokinin are supplied, indicating that additional factors play a role in cell proliferation (*1, 2*). One such factor is the 5–amino acid peptide phytosulfokine (PSK), which has sulfated Tyr residues [Tyr(SO_3H)-Ile-Tyr(SO_3H)-Thr-Gln] (*3*). Together with auxin and cytokinin, PSK induces plant cells to dedifferentiate and reenter the cell cycle at nanomolar concentrations (*3, 4*). PSK is processed from the COOH-terminal region of ~80–amino acid precursor proteins ubiquitously expressed in the leaf, apical meristem, hypocotyl, and root of seedlings, as well as in suspension cells in culture (*5*). It has been proposed that mature PSK is secreted from individual cells in response to changes in the levels of auxin and cytokinin (*4*) and that it functions as an autocrine-type growth factor to regulate cellular dedifferentiation and proliferation in plants.

Graduate School of Bio-Agricultural Sciences, Nagoya University, Chikusa, Nagoya 464-8601, Japan.

*To whom correspondence should be addressed. E-mail: matsu@agr.nagoya-u.ac.jp

Evidence for the existence of high-affinity binding sites for PSK has been provided by binding assays with radiolabeled PSKs (*6, 7*). The observed binding is saturable, reversible, and localized in plasma membrane fractions. In addition, photoaffinity cross-linking analysis has shown that the putative receptors for PSK in rice plasma membrane are 120- and 160-kD glycosylated proteins (*8*). In the present study, we performed purification, molecular cloning, and functional expression of the PSK receptor to gain further insight into the molecular basis of signal transduction triggered by PSK.

For the purification of membrane proteins that specifically interact with PSK, we used microsomal fractions derived from the carrot cell line NC, which has been found to contain a relatively high concentration of high-affinity PSK-binding proteins: ~150 fmol per mg of microsomal proteins, with a dissociation constant (K_d) of 4.2 ± 0.4 nM, as determined by a [^{3}H]PSK binding assay (*9*) (Fig. 1, A and B). This K_d value is consistent with physiological concentrations of PSK in carrot suspension cultures (*10*), and it is also consistent with the PSK concentration that induces a 50% cell division of dispersed mesophyll cells (*3*). Photoaffinity labeling of NC membrane proteins with a photoactivatable PSK analog (*8*) indicated that a 120-kD protein and a minor 150-kD protein specifically interact with PSK (Fig. 1C). Both proteins contain ~10 kD of N-linked oligosaccharide chains that can be cleaved by treatment with peptide *N*-glycosidase F (PNGase F) (Fig. 1C).

We purified these PSK-binding proteins from the microsomal fractions of NC cells by Triton X-100 solubilization and specific ligand-based affinity chromatography using a [Lys^5]PSK-Sepharose column containing a long spacer chain between the ligand and matrix (*9*) (fig. S1). Elongation of the Lys^5 side chain of [Lys^5]PSK does not interfere with its binding affinity and specificity (*11*). Proteins specifically eluted by PSK were further purified by hydroxyapatite column chromatography and concentrated by ultrafiltration (*9*). SDS–polyacrylamide gel electrophoresis (PAGE) and Nile red staining of the proteins in the fractions eluted by PSK showed specific recovery of a major 120-kD protein and a minor 150-kD protein (Fig. 1D). Both of these proteins were absent in the fractions eluted by [2-5]PSK, a synthetic analog of PSK with no biological or binding activities (*12*) (Fig. 1D). PNGase F treatment of these two proteins decreased their apparent sizes to 110 and 140 kD, respectively, suggesting that they are identical to the proteins we detected in photoaffinity cross-linking experiments (Fig. 1D; see also Fig. 1C).

Four independent purifications were performed, yielding 50 μg of the major 120-kD protein from 4800 mg of microsomal proteins, with an overall recovery rate of 40%. The protein was digested with TPCK-trypsin (TPCK, tosyl phenylalanyl chloromethyl ketone), and peptide fragments thus generated were separated by reversed-phase high-performance liquid chromatography (HPLC) (*9*) (Fig. 1E). We analyzed the fragments of the 120-kD protein contained in 15 independent peaks, using a protein sequencer and MALDI-TOF MS (matrix-assisted laser desorption/ionization time-of-flight mass spec-

Combinatorial Synthesis of Genetic Networks

C#lin C. Guet, Michael B. Elowitz, Weihong Hsing, and Stanislas Leibler

Science, 296 (5572), • DOI: 10.1126/science.1067407