In contrast, the autocorrelation of the intrinsic noise (*16*) decays rapidly: $\tau_{intrinsic} < 10$ min $\ll \tau_{corr}$ (Fig. 4E). Thus, the observed slow fluctuations do not result from intrinsic noise; they represent noise extrinsic to CFP expression (see supporting online text). The concentration of a stable cellular factor would be expected to fluctuate with a time scale of the cell cycle period (*7*, *10*). For instance, even though intrinsic fluctuations in production rates are fast, the difference between the total amounts of YFP and CFP in the symmetric branch experiments has an autocorrelation time of $\tau_{total} = 45 \pm 5$ min (*16*). A similar time scale may well apply to other stable cellular components such as ribosomes, metabolic apparatus, and sigma factors. As such components affect their own expression as well as that of our test genes, extrinsic noise may be self-perpetuating.

These data indicate that the single-cell GRF cannot be represented by a single-valued function. Slow extrinsic fluctuations give the cell and the genetic circuits it comprises a memory, or individuality (*29*), lasting roughly one cell cycle. These fluctuations are substantial in amplitude and slow in time scale. They present difficulty for modeling genetic circuits and, potentially, for the cell itself: In order to accurately process an intracellular signal, a cell would have to average its response for well over a cell cycle—a long time in many biological situations. This problem is not due to intrinsic noise in the output, noise that fluctuates rapidly, but rather to the aggregate effect of fluctuations in other cellular components. There is thus a fundamental tradeoff between accuracy and speed in purely transcriptional responses. Accurate cellular responses on faster time scales are likely to require feedback from their output (*1*, *4*, *6*, *10*, *30*). These data provide an integrated, quantitative characterization of a genetic element at the single-cell level: its biochemical parameters, together with the amplitude and time scale of its fluctuations. Such systems-level specifications are necessary both for modeling natural genetic circuits and for building synthetic ones. The methods introduced here can be generalized to more complex genetic networks, as well as to eukaryotic organisms (*18*).

### References and Notes


1. M. A. Savageau, *Biochemical Systems Analysis* (Addison-Wesley, Reading, MA, 1976).
2. M. Ptashne, *A Genetic Switch: Phage Lambda and Higher Organisms* (Cell Press and Blackwell Science, Cambridge, MA, ed. 2, 1992).
3. H. H. McAdams, L. Shapiro, *Science* **269**, 650 (1995).
4. E. H. Davidson *et al.*, *Science* **295**, 1669 (2002).
5. S. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, *Nature Genet.* **31**, 64 (2002).
6. J. E. Ferrell Jr., E. M. Machleder, *Science* **280**, 895 (1998).
7. M. B. Elowitz, S. Leibler, *Nature* **403**, 335 (2000).
8. T. S. Gardner, C. R. Cantor, J. J. Collins, *Nature* **403**, 339 (2000).
9. A. Becskei, B. Seraphin, L. Serrano, *EMBO J.* **20**, 2528 (2001).
10. N. Rosenfeld, M. B. Elowitz, U. Alon, *J. Mol. Biol.* **323**, 785 (2002).
11. F. J. Isaacs, J. Hasty, C. R. Cantor, J. J. Collins, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7714 (2003).
12. K. S. Koblan, G. K. Ackers, *Biochemistry* **31**, 57 (1992).
13. P. J. Darling, J. M. Holt, G. K. Ackers, *J. Mol. Biol.* **302**, 625 (2000).
14. R. J. Ellis, *Trends Biochem. Sci.* **26**, 597 (2001).
15. M. Mirasoli, J. Feliciano, E. Michelini, S. Daunert, A. Roda, *Anal. Chem.* **74**, 5948 (2002).
16. Materials and methods are available as supporting material on *Science* Online.
17. P. Cluzel, M. Surette, S. Leibler, *Science* **287**, 1652 (2000).
18. G. Lahav *et al.*, *Nature Genet.* **36**, 147 (2004).
19. I. B. Dodd *et al.*, *Genes Dev.* **18**, 344 (2004).
20. B. J. Meyer, R. Maurer, M. Ptashne, *J. Mol. Biol.* **139**, 163 (1980).
21. J. A. Shapiro, *Annu. Rev. Microbiol.* **52**, 81 (1998).
22. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183 (2002).
23. P. S. Swain, M. B. Elowitz, E. D. Siggia, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12795 (2002).
24. J. M. Raser, E. K. O'Shea, *Science* **304**, 1811 (2004).
25. E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, A. van Oudenaarden, *Nature Genet.* **31**, 69 (2002).
26. W. J. Blake, M. Kærn, C. R. Cantor, J. J. Collins, *Nature* **422**, 633 (2003).
27. H. H. McAdams, A. Arkin, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814 (1997).
28. J. Paulsson, *Nature* **427**, 415 (2004).
29. J. L. Spudich, D. E. Koshland Jr., *Nature* **262**, 467 (1976).
30. P. S. Swain, *J. Mol. Bio.* **344**, 965 (2004).


31. We thank Z. Ben-Haim, R. Clifford, S. Itzkovitz, Z. Kam, R. Kishony, A. J. Levine, A. Mayo, R. Milo, R. Phillips, M. Ptashne, J. Shapiro, B. Shraiman, E. Siggia, and M. G. Surette for helpful discussions. M.B.E. is supported by a CASI award from the Burroughs Wellcome Fund, the Searle Scholars Program, and the Seaver Institute. U.A. and M.B.E. are supported by the Human Frontiers Science Program. P.S.S. acknowledges support from a Tier II Canada Research Chair and the Natural Sciences and Engineering Research Council of Canada. N.R. dedicates this work to the memory of his father, Yasha (Yaakov) Rosenfeld.


**Supporting Online Material**

www.sciencemag.org/cgi/content/full/307/5717/1962/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
References and Notes
Movies S1 to S3

29 October 2004; accepted 4 February 2005
10.1126/science.1106914

# Noise Propagation in Gene Networks

**Juan M. Pedraza and Alexander van Oudenaarden***

**Accurately predicting noise propagation in gene networks is crucial for understanding signal fidelity in natural networks and designing noise-tolerant gene circuits. To quantify how noise propagates through gene networks, we measured expression correlations between genes in single cells. We found that noise in a gene was determined by its intrinsic fluctuations, transmitted noise from upstream genes, and global noise affecting all genes. A model was developed that explains the complex behavior exhibited by the correlations and reveals the dominant noise sources. The model successfully predicts the correlations as the network is systematically perturbed. This approach provides a step toward understanding and manipulating noise propagation in more complex gene networks.**

The genetic program of a living cell is determined by a complex web of gene networks. The proper execution of this program relies on faithful signal propagation from one gene to the next. This process may be hindered by stochastic fluctuations arising from gene expression, because some of the components in these circuits are present at low numbers, which makes fluctuations in concentrations unavoidable (*1*). Additionally, reaction rates can fluctuate because of stochastic variation in the global pool of housekeeping genes or because of fluctuations in environmental conditions that affect all genes. For example, fluctuations in the number of available polymerases or in any factor that alters the cell growth rate will change the reaction rates for all genes. Recent experimental studies (*2*–*5*) have made substantial progress identifying the factors that determine the fluctuations in the expression of a single gene. However, how expression fluctuations propagate from one gene to the next is largely unknown. To address this issue, we designed a gene network (Fig. 1A) in which the interactions between adjacent genes could be externally controlled and quantified at the single-cell level.

This synthetic network (*6*) consisted of four genes, of which three were monitored in single *Escherichia coli* cells by cyan, yellow, and red fluorescent proteins (CFP, YFP, and RFP). The first gene, *lacI*, is constitutively transcribed and codes for the lactose repressor, which down-regulates the transcription of the second gene, *tetR*, that is bicistronically transcribed with *cfp*. The gene product of *tetR*, the tetracycline repressor, in turn down-regulates the transcription of the third gene, reported by YFP. The fourth gene, *rfp,* is under

Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed: E-mail: avano@mit.edu

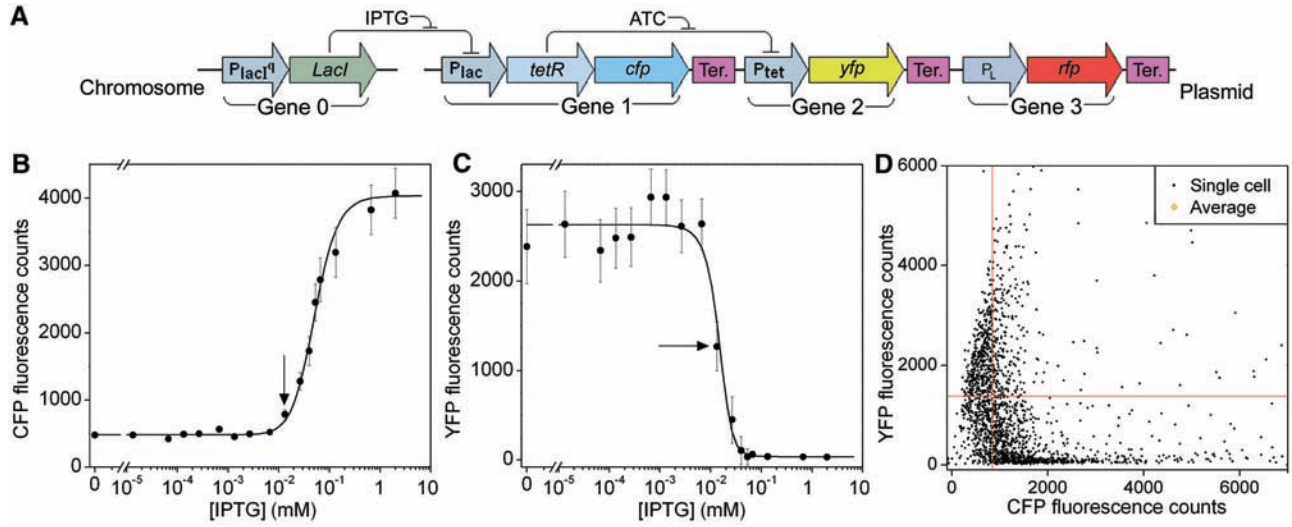


**Fig. 1.** (**A**) A schematic design of the network. (**B** and **C**) Average CFP and YFP expression as a function of IPTG concentration in the steady state. Each experimental data point was obtained from ~2000 single-cell measurements. The solid lines are fits obtained from the Langevin model (*23*). (**D**) Scatter plot of the fluorescence levels for the entire population at [IPTG] = 13 μM. This corresponds to the points marked by the arrows in (B) and (C). The red lines indicate the average CFP and YFP expression.

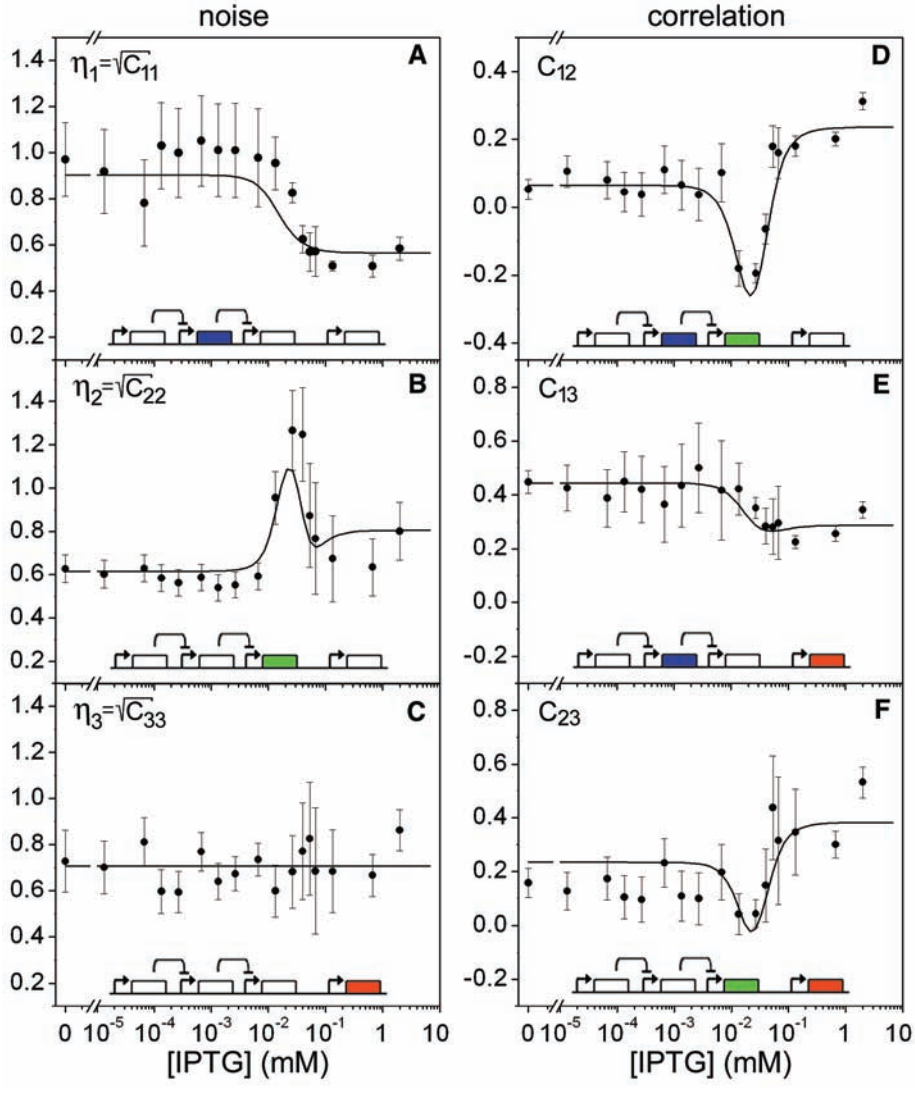


**Fig. 2.** (**A** to **C**) Coefficient of variation $\eta_i = \sqrt{C_{ii}}$ of the expression in genes 1 to 3 as a function of IPTG concentration in the steady state. (**D** to **F**) Correlation between the expression levels of genes 1 and 2, 1 and 3, and 2 and 3, respectively. The solid lines are predictions from the Langevin model (*23*).

the control of the lambda repressor promoter $P_L$, which is a strong constitutive promoter. Because this gene is not part of the cascade, this reporter was used to evaluate the effect of global fluctuations. This cascade was used to measure how fluctuations in an upstream gene (*tetR*, reported by CFP) transmit downstream (and are reported by YFP). The inducers isopropyl-β-D-thiogalactopyranoside (IPTG) and anhydrotetracycline (ATC) bind to and inhibit the repression of the lactose and tetracycline repressors, respectively, and were used to tune, respectively, the expression of the upstream gene and the coupling between the two genes.

We assayed the response of single cells to various amounts of inducers by using automated fluorescence microscopy. In each experimental run, the level of the three fluorescent reporters was quantified for ~2000 individual cells. Figure 1B shows that the average signal of the upstream gene displayed a sigmoidal response to changes in the concentration of IPTG in the growth media. In response, the average signal of the downstream gene (Fig. 1C) behaved inversely and decreased sharply at larger IPTG concentrations. The enhanced sensitivity of the YFP response, compared to the CFP response, when IPTG is varied demonstrates the utility of cascades for generating steep switches (*7*–*10*). However, the average expression alone does not capture the population behavior, because the expression of most cells is quite different from the average (Fig. 1D). Even for a fixed IPTG concentration, the fluctuations in gene expression resulted in a broad distribution that reflects the interaction between the upstream and downstream genes.

To quantify the expression fluctuations and the degree of correlation between different genes, we computed the correlation

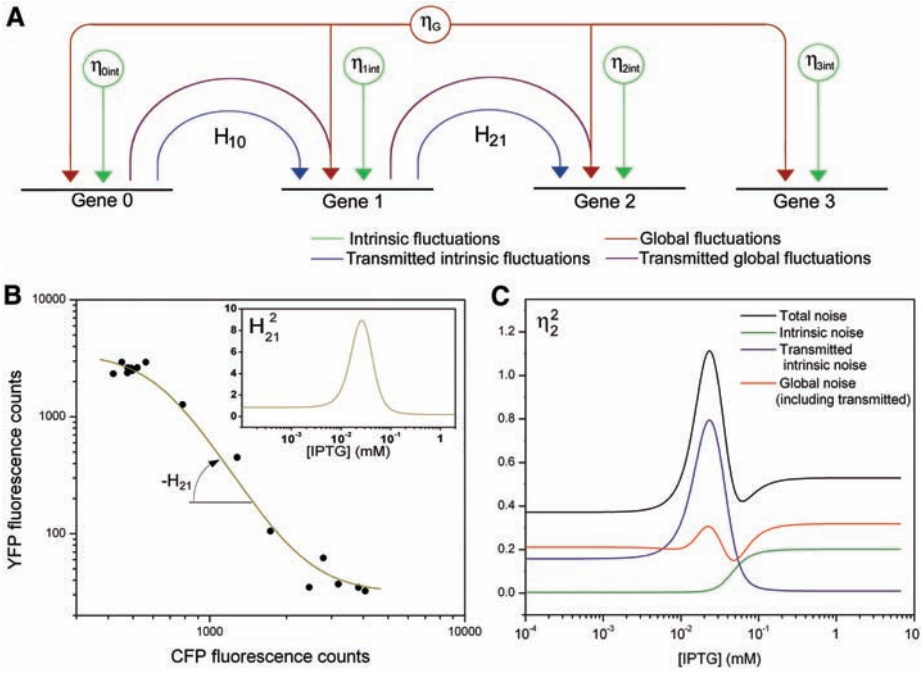


**Fig. 3.** (**A**) A sketch of the propagation of the fluctuations, showing how the two sources of noise, intrinsic and global, can result in many components. (**B**) The logarithmic gain $H_{21}$ is obtained as the negative of the slope in log-log space of the mean expression of YFP as a function of mean CFP expression. (Inset) The square of $H_{21}$ as a function of IPTG. (**C**) Noise in the downstream gene (Fig. 2B) decomposed into the different sources of noise. The total noise (black) is the result of the intrinsic noise in this gene (green), the transmitted noise from the intrinsic fluctuations in upstream genes (blue), and the global noise (red).

$C_{ij} = \frac{\langle F_i F_j \rangle - \langle F_i \rangle \langle F_j \rangle}{\langle F_i \rangle \langle F_j \rangle}$ from the fluorescence levels $F_i$ in individual cells. The brackets $\langle \ldots \rangle$ denote averaging over all cells in the population, and the indices *i* and *j* refer to the gene number as defined in Fig. 1A. Because each cell is characterized by three different expression values ($F_1$, $F_2$, and $F_3$), the statistical properties of this network are summarized by the three self-correlations, $C_{11}$, $C_{22}$, and $C_{33}$, and the three cross-correlations, $C_{12}$, $C_{13}$, and $C_{23}$. The self-correlation is identical to the square of the coefficient of variation, $\eta_i$, which is defined as the standard deviation of the expression distribution normalized to the mean expression. These six correlations were plotted as a function of the IPTG concentration (Fig. 2). The correlations behave in a nonintuitive manner. For example, the noise properties of the upstream gene, reflected in $\eta_1$ (Fig. 2A), are very different from those of the downstream gene, reflected in $\eta_2$ (Fig. 2B), even though both genes are repressed by a single upstream repressor (Fig. 1A). The correlations $C_{13}$ and $C_{23}$ are also dependent on IPTG concentration (Fig. 2, E and F). Because RFP is not part of the cascade, one might expect a correlation that is independent of IPTG.

To clarify these issues, we developed a stochastic model that allows for a systematic interpretation of the data in terms of the different components of the noise. The coefficients of variations and correlations can be derived from the model analytically, enabling a direct fit to the entire experimental data set (*11*). Our model is based on the Langevin approach (*7*, *12*, *13*), in which the deterministic differential equations describing the dynamics of the system are modified by adding stochastic terms (*6*, *14*) that reflect the two sources of noise: intrinsic fluctuations due to low numbers of molecules and global fluctuations in cellular components that change the reaction rates for all genes.

Using the resulting expressions (*15*) for the correlations, we can decompose the noise in each gene into three components: intrinsic noise in that specific gene, transmitted intrinsic noise from the upstream genes, and global noise modulated by the network (Fig. 3A). The intrinsic noise (Fig. 3A, green arrows) arises mostly from low copy numbers of mRNAs (*2*, *3*, *16*). The second noise component, the transmitted intrinsic noise (Fig. 3A, blue arrows), includes the transmitted fluctuations of each of the upstream genes in the network and depends on three factors: the intrinsic noise for that upstream gene; the effect of temporal averaging (*6*, *16*), which depends on the lifetimes of the proteins; and the susceptibility of the downstream gene to the upstream one. We characterize this susceptibility through the logarithmic gain $H_{ji}$ (*16*, *17*) (Fig. 3B). The logarithmic gain reflects how the average expression of the downstream gene *j* changes as the expression of the upstream gene *i* is varied. For example, the main term in the transmitted intrinsic noise from gene 1 to gene 2 (Fig. 1A) is proportional to the squared logarithmic gain $H_{21}^2$ (Fig. 3B, inset). The pronounced peak in $H_{21}^2$ occurs at an IPTG concentration for which the response of the downstream gene is most sensitive to changes in the upstream signal. Consistently, the downstream fluctuations reach a maximum at this concentration (Fig. 2B) (*6*). The last component of the noise reflects the effect of the global fluctuations. It includes the direct effect on the gene, the transmitted effect from the upstream genes (Fig. 3A), and the effect of the correlated transmission, which depends on the interactions. The latter illustrates the main difference between transmitted intrinsic and transmitted global noise. The different intrinsic noise sources are uncorrelated, whereas the global fluctuations arise from the same sources (Fig. 3A). This means that the transmitted global noise (Fig. 3A, purple arrows) does not simply add to the direct global noise (Fig. 3A, red arrows). Because both fluctuations came from the same sources, correction terms arise that depend on the strength (and sign) of the interaction (*15*).

In Fig. 3C, these different noise components are shown for gene 2. The intrinsic component (Fig. 3C, green line) varies as the inverse of the square root of the mean, resulting in increased noise at higher IPTG concentrations. The transmitted intrinsic component (Fig. 3C, blue line) corresponds roughly to the square of the logarithmic gain (Fig. 3B, inset) times the noise in the upstream gene (Fig. 2A) (*18*). The global noise component (Fig. 3C, red line) is not constant but rather shows the modulation as explained above. Thus, the main features of the noise in this gene are determined by the network interactions, rather than by its own intrinsic noise characteristics.

The effect of modulating the global noise is also demonstrated by the behavior of the correlations between noninteracting genes (Fig. 2, E and F). A global fluctuation that raises the expression of RFP will also raise the expression of YFP and CFP. An increased CFP expression will result in a decreased YFP expression by an amount that depends on the interaction between gene 1 and gene 2 and hence will vary with IPTG (*19*). This can be seen in the expression for the correlations (*15*). A consequence of this modulation is that the correlations $C_{12}$ and $C_{23}$ display qualitatively similar behavior as IPTG is varied (Fig. 2, D and F). This indicates that $C_{23}$ is dominated by the global noise that is transmitted from gene 1 to gene 2. Similarly, the correlation $C_{13}$ is dominated by the global noise transmitted from gene 0 to gene 1 and therefore displays a different behavior compared to $C_{12}$ and $C_{23}$ (*6*).

We directly quantified the intrinsic and extrinsic noise for genes 1 to 3 as a function

of the IPTG concentration (Fig. 4, A and B) by measuring the correlation between CFP and YFP in constructs in which both reporters were driven by the same promoter (*3*, *5*, *6*). The total noise was generally dominated by extrinsic fluctuations. The experimentally obtained intrinsic and extrinsic noise of genes 1 and 2 was consistent with the predictions of the model.

To probe the predictive power of the stochastic model, we used it to predict the noise and correlations as the coupling between genes 1 and 2 was altered by adding ATC to the growth media (*6*). We compared these predictions to experimental results. As an example, $\eta_2$ and $C_{12}$ are shown in Fig. 4, C and D. Both $\eta_2$ and $C_{12}$ display rich behavior as a function of both the IPTG and ATC concentrations. As is seen in Fig. 4C, a small perturbation to the network can transform a maximum in the $\eta_2$-IPTG curve (Fig. 4C, black) into a step (red) or even a minimum (green). These features were faithfully predicted by the model (Fig. 4D). Similarly, the model correctly predicts correlation $C_{12}$ (Fig. 4, E and F) and the other correlations (*6*). These experiments demonstrate that the stochastic model is not only descriptive but also has predictive power and can therefore be used as a design tool for synthetic circuits.

Our results show that the noise in a gene affects expression fluctuations of its downstream genes. This transmitted noise can be calculated from the interactions between upstream and downstream genes as quantified by the logarithmic gains. Thus, it is not necessary to have low numbers of molecules to have large fluctuations, because noise could be transmitted from upstream genes. We show that the noise has a correlated global component that is modulated by the network. Thus, even in a network where all components have low intrinsic noise, fluctuations can be substantial and the distributions of expression levels depend on the interactions between genes. Measuring the correlation between a constitutive gene and a gene embedded in a network provides a sensitive probe for correlated sources of noise. This would have been difficult to reveal by monitoring single genes (*2*, *4*) or two copies of the same gene (*3*, *5*). Our results highlight the importance of including stochastic effects in the study of regulatory networks. This will be necessary for understanding faithful signal propagation in natural networks (*20*) as well as for designing noise-tolerant synthetic circuits (*21*).

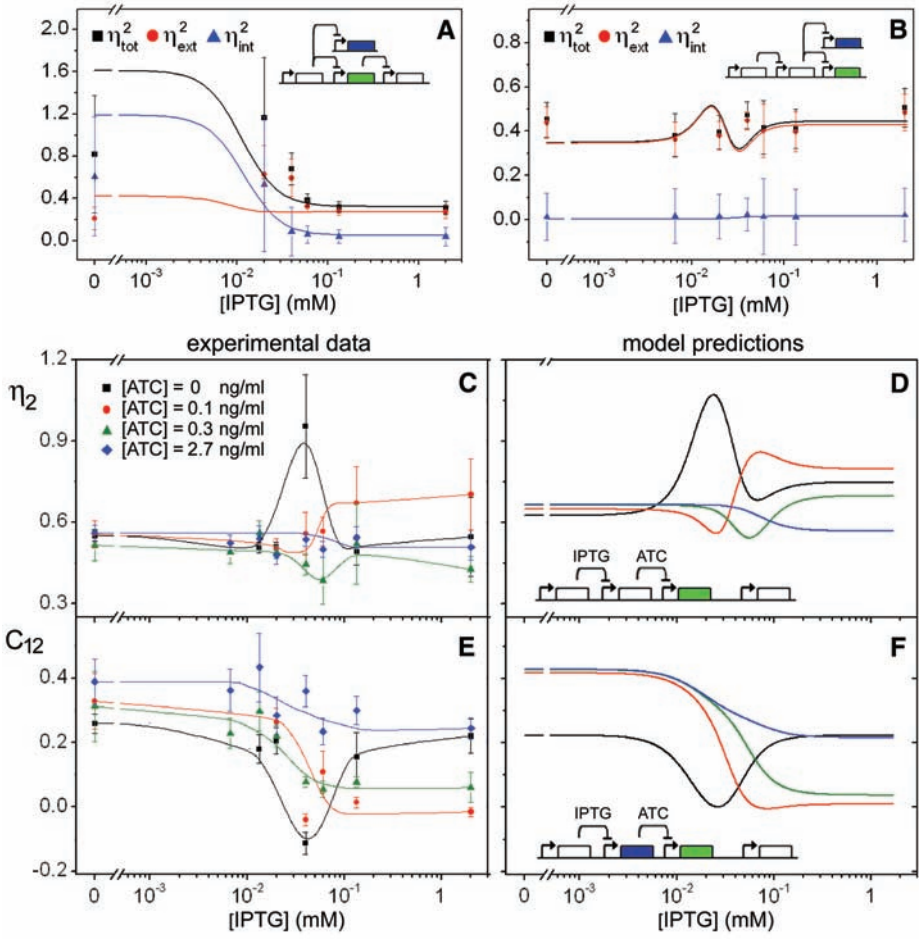


**Fig. 4.** (**A** and **B**) Experimentally determined intrinsic and extrinsic noise as a function of IPTG (*3*, *6*). The solid lines represent predictions by the stochastic model. (A) Two copies of the *lac* promoter are driving CFP and YFP. (B) Two copies of the *tet* promoter are driving CFP and YFP. The model parameters used are those in table S1, except for the basal transcription, which was adjusted to the measured value (*6*). (**C** and **E**) Coefficient of variation $\eta_2$ and correlation $C_{12}$ as a function of IPTG concentration in the steady state for different levels of ATC. The solid lines are guides to the eye. Each experimental data point was obtained from ~1000 single-cell measurements (*23*). (**D** and **F**) Predictions for $\eta_2$ and $C_{12}$ from the Langevin model, given the parameters obtained previously (*6*).

## References and Notes


1. C. V. Rao, D. M. Wolf, A. P. Arkin, *Nature* **420**, 231 (2002).
2. E. M. Ozbudak *et al.*, *Nature Genet.* **31**, 69 (2002).
3. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183 (2002).
4. W. J. Blake, M. Kaern, C. R. Cantor, J. J. Collins, *Nature* **422**, 633 (2003).
5. J. M. Raser, E. K. O'Shea, *Science* **304**, 1811 (2004).
6. Materials and methods are available as supporting material on *Science* Online.
7. M. Thattai, A. van Oudenaarden, *Biophys. J.* **82**, 2943 (2002).
8. P. B. Chock, E. R. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2766 (1977).
9. A. Goldbeter, D. E. Koshland Jr., *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6840 (1981).
10. J. E. Ferrell Jr., E. M. Machleder, *Science* **280**, 895 (1998).
11. We fitted the analytical results to all means, noises, and correlations simultaneously. To test the resulting parameters (table S1), we conducted Monte Carlo simulations of the network using Gillespie's stochastic simulation algorithm (*22*). The results from the Langevin model are consistent with these simulations.
12. T. B. Kepler, T. C. Elston, *Biophys. J.* **81**, 3116 (2001).
13. J. Hasty *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2075 (2000).
14. N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
15. The interaction between two genes is determined by the rate of synthesis from the downstream gene as a function of the concentration of upstream proteins ($y_i$); this will be called the transfer function. Assuming that the interactions are Hill-type repressions, the parameters for the transfer function $f_2(y_1)$ for the repression between genes 1 and 2 can be obtained from the two means, up to the conversion constants from fluorescence counts to protein numbers. To match the notation in previous studies (*16*), we define the logarithmic gain corresponding to genes $i$ and $j$ as $H_{ji} = -\frac{\bar{y}_j}{\bar{y}_i}\frac{1}{\gamma_j}\frac{\partial f_j(y_i)}{\partial y_i}\big|_{\bar{y}_i}$, where $\gamma_j$ is the decay rate of gene $j$ and the overbar denotes the steady state average. The global and plasmid noise are characterized by the parameters $\eta_G$ and $\eta_N$, respectively (*6*). Subscripts 0 to 3 correspond to genes *lacI*, *cfp*, *yfp*, and *rfp* (Fig. 1A). $\eta_{i\ \mathrm{int}}$ denotes the intrinsic noise in gene $i$. In this notation, the total noise in genes 0 to 3 is given by

$$\eta_0^2 = \eta_{0\,\mathrm{int}}^2 + \eta_G^2$$

$$\eta_1^2 = \eta_{1\,\mathrm{int}}^2 + \frac{1}{2}H_{10}^2\eta_{0\,\mathrm{int}}^2 + \eta_G^2\left(1 + \frac{1}{2}H_{10}^2 - H_{10}\right) + \frac{1}{2}\eta_N^2$$

$$\eta_2^2 = \eta_{2\,\mathrm{int}}^2 + \frac{1}{2}H_{21}^2\eta_{1\,\mathrm{int}}^2 + \frac{3}{8}H_{21}^2H_{10}^2\eta_{0\,\mathrm{int}}^2 + \eta_G^2\left(1 + \frac{1}{2}H_{21}^2 + \frac{3}{8}H_{21}^2H_{10}^2 - H_{21} - \frac{3}{4}H_{21}^2H_{10} + \frac{1}{2}H_{21}H_{10}\right) + \eta_N^2\left(\frac{1}{2} + \frac{3}{8}H_{21}^2 - \frac{3}{4}H_{21}\right)$$

$$\eta_3^2 = \eta_{3\,\mathrm{int}}^2 + \eta_G^2 + \frac{1}{2}\eta_N^2$$

The expressions for the correlations are

$$C_{12} = -\frac{1}{2}H_{21}\eta_{1\,\mathrm{int}}^2 - \frac{3}{8}H_{21}H_{10}^2\eta_{0\,\mathrm{int}}^2 + \eta_G^2\left(1 - \frac{1}{2}H_{21} - \frac{1}{2}H_{10} - \frac{3}{8}H_{21}H_{10}^2 + \frac{3}{4}H_{21}H_{10}\right) + \eta_N^2\left(\frac{1}{2} - \frac{3}{8}H_{21}\right)$$

$$C_{13} = \eta_G^2\left(1 - \frac{1}{2}H_{10}\right) + \frac{1}{2}\eta_N^2$$

$$C_{23} = \eta_G^2\left(1 - \frac{1}{2}H_{21} + \frac{1}{4}H_{21}H_{10}\right) + \eta_N^2\left(\frac{1}{2} - \frac{3}{8}H_{21}\right)$$


16. J. Paulsson, *Nature* **427**, 415 (2004).
17. M. A. Savageau, *Biochemical Systems Analysis* (Addison-Wesley, Reading, MA, 1976).
18. The expression for the transmitted noise in $\eta_2$ cannot be rewritten as the transmission of the total steady-state noise in gene 1 only (*6*, *15*).
19. J. Paulsson, M. Ehrenberg, *Q. Rev. Biophys.* **34**, 1 (2001).
20. L. H. Hartwell *et al.*, *Nature* **402** (suppl.), C47 (1999).
21. J. Hasty, D. McMillen, J. J. Collins, *Nature* **420**, 224 (2002).
22. D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
23. Duplicated measurements are averaged. The error bars reflect the standard deviation of run-to-run differences and the error within each measurement as determined by bootstrapping.



24. We thank T. S. Gardner, J. J. Collins, R. Lutz, and H. Bujard for their kind gift of plasmids; M. Thattai, A. Becskei, and H. Lim for helpful discussions and suggestions; and B. Kaufmann for his help with the initial constructs. Supported by NSF CAREER grant no. PHY-0094181 and NIH grant no. R01-GM068957.


**Supporting Online Material**

www.sciencemag.org/cgi/content/full/307/5717/1965/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 and S2
References and Notes


23 December 2004; accepted 18 February 2005
10.1126/science.1109090


# RNA-Dependent Cysteine Biosynthesis in Archaea


**Anselm Sauerwald,[1] Wenhong Zhu,[3] Tiffany A. Major,[4] Hervé Roy,[5] Sotiria Palioura,[1] Dieter Jahn,[6] William B. Whitman,[4] John R. Yates 3rd,[3] Michael Ibba,[5] Dieter Söll[1,2]***



**Several methanogenic archaea lack cysteinyl–transfer RNA (tRNA) synthetase (CysRS), the essential enzyme that provides Cys-tRNA$^{Cys}$ for translation in most organisms. Partial purification of the corresponding activity from *Methanocaldococcus jannaschii* indicated that tRNA$^{Cys}$ becomes acylated with *O*-phosphoserine (Sep) but not with cysteine. Further analyses identified a class II–type *O*-phosphoseryl-tRNA synthetase (SepRS) and Sep-tRNA:Cys-tRNA synthase (SepCysS). SepRS specifically forms Sep-tRNA$^{Cys}$, which is then converted to Cys-tRNA$^{Cys}$ by SepCysS. Comparative genomic analyses suggest that this pathway, encoded in all organisms lacking CysRS, can also act as the sole route for cysteine biosynthesis. This was proven for *Methanococcus maripaludis*, where deletion of the SepRS-encoding gene resulted in cysteine auxotrophy. As the conversions of Sep-tRNA to Cys-tRNA or to selenocysteinyl-tRNA are chemically analogous, the catalytic activity of SepCysS provides a means by which both cysteine and selenocysteine may have originally been added to the genetic code.**


The translation of cysteine codons in mRNA during protein synthesis requires cysteinyl-tRNA (Cys-tRNA$^{Cys}$). Cys-tRNA$^{Cys}$ is normally synthesized from the amino acid cysteine and the corresponding tRNA isoacceptors (tRNA$^{Cys}$) in an adenosine triphosphate (ATP)–dependent reaction catalyzed by cysteinyl-tRNA synthetase (CysRS). Genes encoding CysRS, *cysS*, have been detected in hundreds of organisms encompassing all three living domains (*1*). The only exceptions are certain methanogenic archaea, the completed genome sequences of which encode no open reading frames (ORFs) with obvious homology to known *cysS* sequences (*1*). Because of the discovery that the genomes of a number of methanogenic archaea either lack *cysS* (*Methanocaldococcus jannaschii*, *Methanothermobacter thermautotrophicus*, and *Methanopyrus kandleri*) or can dispense with it (*Methanococcus maripaludis*), the formation of Cys-tRNA$^{Cys}$ in these organisms has been a much studied and increasingly contentious topic (*2*, *3*). A noncognate aminoacyl-tRNA synthetase [aaRS (*4*–*6*)] and a previously unassigned ORF (*7*) were variously implicated in Cys-tRNA$^{Cys}$ formation. Recent studies failed to provide conclusive support for either of these routes, leaving the mechanism of Cys-tRNA$^{Cys}$ formation still in doubt (*2*).

Previous investigations of archaeal Cys-tRNA$^{Cys}$ biosynthesis have been hampered by the significant levels of noncognate tRNA routinely cysteinylated and detected by conventional filter binding assays. This problem was circumvented with a more stringent assay of Cys-tRNA$^{Cys}$ formation: gel-electrophoretic separation of uncharged tRNA from aminoacyl-tRNA (aa-tRNA) and subsequent detection of the tRNA moieties by sequence-specific probing (*8*). Given that *M. jannaschii* is a strict anaerobe, and considering that earlier aerobic purification erroneously identified prolyl-tRNA synthetase (*4*, *5*), we used anaerobic conditions for all procedures unless otherwise indicated. When these procedures were used to monitor acylation of total *M. maripaludis* tRNA by an undialyzed *M. jannaschii* cell-free extract (S-100), tRNA$^{Cys}$ was charged with an amino acid that gave rise to the same mobility shift (*9*) exhibited by standard *M. maripaludis* Cys-tRNA$^{Cys}$ generated by *M. maripaludis* CysRS (*1*) (Fig. 1A, lanes 7 and 8). Further optimization of the reaction at this stage showed that $Zn^{2+}$ and ATP were also required for the successful formation of charged tRNA$^{Cys}$. When the S-

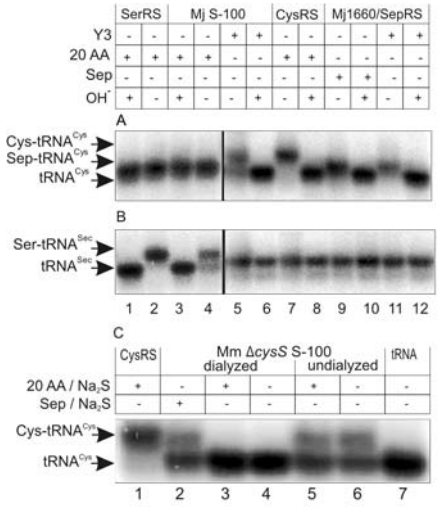


**Fig. 1.** Acid urea gel electrophoresis and Northern blot analysis of total *M. maripaludis* tRNA charged with *M. maripaludis* SerRS, dialyzed *M. jannaschii* S-100, *M. maripaludis* CysRS, and *M. jannaschii* SepRS in the presence of 20 amino acids (20 AA), phosphoserine, or a *M. jannaschii* S-100 cell-free extract filtrate (Y3). Half of each tRNA sample was deacylated by mild alkaline hydrolysis ($^-$OH). The blots were probed with $^{32}$P-labeled oligonucleotides complementary to *M. maripaludis* tRNA$^{Cys}$ (**A**) and *M. maripaludis* tRNA$^{Sec}$ (**B**). Total *M. maripaludis* tRNA charged with dialyzed or undialyzed *M. maripaludis* Δ*cysS* S-100 cell-free extract (*20*) in the presence of 20 amino acids and $Na_2S$, or Sep and $Na_sS$ (**C**). The blot was analyzed with $^{32}$P-labeled oligonucleotides complementary to *M. maripaludis* tRNA$^{Cys}$.


[1]Department of Molecular Biophysics and Biochemistry, and [2]Department of Chemistry, Yale University, New Haven, CT 06520–8114, USA. [3]Department of Cell Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [4]Department of Microbiology, University of Georgia, Athens, GA 30602–2605, USA. [5]Department of Microbiology, The Ohio State University, Columbus, OH 43210–1292, USA. [6]Department of Microbiology, Technische Universität Braunschweig, D-38106 Braunschweig, Germany.

*To whom correspondence should be addressed. E-mail: soll@trna.chem.yale.edu