Using noise to probe and characterize gene circuits

Chris D. Cox*†‡, James M. McCollum§, Michael S. Allen¶‖, Roy D. Dar¶, and Michael L. Simpson‡¶††**

*Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996-2010; †Center for Environmental Biotechnology, University of Tennessee, Knoxville, TN 37996-1605; §Department of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, VA 23284; ¶Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6006; ‖Department of Biological Sciences, University of North Texas, Denton, TX 76203-5220; **Department of Physics and Astronomy, University of Tennessee, Knoxville, TN 37996-1200; and ††Department of Materials Science, University of Tennessee, Knoxville, TN 37996-2200

Communicated by E. Ward Plummer, University of Tennessee, Knoxville, TN, May 21, 2008 (received for review January 24, 2008)

Stochastic fluctuations (or ''noise'') in the single-cell populations of molecular species are shaped by the structure and biokinetic rates of the underlying gene circuit. The structure of the noise is summarized by its autocorrelation function. In this article, we introduce the noise regulatory vector as a generalized framework for making inferences concerning the structure and biokinetic rates of a gene circuit from its noise autocorrelation function. Although most previous studies have focused primarily on the magnitude component of the noise (given by the zero-lag autocorrelation function), our approach also considers the correlation component, which encodes additional information concerning the circuit. Theoretical analyses and simulations of various gene circuits show that the noise regulatory vector is characteristic of the composition of the circuit. Although a particular noise regulatory vector does not map uniquely to a single underlying circuit, it does suggest possible candidate circuits, while excluding others, thereby demonstrating the probative value of noise in gene circuit analysis.

gene circuit analysis | noise analysis | stochastic gene expression

The emerging field of noise biology focuses on the sources, processing, and biological consequences of the inherent stochastic fluctuations in the populations, concentrations, positions, or states of molecules that control cellular function and behavior (1–22). Most of these studies have been directed toward the question: what are the noise characteristics that emerge from a particular well defined gene circuit? An intriguing prospect emerges when the question is turned around: to what extent can measured noise characteristics provide information regarding the underlying gene circuit? A few examples of this type of inference have been published. In two recent flow cytometry studies of *Saccharomyces cerevisiae* (2, 14), the observed relationship between noise and protein expression level was interpreted to be consistent with a predominance of intrinsic noise originating from fluctuations in mRNA levels. Another study (17) inferred that protein mixing times (the time scale required for the protein level in a single cell to transition from higher than the population average to lower than the population average) exceeding one cell generation could originate from upstream noise sources or the presence of feedback loops. In this article, we develop a generalized framework for probing the circuit of a particular gene via its intrinsic noise.

The noise behavior of a given gene circuit is given by:

$$\vec{N} = g(m(S, k)), \qquad [1]$$

where $\vec{N}$ is the noise structure vector (defined below), m is a model of the system comprised of circuit structure (S; architecture or regulatory arrangements) and circuit rate parameters (k; e.g., kinetic rates), and g is an analysis or simulation operation that reveals the noise structure of the model. Eq. **1** states that the noise structure is uniquely determined by circuit structure and parameters. Here, we seek to find another function, $\hat{g}$, such that $m(S, k) = \hat{g}(\vec{N})$; that is, we wish to determine the architecture and kinetic rates from a circuit's noise structure. In general terms, our method is based on a comparison of the measured noise structure of the circuit of interest to the theoretical noise structure of an assumed model (Fig. 1*A*). In this article, we describe the process by which these two signals are processed to yield the noise regulatory vector that points toward a family of possible circuits consistent with the measured signal and away from other possible circuits that are incompatible with the measured signal (Fig. 1*A*).

The noise regulatory vector $\Delta\vec{N}_r$ is defined as:

$$\Delta\vec{N}_r = \vec{N}_m - \vec{N}_A \qquad [2]$$

where $\vec{N}_m$ is the measured noise structure vector for the circuit of interest, and $\vec{N}_A$ is the theoretical noise structure vector for the assumed *a priori* circuit model $m(S_A, k_A)$. The assumed structure S_A is based on *a priori* knowledge of the circuit, which may not be complete. Through experimental observations Ω, we can select parameters k_A, such that the model $m(S_A, k_A)$ maps onto the observations $f(m(S_A, k_A)) \Rightarrow \Omega$, where f is a solution of the model that can be compared with Ω.

This framework is general, in that it may be adapted to any *a priori* circuit structure S_A, appropriate experimental observables Ω, and defined noise structure vector $\vec{N}$. For example, in practice, one may specify the *a priori* circuit to include everything known about a particular gene circuit. In this work, we assume S_A to be a constitutive transcription-translation circuit [Fig. 1*b*, supporting information (SI) *Text*, and Table S1]. This definition is appropriate in the case where little is known about the gene-specific regulatory mechanisms on which S_A is defined. Furthermore, it allows k_A to be uniquely defined from Ω consisting of only abundance and stability of mRNA and protein molecules (see *Methods*).

The complete noise structure of a molecular population or concentration $[M(t)]$ is contained in the autocorrelation function [ACF, $\Phi(\tau)$], which is defined as

$$\Phi(\tau) = E[(M(t) - \langle(t)\rangle)(M(t + \tau) - \langle M(t)\rangle)], \qquad [3]$$

where $\langle M(t)\rangle$ is the average value of $M(t)$, and $E\cdot$ returns the expected value of the term within the bracket. In an ideal case, $\Delta\vec{N}$ would be the difference between measured and *a priori* ACFs, but in practice, there are experimental constraints that limit the accuracy of measured ACFs. The accuracy of the ACF, particularly at larger values of τ, is compromised by the limited number of cells tracked and the limited duration of observation in time-lapse fluorescence microscopy (1, 22). Here, we will consider a noise structure $\vec{N}$ represented by two components: (*i*)

Author contributions: C.D.C., J.M.M., and M.L.S. designed research; C.D.C. and J.M.M. performed research; C.D.C., J.M.M., M.S.A., R.D.D., and M.L.S. analyzed data; and C.D.C., J.M.M., M.S.A., and M.L.S. wrote the paper.

The authors declare no conflict of interest.

‡To whom correspondence may be addressed. E-mail: ccox9@utk.edu or simpsonml1@ornl.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0804829105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

BIOPHYSICS

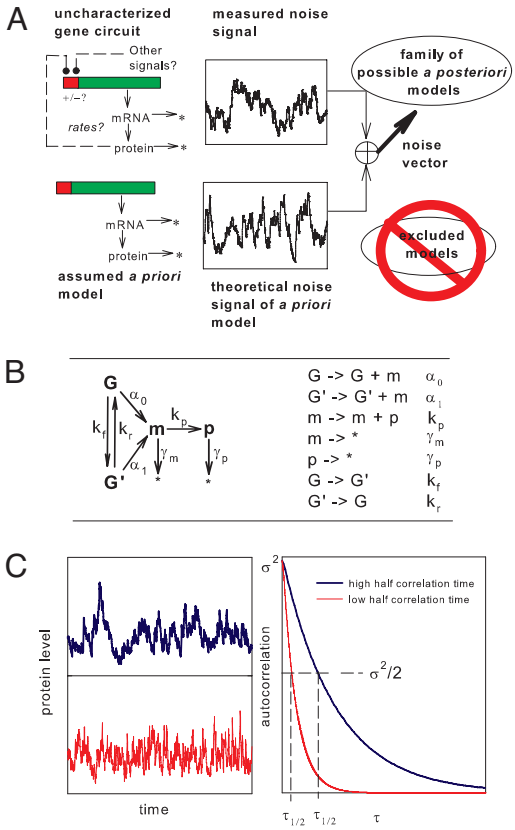

Fig. 1. Noise regulatory vector and its application in the analysis of gene circuits. (*A*) The noise regulatory vector for an uncharacterized gene circuit is determined by comparison of its experimental noise structure to the noise structure of an assumed model. The vector points toward a family of gene circuits that includes the true gene circuit, and away from inappropriate models. (*B*) The assumed structure of the *a priori* model is a constitutive transcription-translation circuit. (*C*) Time series and autocorrelation functions for two stochastic protein populations characterized by identical $\langle p \rangle$ and CV^2 but different half correlation times $\tau_{1/2}$.

noise magnitude, which is a measure of the size of fluctuations; and (*ii*) noise correlation (or frequency content), which is an indication of the dynamic responsiveness of the protein level to changes in the environmental or physiological state of the cell. The relationship between $\Phi(\tau)$ and the components of $\vec{N}$ is shown in Fig. 1*C*. The noise magnitude is quantified by $\Phi(0)$, which is the noise variance σ^2 and is often normalized by the square of the mean to give $CV^2 = \Phi(0)/\langle M(t)\rangle^2$, where CV is known as the coefficient of variation. We quantify correlation using the half correlation time ($\tau_{1/2}$), which is defined by $\Phi(\tau_{1/2}) = \Phi(0)/2$; that is, $\tau_{1/2}$ defines the time for correlation to fall to ½ of its original value. With these two components, we choose to define the noise structure and regulatory vectors in this work as (see *SI Text*):

$$\vec{N} \equiv \log CV^2 \hat{m} + \log \tau_{1/2} \hat{c}$$

$$\Delta\vec{N}_r = \Delta \log CV^2 \hat{m} + \Delta \log \tau_{1/2} \hat{c},$$
$$= \log(CV_m^2/CV_A^2)\hat{m} + \log(\tau_{1/2,m}/\tau_{1/2,A})\hat{c}, \qquad [4]$$

where $\hat{m}$ and $\hat{c}$ are log-scale unit vectors on the magnitude and correlation axes, respectively, and subscripts m and A denote measured and assumed quantities, respectively. The noise vectors are defined in log-space to capture the natural scaling of noise in biological systems. To interpret $\Delta\vec{N}_r$, we compare it to a catalog of theoretical regulatory vectors $\Delta\vec{N}_{\mathbf{rT}} = g(\mathbf{m}(S, k))$ for models over a wide range of S and k, to suggest one or more *a posteriori* models $m(S_P, k_P)$ to better capture the true nature of the gene circuit. In this way, the noise regulatory vector "points" toward a family of possible *a posteriori* models that includes the true gene circuit and away from inappropriate models. In the remainder of this article, we illustrate the probative value of noise in gene circuit analysis by expanding on the conceptual rationale for the above framework and developing the theoretical noise regulatory vectors for several common regulatory motifs.

Results

***A priori* Noise Vectors.** We now present an expanded development of the noise regulatory vector, beginning with the definition of the noise vector $\vec{N}_A$ for the *a priori* model $m(S_A,k_A)$. We illustrate the definition of $m(S_A,k_A)$ by assuming a simple transcription-translation circuit (S_A) and using it to interpret experimental observables (Ω) consisting of a compilation of several genome-scale databases characterizing the abundance and stability of mRNA and proteins in *S. cerevisiae* (see *Methods*). Our model of the intrinsic noise of protein synthesis is based on a system of first-order reactions (Fig. 1*B*; details provided in *SI Text* and Table S1). We further assume that gene activation kinetics are fast ($k_f + k_r \gg \alpha_0, \gamma_m, \gamma_p$), and that $\alpha_1 = 0$. Under these assumptions, the remaining parameters k_A may be obtained directly from Ω (see *Methods*).

We define noise vectors in the context of a 3D plot of the two noise component axes CV^2 and $\tau_{1/2}$ and a mean protein axis $\langle p \rangle$, which we term the 3D noise map. Two orthogonal planes of the 3D noise map for *S. cerevisiae* are shown in Fig. 2*A*, which should be considered an illustrative example due to limitations in the compiled database (see *Methods*). Each point in the noise map in Fig. 2*A* represents the noise characteristics of a particular protein. However, through the action of the regulatory mechanism, the expression level of the gene is likely to change in response to physiological or environmental conditions. The manner in which the noise characteristics change with changing expression level of the gene contains information about the underlying regulatory motif. Bar Even *et al.* (2) and Newman *et al.* (14) interpreted the $CV^2 \propto 1/\langle p \rangle$ scaling observed in their studies, present in our database-derived noise map (Fig. 2*A*), to be consistent with a predominance of intrinsic noise originating from fluctuations in mRNA levels and a relatively low contribution of extrinsic noise at low and moderate values of $\langle p \rangle$. However, the ability to infer regulatory mechanisms from noise characteristics may be enabled by consideration of the $\tau_{1/2}$ component of the noise, which has been ignored in most gene noise investigations to date. For example, analysis of Eqs. **4** and **8** indicates that modulation of $\langle p \rangle$ via changes in protein stability also yields the observed $CV^2 \propto 1/\langle p \rangle$ relationship, under conditions where protein stability is much greater than mRNA stability ($\gamma_m \gg \gamma_p$). It is impossible to distinguish between the two mechanisms of protein modulation based on the CV^2 vs. $\langle p \rangle$ view alone. However, as we show below, the two mechanisms can easily be distinguished when both the CV^2 and $\tau_{1/2}$ components of noise are considered together, because the latter component is affected by changes in protein stability but not by changes in transcription rate. In fact, the distribution of $\tau_{1/2}$ values observed in Fig. 2*B* suggests that variability in protein stability may have contributed to the observed pattern of $CV^2 \propto 1/\langle p \rangle$ scaling.

We incorporate the $CV^2 \propto 1/\langle p \rangle$ scaling in our *a priori* model by stipulating that modulation of $\langle p \rangle$ is achieved through variation of the effective transcription rate. With these assumptions, the noise vector of the *a priori* model is given by:

$$\vec{N}_A \equiv [c_1 - \log(\langle p \rangle)]\hat{m} + c_2\hat{c}, \qquad [5]$$

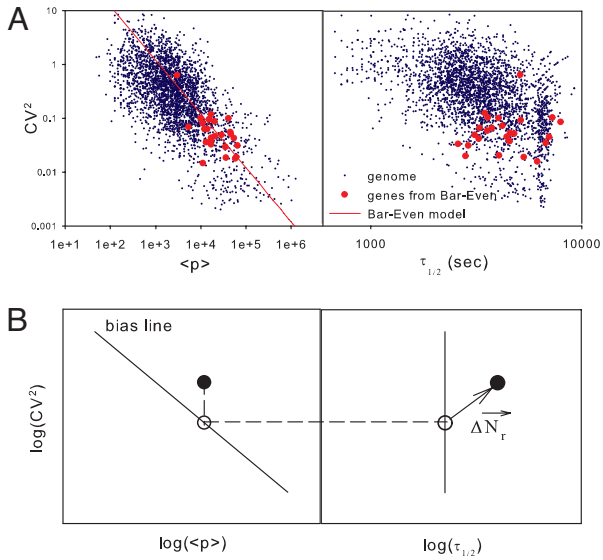

Fig. 2. Relationship of 3D noise map and noise regulatory vectors, $\Delta\vec{N}_r$. (*A*) Noise map for 2,920 ORFs in *S. cerevisiae*. A subset of 25 proteins that were both included in the Bar-Even *et al.* (2) study and present in our compiled database (shown in red) are observed to scatter about the Bar-Even (2) noise model $CV^2 = 1{,}200/\langle p\rangle$ (red line), suggesting a similarity in measured noise trends [Bar-Even (2)] and those inferred from literature data (blue points). (*B*) Graphical definition of the bias line and $\Delta\vec{N}_r$. The bias line represents the behavior of the *a priori* model. To determine $\vec{N}\Delta_r$ for a protein with measured coordinates on the noise map (filled circle), one first locates the bias point (open circle) by projecting vertically to the bias line. $\Delta\vec{N}_r$ is defined by the 2D vector connecting the bias point to the measured point in the $\log(CV^2)-\log(\tau_{1/2})$ plane.

where c_1 and c_2 are protein-specific constants (see Eq. **7** in *SI Text*), which are functions of Ω measured under standard (defined) physiological and environmental conditions. Experimental techniques for measuring the mRNA and protein abundances and half-lives are reviewed in the *SI Text*. The line defined by the locus of all $\vec{N}_A$ over all values of $\langle p\rangle$ is termed the bias line (see Fig. 2*B*). In the case of our *a priori* model, the bias line represents the hypothetical intrinsic noise characteristics of the gene under the assumption that its activity is regulated by rapid, noise-free modulation of the effective transcription rate. The bias line is hypothetical, because the actual regulatory mechanism of the gene may be quite different from the *a priori* model. The bias line is a reference state from which we quantify deviations in noise characteristics attributable to alternative regulatory mechanisms, such as modulation of translation rate, modification of protein, or mRNA stability, slow or noisy modulation of transcription, or positive or negative feedback.

Graphically, the noise regulatory vector $\Delta\vec{N}_r$ is defined as the vector in the $\log(CV^2)$ vs. $\log(\tau_{1/2})$ plane (i.e., $\Delta\vec{N}_r$ as defined here is 2D), which quantifies the deviation of the measured noise characteristics from the bias line at the measured value of $\langle p\rangle$ (Fig. 2*B*, see *SI Text* Eq. **8** for analytical description). Measurements of $\langle p\rangle$, CV^2, and $\tau_{1/2}$ are made using time-lapse microscopy (1, 5, 15, 22, 23). These measurements need not be made under the same experimental conditions used to define the bias line, because the goal is to make inferences of the underlying regulatory mechanism as it responds to changes in physiological and environmental conditions. We now develop theoretical noise regulatory vectors $\Delta\vec{N}_{rT}$ for several common regulatory mechanisms.

Fast Operator Kinetics. We first analyze the simple transcription-translation model shown in Fig. 1*B* for the case where the operator kinetics are fast. This model is identical to the *a priori* model, except we relax the assumption that changes in $\langle p\rangle$ are modulated only by variation of $k_{m,\mathrm{eff}}$ and consider also changes in k_p, γ_m, and γ_p. We first consider the case in which mRNA is more stable than proteins ($\gamma_m/\gamma_p \gg 1$) and high translational efficiency ($b \gg 1$), which is applicable to many prokaryotic and some eukaryotic proteins. Under these conditions, changes in k_p and γ_m can be considered in the context of the resulting change in b. The magnitudes of the components of $\Delta\vec{N}_{rT}$ for changes in these parameters (Table 1) are derived via analysis of Eqs. **4** and **8** (see *Methods* and *SI Text*). In general, modulation of $\langle p\rangle$ by changing translation efficiency affects the noise magnitude, whereas modulation of $\langle p\rangle$ by changing protein stability affects the dynamic responsiveness, compared with modulation of $\langle p\rangle$ by changes in transcription rate. These patterns are reflected in $\Delta\vec{N}_r$ and allow the inference of possible mechanisms of protein modulation and the magnitude of the parameter changes. These inferences allow certain regulatory mechanisms to be excluded and allow additional characterization to be focused on the most likely candidate models. The effects of simultaneous variation of more than one parameter can be predicted by simple addition of the independent vector components.

Table 1. Components of $\Delta\vec{N}_{rT}$ for transcription-translation model with fast regulatory kinetics and $b >> 1$

$\langle p\rangle$ modulated by	$\Delta\log CV_2$	$\Delta\log \tau_{1/2}$
$\Delta k_{m,\mathrm{eff}}$	0	0
Δb ($\gamma_m/\gamma_p >> 1$)	$\Delta\log(b)$	0
$\Delta\gamma_p$ ($\gamma_m/\gamma_p >> 1$)	0	$-\Delta\log(\gamma_p)$
Δb (γ_m/γ_p approaches 1)	$\Delta\log(b) - \chi_{m,\Delta_b}$	χ_{c,Δ_b}
$\Delta\gamma_p$ (γ_m/γ_p approaches 1)	$-\chi_{m,\Delta\gamma}$	$-\Delta\log(\gamma_p) + \chi_{c,\Delta\gamma}$

χ_{i,Δ_j} is a small correction to vector component i (m = magnitude, c = correlation) that has the same sign as the change in parameter j. χ_{i,Δ_j} increases with decreasing γ_m/γ_p.

For many of the proteins in our combined yeast database, the time scale of mRNA decay is of the same order as the cell-doubling time, in which case γ_m/γ_p could approach unity. Numerical analysis of Eqs. **4** and **8** revealed that relaxing the assumption $\gamma_m/\gamma_p \gg 1$ necessitates small corrections to the ideal vectors (Table 1, see *SI Text* and Fig. S1).

Slow Operator Kinetics. When the time scale of operator transitions approaches or exceeds the time scale of transcription and decay processes, operator dynamics add significant noise to the system (3, 11, 19). Long time scales in operator transitions can originate from many different underlying biological mechanisms, such as cell-cycle-dependent gene expression, low frequency noise in activators and repressors, or chromatin remodeling. These systems must be analyzed with care (see *SI Text* and Figs. S2 and S3).

Noise map coordinates for systems characterized by slow operator dynamics are calculated analytically under the assumption $\gamma_m/\gamma_p \gg 1$ (see Eq. **5**, *SI Text*). As expected, all of the noise vectors reside in the upper-right quadrant of the $\Delta\log(CV^2)-\Delta\log(\tau_{1/2})$ plane (Fig. 3), because slow operator dynamics adds noise and slows the dynamics of the system. The mean gene activity level $\overline{G(t)}$ is determined by $k_r/(k_r + k_f)$. Deviations from bias-line behavior are largest at moderate levels of gene activity and disappear as $\overline{G(t)}$ approaches 0 and 1, consistent with earlier work (3, 19). We now define two ratios that characterize the DNA-binding kinetics: $\kappa_1 = k_r/\gamma_p$ and $\kappa_2 = (k_r/\alpha_0)$. In general, κ_1 has a larger effect on the direction of the vectors, whereas κ_2 has a larger effect on the overall magnitude of the vectors. At $\kappa_1 \gg 1$, only the $\Delta\log(CV^2)$ component of the noise vector is significant, because $1/\gamma_p$ is the dominant time scale

BIOPHYSICS

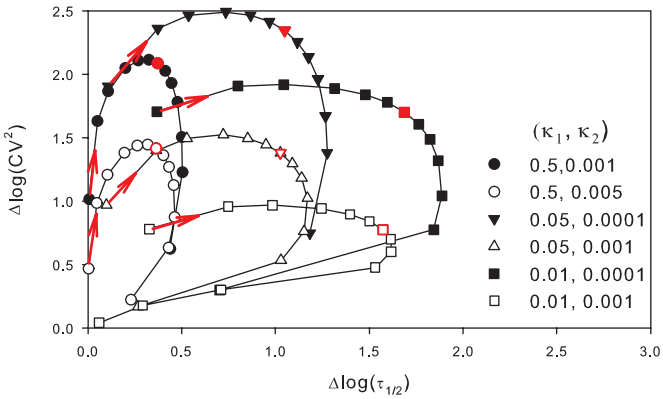

Fig. 3. Noise regulatory vectors for the gene circuit in Fig. 1*B* as affected by slow operator kinetics. Points indicate $\Delta\vec{N}_{rT}$ for $\overline{G(t)}$ values of (starting from and moving in the direction of the red arrow) 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 (red), 0.6, 0.7. 0.8, 0.9, 0.95, and 0.99. The effect of operator kinetics is captured in the ratios $\kappa_1 = k_r/\gamma_p$ and $\kappa_2 = k_r/\alpha_0$.

under both bias-line and experimental conditions. However, as κ_1 decreases below 1, the operator dynamics control the time scale of protein fluctuations, and the $\Delta\log(CV_2)$ component becomes increasingly significant. The overall magnitude of the deviation is negligible for $\kappa_2 \geq 1$ and increases as κ_2 decreases below 1. This can be most easily understood by realizing that small values of κ_2 are associated with small values of b ($\kappa_2 = \kappa_1\gamma_p/\alpha_0 = \kappa_1 b/\langle p\rangle_0$, where $\langle p\rangle_0$ is the mean protein level in the absence of regulation $= \alpha_0 b/\gamma_p$), which in turn results in a decrease in CV_A^2 relative to large κ_2 and hence an increase in $\Delta\log(CV^2)$. In other words, systems with large b are inherently noisy; therefore, the additional noise attributable to slow operator dynamics is smaller relative to the bias line circuit. The distortion of the bottom of some of the curves in Fig. 3 is an artifact related to the operational definition of $\tau_{1/2}$ (see *SI Text* and Fig. S3).

Autoregulation. So far, we have demonstrated that parameter deviations from the *a priori* model $m(S_A, k_A)$ are quantitatively captured in $\Delta\vec{N}_r$. We now use stochastic simulation (6, 12) of negatively and positively regulated gene circuits (see models in *SI Text* and Table S1) to quantify how these specific deviations of model structure S relative to the *a priori* model are reflected in $\Delta\vec{N}_r$. Negative autoregulation is a mechanism of maintaining homeostatic protein levels that is reported to decrease both CV^2 (20, 24) and $\tau_{1/2}$ (1, 18).

The noise behavior of negatively autoregulated gene circuits strongly depends on the speed of the operator kinetics, which we found to be conveniently characterized by kinetic ratios analogous to those defined in the previous section ($\kappa_1 = k_{ar}/\gamma_p$ and $\kappa_2 = k_{ar}/\alpha_0$, , where k_{ar} is the rate constant for dissociation of the autoregulator–DNA complex). Specification of κ_1 and κ_2 also fixes the ratio $b/\langle p\rangle_0$ and results in noise vectors that are independent of $\langle p\rangle_0$. Classical negative autoregulation behavior (a decrease in both CV^2 and $\tau_{1/2}$) occurs as both κ_1 and κ_2 approach unity, in which case the operator dynamics may be considered to be fast (Fig. 4*A*). For a given value of $\overline{G(t)}$, the regulation vectors are observed to rotate clockwise and grow larger in magnitude with decreasing κ_1 and κ_2 as the DNA-binding kinetics become slower (Fig. 4*A*). For the case of $\kappa_1 \geq 1$, the predominant effect of decreasing κ_2 below a value of one is to increase $\Delta\log(CV^2)$ (Fig. 4*A*). However, for $\kappa_1 < 1$, a decrease in κ_2 causes both $\Delta\log(CV^2)$ and $\Delta\log(\tau_{1/2})$ to increase, effectively increasing the magnitude of the vector (Fig. 4*A*). Therefore, when κ_1 and κ_2 are both <1, the former can be considered to primarily change the direction of the noise vector,

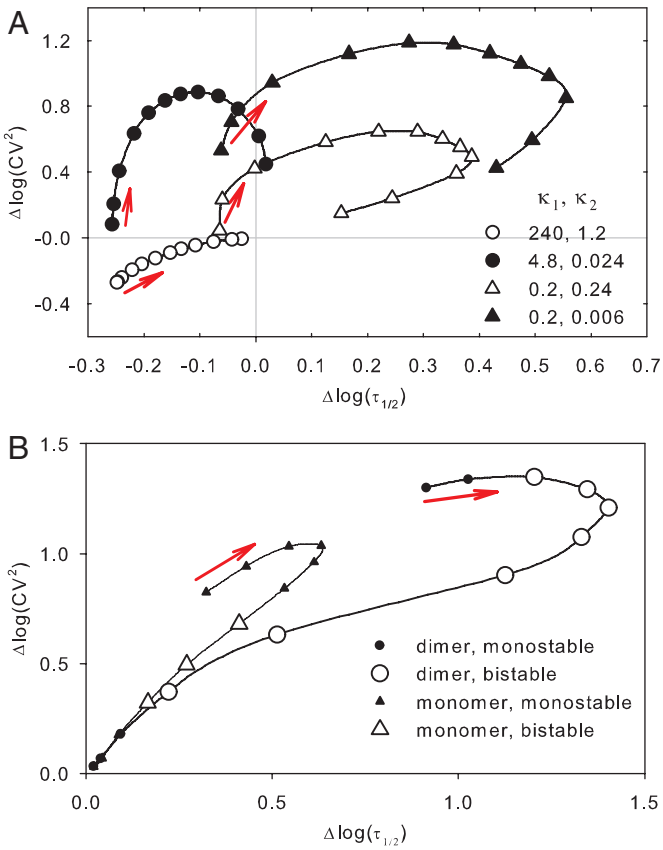

Fig. 4. Noise regulatory vectors for autoregulation. Points indicate $\Delta\vec{N}_{rT}$ at gene activation levels of (starting from and moving in the direction of the red arrow) 0.02, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7. 0.8, 0.9, and 0.95. (*A*) Negative autoregulation. The effect of decreasing κ_2 is to increase $\Delta\log(\tau_{1/2})$ when κ_1 1 and to increase both $\Delta\log(CV^2)$ and $\Delta\log(\tau_{1/2})$ when $\kappa_1 < 1$. (*B*) Positive autoregulation. The effect of multimerization and regimes of mono- and bistability.

whereas the latter primarily controls its magnitude, as was the case for the nonautoregulated vectors considered in the previous section. Overall, negative autoregulation may result in either an increase or a decrease in CV^2 and $\tau_{1/2}$, depending on whether the repression kinetics are fast or slow. This likely explains a few reports in the literature in which negative autoregulation appears to either have little effect on noise or even increases it (e.g., ref. 9).

Because the role of negative autoregulation is thought to be to

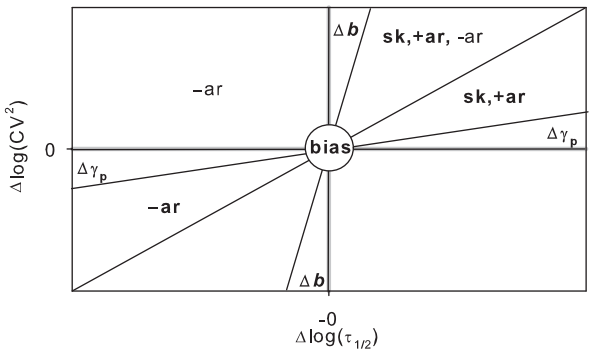

Fig. 5. Summary of noise vector domains for various regulatory motifs. (−ar, negative autoregulation; +ar, positive autoregulation; sk, slow gene activation kinetics). Bold font denotes domains of primary influence.

reduce the magnitude and correlation components of noise (1, 18, 20, 24), we sought to identify conditions in which these effects were enhanced. First, we recognized that lower values of b correspond to stronger effects of negative autoregulation at high levels of repression (see *SI Text* and Fig. S4). At high repression, transcription events are relatively rare; however, when b is high, a large number of proteins are translated from each transcription event. These proteins persist until they are removed by decay or dilution. The feedback system is no longer able to tightly regulate the protein level around $\langle p \rangle$, rather the instantaneous value of protein level varies widely from near zero to around b, with small values of $\langle p \rangle$ being achieved by shutting down transcription for long periods of time. The noise dynamics of the system then becomes controlled by protein decay/dilution rather than by feedback in strongly repressed systems with high b.

We also considered the case in which the autorepressor must first form a multimer before binding to the operator and the case of cooperative binding of multiple monomers to operator sites. Repression by multimers and cooperative binding were shown to have similar effects when the kinetics of complex formation were sufficiently rapid (*SI Text* and Fig. S5). Compared with regulation with the monomer, regulation by a dimer results in a small additional decrease in $\Delta\log(CV^2)$ and $\Delta\log(\tau_{1/2})$ when the operator kinetics are fast (Fig. S4).

A positively autoregulated gene circuit is characterized by a low basal state and a high induced state. The system is schematically identical to the negative autoregulation case except now G and G' are the basal (transcription rate α_0) and induced (transcription rate α_1) states of the gene, respectively. Depending on the parameter regime, the circuit may be monostable only in the high or the low state, or it may possess bistability in which both the high and low states are stable (10, 16, 25). Noise regulatory vectors for a wide range of parameters were generated by using stochastic simulation. Both the CV^2 and $\tau_{1/2}$ components of $\Delta\overrightarrow{N}_{rT}$ are positive over the entire parameter space, consistent with previously known characteristics of the system (16, 25). Unlike negative autoregulation, the noise regulatory vectors for positive autoregulation are strongly affected by multimerization (Fig. 4*B*). In the absence of other nonlinearities (e.g., population-dependent protein degradation), deterministic analysis predicts that autoregulation by a monomer is always monostable (25), whereas higher multimers can introduce bistability. Points characterized by bistability (as determined by histograms of the simulation results) are shown as open symbols. It is noteworthy that, in a stochastic context, autoregulation by the monomer was able to generate bistability. Stochastic bistability occurs when the positive-feedback effect is strong enough to sustain an induced state, but stochastically it is possible for the protein concentration to decay to zero. Nevertheless, the deterministically bistable dimer system is characterized by bistability over a much greater range of average induction levels $\overline{G'}$ (Fig. 4*B*). The sensitivity of $\Delta\overrightarrow{N}_{rT}$ to other parameter values in positively regulated gene circuits is provided in *SI Text* and Fig. S6.

Discussion

The noise regulatory vector $\Delta\overrightarrow{N}_r$ is a framework for interpreting measurements of gene noise to determine the fidelity of an assumed gene circuit model. Unlike conventional goodness-of-fit tests, $\Delta\overrightarrow{N}_r$ can suggest necessary modifications of the parameter values k and/or structure S of the assumed model to better represent the noise behavior of the circuit. These suggestions arise from comparison of experimentally measured $\Delta\overrightarrow{N}_r$ to a catalog of theoretical noise regulatory vectors $\mathbf{\Delta\overrightarrow{N}_{rT}} = g(\mathbf{m}(S, k))$ representing many different models. The "first edition" of such a catalog is given in Fig. 5, which summarizes the mapping of $\Delta\overrightarrow{N}_{rT}$ for the regulatory motifs considered in this work to various spatial domains in the $\log(\tau_{1/2})-\log(CV^2)$ plane. Additional motifs can be added as they are developed. Interestingly, there are some domains in which multiple motifs are operative, whereas others cannot be reached by any one of the motifs analyzed here, acting alone. Although the domains are not unique to a single regulatory motif, it allows additional experimental characterization to focus on the most likely candidates. Although the direction of $\Delta\overrightarrow{N}_r$ points to one or more motifs, its magnitude provides information that constrains the range of viable biokinetic parameters.

The noise vectors may be particularly useful when they are measured over a wide range of experimental conditions to assess specific contexts in which different regulatory motifs may be operative. In this case, the workflow in Fig. 1 may be used in an iterative manner in which the *a posteriori* gene circuit of one iteration becomes the *a priori* circuit for the next.

Noise regulatory vectors could be easily applied to a number of experimental studies in the recent literature. For example, recent studies have sought to build predictable gene circuits with complex feedback (7) or combinatorial promoters with multiple repressor-binding sites (13) based on understanding obtained by studying their simpler subcomponents. In this case, the *a priori* model would be based on the characterized components. A noise regulatory vector close to zero would result if the composed system performed as predicted. This is a particularly stringent test because the noise regulatory vector depends on both CV^2 and $\tau_{1/2}$. Deviations would result in noise vectors that, when compared with an appropriately constructed catalog of theoretical noise vectors, would point toward modifications of the model needed to capture system behavior.

In this work, we have considered noise vectors with CV^2 and $\tau_{1/2}$ components. One interesting extension of the method is the use of higher-dimensional noise vectors. Additional dimensions may be able to better capture nuances in the shape of the autocorrelation function that correlate strongly to particular regulatory motifs. It may also be possible to include cross-correlation components to the noise vector that indicate regulatory relationships between two genes within the same circuit.

One final interesting observation concerns the relative magnitude of b in eukaryotes (b up to 1,000 or more; ref. 2). and our compiled database) vs. prokaryotes [typically <100 (20)]. The noise effects (relative to the bias circuit) of slow operator kinetics and positive and negative autoregulation were shown to be amplified when b is small. It appears that negative autoregulation, a common regulatory that is beneficial in reducing the effects of noise, has coevolved with relatively small values of b in prokaryotes. Conversely, regulatory processes that increase noise such as chromatin remodeling, cell-cycle-dependent gene regulation, and positive autoregulation seem to have coevolved with more efficient translation systems (high b) in eukaryotes. In the latter case, the evolution of noise-tolerant cell function would perhaps accommodate both noisy regulatory processes and noisy constitutive gene expression associated with efficient translation.

Methods

Construction of 3D Noise Map. As an illustrative example, we used our assumed transcription-translation circuit (S_A) to interpret experimental observables (Ω) consisting of a compilation of several genome-scale databases characterizing the abundance and stability of mRNA and proteins in *S. cerevisiae*. We combined separate databases for mRNA abundance $\langle m \rangle$ (26), mRNA half-life ($t_{1/2,m}$) (27), protein abundances $\langle p \rangle$ (28), and protein half-life ($t_{1/2,p}$) (29) by matching data across ORFs. Because the measurements in the compiled database were made by different researchers at different times under different experimental conditions, caution should be used in their use. Our model of the intrinsic noise of protein synthesis is based on a system of first-order reactions (Fig. 1*B*; details provided in *SI Text* and Table S1). As part of our definition of $m(S_A, k_A)$, we assume that gene activation kinetics are fast ($k_f + k_r \gg \alpha_0, \gamma_m, \gamma_p$), and that $\alpha_1 = 0$. Under these conditions, the steady-state mean protein level $\langle p \rangle$ is given by:

BIOPHYSICS

$$\langle p \rangle = k_{m,eff} b/\gamma_p$$

$$k_{m,eff} = \alpha_0 + k_f(\alpha_1 - \alpha_0)/(k_f + k_r), \quad [6]$$

where $k_{m,\mathrm{eff}}$ is the effective transcription rate, and $b = k_p/\gamma_m$ is the average number of proteins synthesized per transcript. Values of CV^2 ($= \Phi_P(0)/\langle p \rangle^2$) and $\tau_{1/2}$ ($\Phi_P(\tau_{1/2}) = \Phi_P(0)/2$) were calculated from the autocorrelation function of protein level [$\Phi_p(\tau)$; see Eq. **8**] for each of the 2,920 ORFs in the combined database (Dataset S1).

Relationship of Parameters k_A to Observables Ω. The values of model parameters k_A are assumed to be related to Ω according to:

$$\gamma_m = \ln(2)/t_{1/2,m} + ln(2)/t_{\mathrm{double}}$$

$$\gamma_p = \ln(2)/t_{1/2,p} + \ln(2)/t_{\mathrm{double}} \quad [7]$$

$$k_{m,eff} = \alpha_0(1 - \overline{G'(t)}) = \langle m \rangle \gamma_m$$

$$b = \langle p \rangle \gamma_p / k_{m,eff},$$

under the assumptions $\alpha_1 = 0$ and $t_{\mathrm{double}} = 90$ min in yeast.

Calculation of Autocorrelation Functions. Under conditions in which gene activation kinetics are fast ($k_f + k_r \gg \alpha_0, \gamma_m, \gamma_p$), and extrinsic noise is negligible (2), the protein autocorrelation function $\Phi_P(\tau)$ for model structure S_A (shown in Fig. 1*B* and *SI Text* and Table S1) is given by:

$$\Phi_P(\tau) = \langle p \rangle \left\{ \left(1 + \frac{b}{1 - (\gamma_p/\gamma_m)^2}\right) \exp(-\gamma_p \tau) + \left(\frac{b(\gamma_m/\gamma_p)}{1 - (\gamma_m/\gamma_p)^2}\right) \exp(-\gamma_m \tau) \right\}. \quad [8]$$

The analytical expression for $\Phi_p(\tau)$ in the case where gene activation kinetics are slow, from which Eq. **8** is derived, is given by Eq. **5** in *SI Text*.

Stochastic Simulation and Construction of Noise Maps. Exact stochastic simulations of negative and positive autoregulation models were performed by using the Sorting Direct Method (12), an optimized version of Gillespie's original Stochastic Simulation Algorithm (6). Values of CV^2 and $\tau_{1/2}$ were determined from the autocorrelation function, which was calculated by Eq. **3**. Parameter values were swept to generate 3D noise maps. Noise regulatory vectors were constructed by cubic spline interpolation of the 3D maps, after smoothing with a running average function ($n = 5$).

ACKNOWLEDGMENTS. We thank Natalie Ostroff for thoughtful review of the manuscript. R.D.D., M.S.A., and M.L.S. acknowledge support from the Center for Nanophase Materials Sciences, which is sponsored by Basic Energy Sciences, Office of Science, U. S. Department of Energy. J.M.M. acknowledges the use of Miami University's 128 node Redhawk Computing Cluster, which was used for the stochastic simulations.

1. Austin DW, *et al.* (2006) Gene network shaping of inherent noise spectra. *Nature* 439:608–611.
2. Bar-Even A, *et al.* (2006) Noise in protein expression scales with natural protein abundance. *Nat Genet* 38:636–643.
3. Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* 422:633–637.
4. Cox CD, *et al.* (2006) Frequency domain analysis of noise in simple gene circuits. *Chaos* 16.
5. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
6. Gillespie DT (1977) Exact stochastic simulation of coupled chemical-reactions. *J Phys Chem* 81:2340–2361.
7. Guido NJ, *et al.* (2006) A bottom-up approach to gene regulation. *Nature* 439:856–860.
8. Hasty J, McMillen D, Isaacs F, Collins JJ (2001) Computational studies of gene regulatory networks: In numero molecular biology. *Nat Rev Genet* 2:268–279.
9. Hooshangi S, Weiss R (2006) The effect of negative feedback on noise propagation in transcriptional gene networks. *Chaos* 16.
10. Isaacs FJ, Hasty J, Cantor CR, Collins JJ (2003) Prediction and measurement of an autoregulatory genetic module. *Proc Natl Acad Sci USA* 100:7714–7719.
11. Kepler TB, Elston TC (2001) Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys J* 81:3116–3136.
12. McCollum JM, Peterson GD, Cox CD, Simpson ML, Samatova NF (2006) The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior. *Comput Biol Chem* 30:39–49.
13. Murphy KF, Balazsi G, Collins JJ (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci USA* 104:12726–12731.
14. Newman JRS, *et al.* (2006) Single-cell proteomic analysis of S-cerevisiae reveals the architecture of biological noise. *Nature* 441:840–846.
15. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307:1962–1965.
16. Scott M, Hwa T, Ingalls B (2007) Deterministic characterization of stochastic genetic circuits. *Proc Natl Acad Sci USA* 104:7402–7407.
17. Sigal A, *et al.* (2006) Variability and memory of protein levels in human cells. *Nature* 444:643–646.
18. Simpson ML, Cox CD, Sayler GS (2003) Frequency domain analysis of noise in autoregulated gene circuits. *Proc Natl Acad Sci USA* 100:4551–4556.
19. Simpson ML, Cox CD, Sayler GS (2004) Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. *J Theor Biol* 229:383–394.
20. Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* 98:8614–8619.
21. Volfson D, *et al.* (2006) Origins of extrinsic variability in eukaryotic gene expression. *Nature* 439:861–864.
22. Weinberger LS, Dar RD, Simpson ML (2008) Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat Genet* 40:466–470.
23. Megason SG, Fraser SE (2007) Imaging in systems biology. *Cell* 130:784–795.
24. Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. *Nature* 405:590–593.
25. Becskei A, Seraphin B, Serrano L (2001) Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion. *EMBO J* 20:2528–2535.
26. Arava Y, *et al.* (2003) Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 100:3889–3894.
27. Wang YL, *et al.* (2002) Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci USA* 99:5860–5865.
28. Ghaemmaghami S, *et al.* (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
29. Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci USA* 103:13004–13009.