NIH Public Access
Author Manuscript
IET Syst Biol. Author manuscript; available in PMC 2012 July 06.

Published in final edited form as:
IET Syst Biol. 2010 November ; 4(6): 379–392. doi:10.1049/iet-syb.2009.0070.

Defining Cooperativity in Gene Regulation Locally Through Intrinsic Noise

Mark Maienschein-Cline,
Department of Chemistry and James Franck Institute, The University of Chicago, Chicago, IL 60637

Aryeh Warmflash, and
Center for Studies in Physics and Biology, The Rockefeller University, New York, NY, USA

Aaron R. Dinner
Department of Chemistry and James Franck Institute, The University of Chicago, Chicago, IL 60637

Abstract

Regulatory networks in cells can comprise a variety of types of molecular interactions. The most basic are pairwise interactions, in which one species controls the behavior of another (for example, a transcription factor activates or represses a gene). Higher-order interactions, while more subtle, can be important for determining the function of networks. Here, we systematically expand a simple master equation model for a gene to derive an approach for robustly assessing the cooperativity (effective copy number) with which a transcription factor acts. The essential idea is that moments of a joint distribution of protein copy numbers determine the Hill coefficient of a cis-regulatory input function (CRIF) without nonlinear fitting. We show that this method prescribes a definition of cooperativity that is meaningful even in highly complex situations in which the regulation does not conform to a simple Hill function. To illustrate the utility of the method, we measure the cooperativity of the transcription factor CI in simulations of phage-λ and show how the cooperativity accurately reflects the behavior of the system. We numerically assess the effects of deviations from ideality, as well as possible sources of error. The relationship to other definitions of cooperativity and issues for experimentally realizing the procedure are discussed.

1. INTRODUCTION

The relationships between different species in a regulatory reaction network can assume several layers of complexity. At the most basic, discerning the linking of species by interactions yields an underlying network structure; determining the causal direction of each interaction further informs how the network functions. Such pairwise information can be determined statistically by Bayesian methods [1, 2] and related [3] analyses; other statistical methods have also been applied to this end [4]. However, more complex interactions between species can exist. For example, it may be that several copies of the species A must act together to promote the production of B. This cooperativity can have important implications for the behavior of a network at different concentrations of species A: at high cooperativity, there is a sharp threshold for production of B, whereas at low cooperativity the production of B scales more gradually with A. Thus, knowledge of the cooperativity is important for quantitatively modeling and, more generally, assessing a network's function; for example, determining whether a system is bistable [5–8]. Gene regulatory networks commonly involve cooperativity [9–13] and thus serve as good systems for investigating means of estimating it.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

There is no one definition of cooperativity, and in each case the measurement method used reflects the particular choice of definition. A common approach to measuring cooperativity assumes a Hill function form of regulation [14] and performs a non-linear fit to the curve of regulator versus product concentrations to obtain the Hill coefficient. This method requires a large exploration of the range of the regulator concentrations to tease out the full shape of the curve, which can be challenging for many systems. Additionally, complex systems may not exhibit Hill-like regulation, and thus fits based on this form can be misleading. Alternatively, Goldbeter and Koshland [15] define cooperativity from the ratio of regulator concentrations required to give 90% and 10% of the maximal transcription activity. Although their derivation begins with a Hill function for the form of theoretical regulation, the definition of cooperativity is applicable to other forms of regulation. However, this choice still requires exploration of a large range of regulator concentrations, as determination of the upper and lower bounds of regulation is required. In this sense, it is a global measure of cooperativity. It does not provide insight into the local degree of sensitivity of regulation.

The above analyses of gene regulatory networks focus on relationships between average expression rates and average concentrations of transcription factors. However, because proteins in a cell are often present in relatively low abundance, stochastic fluctuations in the copy numbers of molecules can have important effects on the dynamics of gene regulation [16–20] and the analysis of the noise produced by such fluctuations has been the subject of many recent advances in the analysis of regulatory networks [21–24]. Based on this fact, we recently suggested that higher-order correlation functions of copy numbers can provide model-free signatures of complex regulatory interactions. We specifically showed how to elucidate combinatorial contributions to regulatory interactions involving two or more transcription factors (whether they act sub- or super-additively) [25]. The approach is based on analytically derived relationships between the moments of the distribution from an expanded master equation [26] and the theoretical form of the combinatorial interaction, which is encoded in the *cis*-regulatory input function [27]. In this paper we extend this paradigm to define cooperativity in gene regulation. By systematically expanding a master equation [26], we identify linear relationships between functions of measurable moments of the steady-state distribution of protein copy numbers and the cooperativity manifested in gene responses to changes in transcription factor levels. We illustrate the method by applying it to simulation data of a number of different possible regulatory scenarios (e.g., oligomerization followed by DNA binding versus different copies of the transcription factor binding independently to the DNA). These examples suggest that our approach can be used to meaningfully quantify cooperativity even in complex situations.

Our definition, which again starts with a Hill function, is also suitable to a larger range of regulation mechanisms. However, because our method uses the slope of a best-fit line to determine cooperativity, the range of regulator concentration that needs to be sampled can be much smaller. This also allows for the determination of cooperativity locally, reflecting the natural behavior of the system. Unlike the previously mentioned methods, which compare only average values, ours uses the intrinsic noise of the system to obtain a covariance between regulator and product. While this makes our method more susceptible to experimental noise, the increasing availability of high-throughput methods capable of generating large amounts of data works to our advantage, as large data sets allow for better statistics but do not imply a larger range of available regulator concentrations (required by the other methods but not by ours). Comparisons of our method with those previously discussed are performed, and it becomes apparent that the methods are not simply different ways of measuring the same cooperativity, but actually reflect different definitions of the same phenomenon. Issues in realizing our approach experimentally are discussed.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2. METHODS

In this section we introduce a reaction scheme for cooperative gene expression and its associated *cis*-regulatory input function (CRIF). The CRIF relates the rate of expression to the transcription factor levels [28–30]. We then show how measurable moments of the protein copy number distribution can be used to elucidate the Hill coefficient (exponent) that quantifies the effective copy number with which the transcription factor acts. The notion of cooperativity is generalized in Section 3.3.

2.1. Reaction scheme for cooperative gene expression

Here, we consider a single transcription factor, A, which must oligomerize to form an n-mer to bind to the DNA and activate transcription of the gene for protein B:

$$\begin{aligned} \varnothing &\rightleftharpoons A \\ A+A &\rightleftharpoons A_2 \\ &\vdots \\ A+A_{n-1} &\rightleftharpoons A_n \\ A_n+D &\rightleftharpoons A_nD \\ A_nD &\rightarrow R+A_nD \\ R &\rightarrow \varnothing \\ R &\rightarrow B+R \\ B &\rightarrow \varnothing \end{aligned} \tag{1}$$

Above, D is the gene, and R is the mRNA for B; arrows from and to Ø represent production and degradation, respectively. In the following treatment of this network, we use rate constants g_X and k_X for the production/binding and degradation/unbinding reactions (X is an arbitrary species or bound complex), respectively. Additionally, we choose to write the rate constant for the reaction $R \rightarrow B + R$ as g_B/α, where α is an arbitrary standard copy number or concentration, such that g_B is a zeroth order rate constant (the same order as g_A).

Because binding and unbinding reactions are generally fast compared to transcription, translation, and degradation, we can assume a quasi-steady-state in all intermediate species [A_i ($i > 1$), A_nD, and R] [31] and model the transcription of B as a single (effective) event. We thus simplify the reaction network to

$$\varnothing \xrightarrow{g_A} A \xrightarrow{k_A} \varnothing \qquad \varnothing \xrightarrow{g_B f} B \xrightarrow{k_B} \varnothing \tag{2}$$

where $f \equiv f(N_A)$ is the CRIF (unitless) denoting the dependence of the rate of transcription of B on the amount of A. In the quasi-steady-state limit of the reaction scheme in Eq. 1, the CRIF takes the form

$$f(N_A) = \frac{\nu N_A^n}{\mathcal{K}_n + N_A^n}. \tag{3}$$

In this CRIF, the Hill coefficient n quantifies the cooperativity. N_A is the copy number of species A; $\mathcal{K}_n = K_{A_nD} \prod_{i=2}^{n} K_{A_i}$, and $K_X = k_X/g_X$; $\nu = N_{D,tot}/\alpha K_R$, where $N_{D,tot}$ is the total (bound and unbound) copy number of D.

2.2. Moments of the steady-state distrbution

Our goal is to relate the CRIF (f) and its derivatives to measurable moments of the steady-state distribution. To this end, we need to account for the fluctuations that give rise to the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

distribution. Returning to the simplified reaction scheme presented in Eq. 2, we can describe the dynamics by a master equation:

$$\frac{dP(N_A,N_B)}{dt}=\left(\left[g_A\Omega\left(\widehat{E}_A^{-1}-1\right)+g_B\Omega f(N_A)\left(\widehat{E}_B^{-1}-1\right)\right]+\left[k_A\left(\widehat{E}_A-1\right)N_A k_B\left(\widehat{E}_B-1\right)N_B\right]\right)P(N_A,N_B) \quad (4)$$

Here N_X is again the copy number of species X, and $\hat{E}_X$ is a step operator such that, for h an arbitrary function of N_X, $\hat{E}_X h(N_X) = h(N_X+1)$ and $\widehat{E}_X^{-1}(N_X)=h(N_X-1)$. In Eq. 4, the terms in the first set of square brackets on the right hand side represent production of species A and B, and the terms in the second set of square brackets represent their degradation.

The parameter Ω in Eq. 4 is introduced as an arbitrary extensive variable, used to quantify the size of the system. Its use reflects the choice of units in the rate constants, and we include it to facilitate the expansion of the master equation following van Kampen [26]. Fluctuations are expected to scale as $\Omega^{1/2}$ [26, 32, 33], motivating the change from discrete to continuous variables: $N_X = \Omega\phi_X + \Omega^{1/2}\xi_X$. Here, ϕ_X is the solution to the deterministic rate equations and ξ_X measures instantaneous deviations, which we take to be equivalent to the difference between the concentration of X in a single cell and its average over the population. The resulting master equation, now as a function of ξ_A and ξ_B, contains the CRIF as a function of the average concentration only [$f(\phi_A)$]. Relatedly, there is a factor of Ω difference between $f(\langle N_A\rangle)$ and $f(\phi_A)$, which reflects the units of the rate constants in the CRIF.

As discussed further below, for our purposes we need selected first and second moments of the joint steady-state distribution of N_A and N_B. We obtain these by expanding the step operator $\hat{E}$ and f in the master equation in powers of $\Omega^{-1/2}$, as detailed elsewhere [25, 26]. The first moments (means) are trivial:

$$\begin{aligned}\langle N_A/\Omega\rangle &= \phi_A\\ \langle N_B/\Omega\rangle &= \phi_B.\end{aligned} \quad (5)$$

From the deterministic rate equations (Eq. 2), we can further write $\phi_A = g_A/k_A$ and $\phi_B = g_B f(\phi_A)/k_B$ (yielding a relationship between ϕ_B and f). The variance of A and the covariance of A and B (multiplied by a factor of Ω) are

$$\begin{aligned}\Omega\left\langle(N_A/\Omega-\phi_A)^2\right\rangle &= \left\langle\xi_A^2\right\rangle=\tfrac{1}{2}\left(\tfrac{g_A}{k_A}+\phi_A\right)=\tfrac{g_A}{k_A}\\ \Omega\left\langle(N_A/\Omega-\phi_A)(N_B/\Omega-\phi_B)\right\rangle &= \langle\xi_A\xi_B\rangle=\tfrac{g_B}{k_A+k_B}\left\langle\xi_A^2\right\rangle\tfrac{\partial f}{\partial\phi_A}.\end{aligned} \quad (6)$$

Note that the second line of Eq. 6 is also the $j = 1$ case of Eq. 3 in our recent study on combinatorial regulation in gene networks [25]. Substituting the variance of A into the equation for the covariance, we obtain

$$\langle\xi_A\xi_B\rangle=\frac{g_A g_B}{k_A(k_A+k_B)}\left(\frac{\partial f}{\partial\phi_A}\right), \quad (7)$$

which yields a relationship between $\langle\xi_A\xi_B\rangle$ and $\partial f/\partial\phi_A$. The means and variance of A computed above are exact from the expanded master equation, but the covariance is the first term in a series in powers of $\Omega^{-1/2}$. For most experimental systems, this approximation is expected to be a less significant source of error than (1) the system being not well described by the quasi-steady-state approximation such that the form of the CRIF is incorrect, or (2) $\partial f/\partial\phi_A$ being close to 0, as in the saturation limit or at low levels of A in cooperative systems, so the covariance is difficult to estimate from available data.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2.3. Hill coefficient

Our goal is to identify a linear expression for the Hill coefficient n. To isolate the n dependence of the CRIF, we differentiate Eq. 3 (following the change of variables from N_A to ϕ_A) and divide the result by f^2. Taking the logarithm of the ratio:

$$\ln\left(\frac{\partial f/\partial \phi_A}{f^2}\right)=\ln\left(\frac{n\mathcal{K}_n}{v}\right)-(n+1)\ \ln \phi_A. \tag{8}$$

Substituting Eqs. 5 and 7, Eq. 8 becomes

$$\ln\left(\frac{\langle \xi_A \xi_B\rangle}{\phi_B^2}\right)=C-(n+1)\ \ln \phi_A, \tag{9}$$

where

$$C=\ln\left[\left(\frac{n\mathcal{K}_n}{v}\right)\frac{g_A k_B^2}{g_B k_A\,(k_A+k_B)}\right]. \tag{10}$$

Thus, by plotting $\ln\left(\langle \xi_A \xi_B\rangle/\phi_B^2\right)$ as a function of $\ln \phi_A$ for constant C, one can obtain n from the slope of the best fit line.

2.4. Extension to repression

It is relatively straightforward to derive an analogous expression for the case that A represses the transcription of B. Schematically, this change requires replacing the reaction for the production of R in Eq. 1 with

$$D \rightarrow D+R \tag{11}$$

All other reactions are the same. In this case, the CRIF (following the change of variables to ϕ_A) is

$$f(\phi_A)=\frac{v\mathcal{K}_n}{\mathcal{K}_n+\phi_A^n} \tag{12}$$

where v and $\mathcal{K}_n$ are the same as in Eq. 3. Proceeding as above,

$$\ln\left(-\frac{\partial f/\partial \phi_A}{f^2}\right)=\ln\left(\frac{n}{\mathcal{K}_n v}\right)+(n-1)\ \ln \phi_A \tag{13}$$

or

$$\ln\left(-\frac{\langle \xi_A \xi_B\rangle}{\phi_B^2}\right)=C+(n-1)\ \ln \phi_A. \tag{14}$$

The constant C is the same as in Eq. 10, but with a factor of $\mathcal{K}_n^{-1}$ instead of $\mathcal{K}_n$. The negative sign in the logarithm on the left hand side reflects the fact that $\partial f/\partial \phi_A$ is proportional to $\langle \xi_A \xi_B\rangle$, which is negative when A is a repressor.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2.5. Generating data

Eqs. 9 and 14 enable one to obtain the Hill coefficient n from the slope of the line that best fits $\ln\left(\langle \xi_A \xi_B \rangle / \phi_B^2\right)$ as a function of $\ln \phi_A$. A sufficient number of measurements of $\ln\left(\langle \xi_A \xi_B \rangle / \phi_B^2\right)$ as a function of $\ln \phi_A$ are thus needed. Acquiring points along this line requires modulating one or more of the rate constants, as altering only the initial concentrations does not affect the steady state. Specifically, we should vary g_A or k_A if we are viewing $\ln \phi_A$ as the independent variable. Importantly, these (and other changes) to the kinetic parameters must be made in such a way that the linear forms of Eqs. 9 and 14 are preserved despite the dependence of the intercept C on these parameters (Eq. 10).

As a thought experiment, consider scaling both g_A and g_B by a factor $g_A = g_A^{(0)} s$ and $g_B = g_B^{(0)} s$, where $g_X^{(0)}$ denotes the original value of g_X. Since $\phi_A = g_A / k_A$, we have $\phi_A \propto s$. This choice leaves C independent of s because the numerator and denominator of Eq. 10 are scaled to the same extent. Similar results are obtained by scaling both k_A and k_B by s (in which case $\phi_A \propto 1/s$). From a more practical perspective, consider only scaling g_A. Again $\phi_A \propto s$, but in this case the intercept in Eq. 9 is proportional to s as well. Under these circumstances, the right hand side of this equation can be rewritten as

$$\begin{aligned} \ln\left[\left(\frac{n\mathcal{K}_n}{v}\right)\frac{g_A k_B^2}{g_B k_A (k_A + k_B)}\right] - (n+1)\ \ln \phi_A \quad &= \ln C_1 s - (n+1)\ \ln C_2 s \\ &= C_3 - n \ln s \end{aligned} \tag{15}$$

where the C_i ($i = 1, 2, 3$) are again independent of s. In other words, we still expect a linear relationship in a log-log plot, but scaling g_A alone increases the slope by 1. Thus one can vary only g_A, but care is needed in interpreting the slope. Other possible scalings do not result in linear log-log plots. Because we ultimately plot two observables (resulting from averages and covariances) against each other, the precise value of the scaling factor s does not need to be known in experiment. Note that because the only difference between the activator and repressor constants C (in Eqs. 9 and 14) is the placement of $\mathcal{K}_n$ (in the numerator or denominator), the above discussion of scaling methods applies equally to each case. Potential experimental realizations of the various scalings are considered in Section 5.

3. NUMERICAL EXAMPLES

3.1. Systems with Hill-function regulation

To illustrate Eqs. 9 and 14, which make the quasi-steady state assumption, we used the Gillespie algorithm [34] to simulate the idealized reaction network in Eq. 1, as well as its analog for a repressor, for $n = 1$, 2, 3, and 4. The rate constants (prior to scaling) are given in Table I. We scaled g_A by s, for $s = 0.5$, 0.75, 1, 1.25, 1.5. Simulations were run for 10^6 time units, printing a frame once every time unit. Plots of these simulations are shown in Fig. 1.

Statistical error was estimated using a bootstrap method [35]: 2000 replications of the numerical data were generated by random sampling, with replacement, of the 10^6 frames, omitting the first 2000 frames to ensure sampling from the steady-state distribution. Averages and subsequent best-fit slopes were recomputed for each replication, and a distribution of slopes was obtained. Ranges reported are for the two-tailed, 95% confidence interval of the distribution. This procedure tends to underestimate the errors because it assumes that the sampled distribution is a, good representation of the real distribution. Not surprisingly, higher moments converge more slowly than lower ones, such that the errors are primarily in the vertical dimension in Fig. 1. Slopes and confidence intervals are shown in

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table. II; for completeness we also include results from simultaneously scaling either g_A and g_B or k_A and k_B.

As expected, when A is an activator (repressor) and scaled by g_A the best fit slope is close to $-n$ (n). These results can be compared with those obtained by nonlinearly fitting the average expression of B as a function of ϕ_A (Fig. 2 and Table III). As an example, we use the simulation data in Fig. 1A; fits were performed using the nonlinear least-squares Levenberg-Marquardt algorithm [36]. The variables optimized are n, v and $\mathcal{K}_n$. We see that the results are sensitive to the initial guesses for their values. Substituting the correct values for these parameters results in a very good fit, but the fitting routine cannot reliably identify these values starting from arbitrary ones. Specifically, for $n = 1$, the fit routine reliably returns the same answer as the best-fit slope, but, as n increases, the ability of the nonlinear fitting routine to pick out the right value of n worsens. Exhaustive exploration of integer values of n is meaningful in this simple example, but not in the more complex cases considered in the following sections.

In summary, our procedure yields much smaller uncertainties in n than does directly fitting the CRIF. Moreover, our procedure requires exploration of fewer values and a smaller range of ϕ_A, which could be experimentally advantageous since there are likely to be limits to how much a species can be scaled in practice. For a best-fit line, one only needs two points in principal, so long as there is a sufficient change in the independent variable that the slope of the line is distinguishable from experimental or statistical noise. By contrast, non-linear fitting would require sampling of a much broader range of the independent variable in order to even observe the general shape of the curve.

3.2. Measurement uncertainty

We explore the effects of two sources of experimental uncertainty on measured cooperativities separately: the presence of noise in data collection, and the number of data points collected. These tests were performed on the scenario considered in Fig. 1A, in which the action of an activator is probed by scaling its production. To mimic experimental noise, uncorrelated Gaussian noise was added to the copy numbers of species A and B, with the standard deviations of the noise set as percentages of the steady-state averages. For example, a standard deviation of 5% would indicate that a species with an average copy number of 135 would have noise added with standard deviation 6.75 (and mean 0). To understand how the error scaled with the size of the data set, L points ($L < 10^6$, the total length) were randomly sampled from the simulations (after the first 2000 points were removed); in this case, no Gaussian noise was added. Bootstrap estimations of the 95% confidence interval were calculated as discussed above for each of the possible sources of error, and the magnitude of the interval ranges are plotted in Fig. 3. As expected, the statistical error increases with the magnitude of the noise and decreases with the extent of sampling (it scales roughly as $1/\sqrt{L}$). Naturally, noisier measurements and smaller data sets compound each other, but there is no systematic deviation of the average.

3.3. Extension to independent binding

3.3.1. A generalized, single-species CRIF—The reaction scheme presented in Eq. 1 is one possible molecular mechanism leading to cooperativity. An alternative is that n copies of A each bind individually to the regulatory region of gene D, such that the intermediate species are A_iD, $1 \le i \le n$, instead of A_i, $1 < i \le n$, and A_nD. In the case that copies of A bind individually, it is also necessary to specify how different occupancies influence transcription. For example, transcription could occur only when all n copies are present (corresponding to an AND logic operation), or when any one transcription factor is bound (an OR logic operation). Intuitively, the AND-like case may seem more like an actual

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

cooperativity; however, both cases (and ones in between) can be considered within the same framework. As we show, none of the idealized cases exactly manifests cooperativity equal to the maximum number of copies of A that can bind, but we detail under which circumstances the effective cooperativity approaches this maximum. Indeed, our analysis shows that the manifested cooperativity is a truer reflection of the actual behavior of the system than the maximum possible number of bound copies.

We can write the CRIF for an arbitrary single-transcription factor situation as:

$$f(\phi_A) = \frac{\sum_{i \in S} v_i \phi_A^i / \mathcal{K}_i}{\sum_{j \in T} \phi_A^j / \mathcal{K}_j}. \quad (16)$$

Here, S is the set of transcriptionally active gene-regulator complexes, and T is the set of all gene-regulator complexes, whether active or not. Both of these sets can contain the unbound gene as well (i or $j = 0$ in the notation above). By definition, $S \subset T$. For each term in both sums, the power of ϕ_A represents the number of copies of A that have bound to the gene and v_i is the corresponding transcription rate in the numerator. For this general case, the cooperativity can no longer be described by a single integer exponent. Nevertheless, we can use our analysis to quantify cooperativity and obtain an effective Hill coefficient, n_{eff}.

The constant $\mathcal{K}_i$ is an extension of the definition given in Section 2: $\mathcal{K}_i = \prod_{j \in M_i} K_{X_j}$, where M_i is the set of all intermediate species X_j leading up to the ith transcription factor bound to the gene and K_{X_j} is again the ratio of degradation to production rate constants for X_j (we define $\mathcal{K}_0 = 1$). Essentially, $\mathcal{K}_i$ is a measure of the probability of forming the gene-regulator species with i copies bound. Eq. 16 can be extended to an arbitrary number of regulators by summing over all combinations and powers of each transcription factor.

3.3.2. Idealized independent binding cases—We now consider two idealized cases explicitly: an activator that can act strictly in either an AND-like or an OR-like manner. Clearly, the mode of regulation in a real system may be more complex and may not be known, but these examples suffice for illustration. To obtain a CRIF for these cases, we define S and T appropriately in Eq. 16. In all instances, because monomers of A bind to the gene one at a time, we have a total of $n + 1$ gene-transcription factor species in the system (D, AD, . . ., A_nD). Thus $T = \{0, 1, \ldots, n\}$. For the AND-like activator, the only transcriptionally active species is A_nD, so $S = \{n\}$. For the OR-like activator, all species are active except for the unbound gene, so $S = \{1, 2, \ldots, n\}$. In each case, the CRIF now contains sums in the denominator and, for the OR-like activator, in the numerator. To continue our analytical treatment, we assume that the binding rate for each copy of A is unaffected by the number already bound, so $K_{A_iD} = K_{A_jD}$ for any i and j, and $\mathcal{K}_i = \prod_{j=1}^{i} K_{A_jD} = K_{AD}^i$. This allows us to write each CRIF in closed form by geometric series substitution and calculate $\ln(\partial f/\partial \phi_A / f^2)$ to obtain an equation similar to Eq. 8. It di ers in that the constant $\ln(n\mathcal{K}_n/v)$ is replaced by a ϕ_A-dependent term. This term is complicated but has a sigmoid form with respect to the ratio ϕ_A/K_{AD}: it monotonically increases from 0 to $n - 1$ for the AND-like activator and decreases from $n - 1$ to 0 for the OR-like activator.

Thus, if we consider ϕ_A fixed but K_{AD} variable, then in the limit of high K_{AD} (corresponding to weak binding to the gene) this sigmoid expression goes to zero for the AND-like activator, the log-log plot in the analog equation becomes more linear, and the effective cooperativity goes to the maximal number of binding sites. In the limit of strong binding to the gene (small K_{AD}), the sigmoid expression goes to $n - 1$ and the effective

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

cooperativity goes to 1 regardless of the number of binding sites. Since the trend of the sigmoid expression is opposite for the OR-like activator, the limits are opposite as well: effective cooperativity is maximal at strong binding and minimal at weak binding. Fig. 4 shows these trends for 1, 2, 3, and 4 binding sites.

To understand these limits physically, consider the case of the AND-like activator with 4 binding sites. In the limit of strong binding, because there are many more transcription factors in the system than available binding sites, most of the time all 4 sites will be filled and transcription activated. Occasionally one transcription factor will dissociate from the gene, and transcription will be deactivated. However, at this point it is much more likely that the fourth site will be refilled than that another transcription factor will dissociate, so the transition between active and inactive is decided by the fluctuation in a single transcription factor. Thus, although there are four binding sites, the effective cooperativity is 1, as our method reflects. Conversely, in the limit of weak binding, most of the time no binding sites are filled, and transcription is inactive. In order for the transition for active to occur, four transcription factors must bind, so the effective cooperativity is equal to the number of binding sites, again as our method reflects. Similar reasoning can explain the trends in the OR-like activator, as well as repressor cases.

The key element of this section is that the effective cooperativity may depend both on the mode of regulation and the stability of the transcriptionally active gene-regulator species relative to the inactive ones. Generally, a highly stable transcription factor-gene complex will be relatively unaffected by fluctuations or scalings in the regulator, and effective cooperativity will decrease. This fact can be used with other information to gain an additional understanding of regulator dynamics. For example, if the mode of regulation is known by some other means, a measurement of effective cooperativity will give the researcher an idea of how strongly the regulator binds to the gene. Despite the fact that the actual mode of regulation and the stabilities of these regulator-gene species will not be known in most experiments, our method will return a meaningful effective cooperativity that reflects the behavior of the system.

3.3.3. Comparison to an alternative measure of cooperativity—It is illuminating to compare the effective cooperativities determined from the independent binding cases to an alternative definition of cooperativity. Goldbeter and Koshland [15] define a response sensitivity parameter R_s based on the ratio of regulator concentrations near the upper and lower limits of activity and relate it to the Hill coefficient to obtain a relatively model-free measure of cooperativity n_s. While this method does not rely on non-linear fitting, it requires a complete exploration of the range of ϕ_A to determine the limits of activity. Many experimental systems may be difficult to scale to this degree, but for our idealized systems it is easy to compute these parameters analytically. Unlike n_{eff}, n_s is invariant to K_{AD} because scaling K_{AD} up or down does not change the limits of the CRIF; such scaling is equivalent to changing the units of ϕ_A. Of course, for n_{eff} if we restrict attention to a limited (absolute) domain of ϕ_A then the cooperativity can appear to change with K_{AD} (Fig. 4).

If we compute n_s for the AND-like and OR-like activator independent binding systems discussed above, we obtain the same values for both systems: (n, n_s) = (2, 1.36), (3, 1.56), (4, 1.69). Additionally, computation of n_s for higher values of n in the independent binding networks reveals that it is bounded above by 2. This restriction on the range of n_s can be intuited from comparisons of plots of the CRIFs for the oligomerization network with those for the independent binding networks (Fig. 5). As n increases, the oligomerization CRIF sharpens both before and after $\phi_A = K_{AD} = 100$ (Fig. 5A), whereas the independent binding CRIFs only sharpen before (AND-like, Fig. 5B) or after (OR-like, Fig. 5C) this point. In the independent binding case, each additional binding site has a diminishing effect on the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

cooperativity, measured by n_s. This decrease is also the same for the idealized AND-like and OR-like cases, resulting in equivalent cooperativities. Mathematically, the limit of 2 is a consequence of the specific definition of R_s and can be seen by calculating it analytically for the independent binding CRIFs (results not shown). Although this discussion has focused on a comparison of our definition of cooperativity to that proposed by Goldbeter and Koshland, it should be apparent that non-linear fitting to a Hill function is an inappropriate representation of the form of regulation for the independent binding cases, as the CRIFs in Fig. 5B and C are quite different from those in Fig. 5A.

4. APPLICATION TO A REALISTIC SYSTEM: PHAGE-λ

To test our definition of cooperativity on a non-idealized system, we consider the CI protein in bacteriophage-λ. CI is responsible for stabilizing the lysogenic phase of phage-λ by repressing lytic gene expression (notably Cro) [37]; this repression is caused by binding of CI dimers to the first or second sites of the O_R operator (O_R1 or O_R2), preventing activation of the P_R promoter by blocking RNA polymerase. This binding event is stabilized by the binding of other CI dimers to adjacent O_R sites (forming a CI tetramer) and by long-range DNA looping between the O_R and O_L operators, facilitated by a CI octamer [37, 38]. The result is a set of highly cooperative interactions between CI dimers for repression of P_R. CI also autoregulates, activating its transcription through P_{RM} by binding to O_R2 [39] or repressing its transcription by binding to O_R3 [37] (Fig. 6).

To simulate this system, we adapt a reaction network model recently introduced by Morelli and co-workers [5] for phage-λ which includes DNA looping and transcription from both P_{RM} and P_R (full model in their supplementary information). In this study, the authors examined the effect of DNA looping and non-specific DNA binding on the robustness of the bistability of phage-λ, which involves transitions between a CI-dominated (lysogenic) state and a Cro-dominated (lytic) state. Because we are interested in measuring the cooperativity of CI, and because the Cro and non-specific DNA binding reactions are numerically expensive, we simulate the lysogenic state without Cro or non-specific DNA binding. However, to measure the effect of CI regulation we add a non-interacting reporter protein B that is transcribed with the kinetics of Cro, as might be accomplished by addition of a plasmid. We simulate the system with and without loop formation (in simulations without looping, only O_R is simulated) to examine the effect of DNA looping on cooperativity. As Morelli and co-workers note, the strength and dynamics of the loop formation in vivo are not known; their study examines a range of ΔG values for looping, and we choose an intermediate value for ours (–1.5 kcal/mol).

We choose to scale the degradation rate of the mRNA for CI (this is analogous to scaling the rate constant for formation of CI, i.e. g_A in the above examples). The value used for this rate constant by Morelli and co-workers is 0.005776, and we scale it from 0.003 to 0.0095 in increments of 0.0005. Simulations for this system were run for 10^7 seconds, printing a frame every 10 s; the first 5000 frames are omitted from the averages, since the system takes longer to converge to steady-state than the idealized systems. Due to the stability of the lysogenic phase, the most frequent copy number of B is zero as instances where B is transcribed are relatively rare, and the average copy number of B is very low (order of 10^{-3} to 10^{-1}). Thus the covariance of CI and B is also quite small and slow to converge, and it is not uncommon for simulations run for a finite amount of time to have a positive covariance, which makes the logarithm undefined. In our analysis, these simulations were removed from the best-fit line procedure, and the line was fit using the remaining data points. However, the scaling procedure described for phage-λ creates 14 scalings (and corresponding data points in the log-log plot), rather than the 5 used in the idealized simulations, to lessen the effect of the elimination of one or two data points. Since CI is a repressor of B and we are scaling by

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

g_A alone, following the discussion in Section 2.5 we expect a positive slope that gives n_{eff} (rather than $n_{eff} - 1$). Because different experimental methods might be able to measure only the monomer CI, or only the dimer CI_2, or only the sum of the two, we consider the cooperativity of CI, CI_2, and their sum (redefining ϕ_A in our analysis accordingly).

The results of these simulations are given in Table IV. CI in the non-looping system has an effective cooperativity of 5.62. This is close to the maximal cooperativity, given the mechanism of repression, as anywhere from 1 to 3 dimers (2 to 6 CI monomers) can act to repress transcription of B. Although it only takes one CI dimer to repress transcription of B, this effect is stabilized by the binding of another dimer to an adjacent site on O_R, forming a tetramer; this embodies an AND-like regulation. Because there are two ways for tetramers to form on O_R (binding to O_R1 and O_R2, or O_R2 and O_R3) and binding to either O_R1 or O_R2 represses P_R; thus there is an OR-like feature to the regulation as well.

One might expect that the cooperativity would be larger when looping is added based on the fact that then up to twelve copies of CI can be bound to the DNA. However, we find that the cooperativity is the same or lower within our confidence interval (Table IV). In this case, the explanation of the lower effective cooperativity is directly related to the discussion in Section 3.3.2: the high stability of the loop means that loop formation is relatively unresponsive to fluctuations or scaling in CI copy number. Because any loop structure represses transcription of B, the effective cooperativity of CI is thus decreased. This illustrates how our definition of cooperativity reflects the functional behavior of the system, rather than the underlying structural properties.

The trend in the CI dimer cooperativities from non-looping to looping follow that of the monomer, but are less than those for the monomers (2.12 and 1.68 less for the non-looping and looping systems, respectively). This reflects the fact that the dimers are already a cooperative interaction, so the remaining cooperativity to regulate transcription is less. The cooperativity of the total free CI (sum of monomer and twice the dimer copy numbers) is intermediate to the monomer and dimer cooperativities, but closer to the dimer value because the average copy number of dimer is several times high than that for the monomer, and fluctuations are larger and more long-lived in the dimer.

5. DISCUSSION

To understand the dynamics of a regulatory network, it is important to elucidate both the overall structure of the network and the details of individual regulation events. From the former we can ascertain the function of the network by determining what processes are regulated by different species, and how those species in turn are regulated. From the latter, we learn the specific mode of regulation in a single interaction, which can have important consequences for the dynamics of the network. Cooperativity has been demonstrated to be important in gene regulation [10, 13, 37, 40, 41] and signaling [6–8, 15]; this nonlinearity can lead to bistability, among other consequences. If many regulators are required to work in unison to activate or repress a reaction, the threshold between very low activity and very high activity can be quite sharp, and this activity will propagate downstream through the network. For systems with high cooperativity, if the regulator concentration is near the threshold then small changes in concentration can greatly affect the larger function of the network, but if the regulator concentration is far from the threshold then the network will be relatively robust to changes in regulator concentration [8, 18, 20, 24]. Thus cooperativity can have an important impact on the effect of noise in system dynamics.

Often, cooperativity is estimated by assuming a Hill form of regulation (as in Eq. 3) and nonlinearly fitting the average expression to determine the exponent (Hill coefficient) [9,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

10]. Such an approach is sensitive to the initial guesses for the parameters and oversimplifies the regulation in many cases. For example, as seen in Section 3.3, the *cis*-regulatory input function can have multiple terms in both the numerator and the denominator. It is not straightforward to non-linearly fit data in this more general case. Other methods that rely more implicitly on the form of regulation, such as R_s from Goldbeter and Koshland [15], still require a full exploration of the range of regulator concentrations. In contrast, we have introduced a means of determining transcription factor cooperativity from a best-fit line. This approach is relatively robust and does not require generating as many ϕ_A values, which is likely to be of significant experimental advantage.

The analysis clarifies the distinction between effective cooperativity (from the slope that is measured at an intermediate value of K_{AD}, for example) and maximal cooperativity (the largest exponent in the denominator of the CRIF). For any system, the measured cooperativity n_{eff} is a reflection of the number of transcription factors being used simultaneously in the system and typically will not be equal to n (the largest denominator exponent) that appears in the CRIF. The phage-λ example characterizes this difference particularly well: although DNA-looping necessarily involves more CI acting together, the robustness of the loop means that the effective cooperativity (reflecting the transition of a number of free CI to bound CI) decreases. The R_s parameter (which allows computation of a cooperativity n_s) discussed previously is also a general measure of the response sensitivity to changes in ϕ_A. However, R_s is a global, rather than local, measure, as it relies on the distance between the upper and lower limits of transcription and may ignore the effects of steep regions of the CRIF. By contrast, due to the dependence of our n_{eff} on K_{AD} for independent binding systems, a potentially interesting experiment is also possible: in the event that the ratio K_{AD} is experimentally adjustable, a change in slope at different values could be used to infer the presence of an independent binding, as opposed to an oligomerization, mechanism.

Our analysis requires measurement of a steady-state distribution of protein copy numbers. Deviations from the average behavior across a population can occur either due to the stochastic nature of gene transcription, as individual molecular events occur discretely at random times (intrinsic noise), or from variations in the global cellular state (extrinsic noise) [16]. The latter could reflect differences in cell size, copy numbers of RNA polymerases and ribosomes, or cell cycle stages, as well as variations in unobserved upstream regulators. Because our purpose is to extract information from moments of the distribution that arise from intrinsic noise, the ability to detect and control for extrinsic noise is important. For example, variations in the number of RNA polymerases affect all protein concentrations globally, which results in correlated fluctuations in copy numbers of species that are not linked mechanistically. It has been shown that such external effects can systematically shift the average behavior differently depending on the specific mode of regulation [24], and thus extrinsic noise must be dealt with carefully.

The most direct means of estimating the relative contributions of intrinsic and extrinsic noise is to distinguish and compare the outputs from two (or more) copies of a gene within a single cell [16, 42]. Some specific factors contributing to extrinsic noise, such as variations in cell size, can also be quantified easily; others can be inferred from amounts of molecular species that are unrelated to the network of interest [25, 43]. Interestingly, it has been suggested recently that intrinsic and extrinsic noise can be distinguished by analysis of the time-dependence of either the divergence of daughter cells from the same mother cell [44] or of the cross-correlation between a regulatory protein and its target [43]. In the latter, the essential idea is that extrinsic noise affects all species simultaneously, so there is no lag, whereas intrinsic noise requires some time to filter from a fluctuation in a regulator to a fluctuation in its target [43]. Using such time-dependent techniques, it may be possible to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

separate extrinsic from intrinsic noise, or at least assess the relative magnitudes of each. The degree to which intrinsic and extrinsic factors contribute to noise in gene expression remains open to debate [42, 44], and it is likely to be system specific.

If one can detect and separate extrinsic and intrinsic noise, one can sort the data into bins according to the relative levels of extrinsic noise in each cell and then compute the statistics of interest separately for each bin. Indeed, the scaling procedure detailed in Section 2.5 is essentially a version of controlled extrinsic noise: an external perturbation is introduced that shifts the steady-state average. Of course, because there is extrinsic noise in a real system one would have to ensure that external scaling factors did not significantly couple to existing extrinsic noise; this would be most important to consider if the levels of upstream regulators were significantly changed by the scaling. This consideration raises the possibility that one could avoid perturbation altogether by exploiting naturally arising bistability (or, more generally, multistability) within a system. To the extent that there was no significant interconversion between subpopulations of cells expressing different amounts of the regulator of interest, and the systematic changes in the output arose only from the changes in the regulator, the needed moments for fitting a line could be calculated by treating each of the the differentiated subpopulations as a separate ensemble of cells.

In the absence of such a situation, it is necessary to scale the production or degradation of participating species artificially. How might such perturbations be implemented experimentally? The production of the mRNA of a specific species can be controlled by an inducible promoter or, in some cases by varying the copy number of the gene itself (e.g. by varying the copy number of a plasmid); mRNA levels can also be modulated by direct injection, addition of specific sequences affecting the stability, or by knockdown (e.g., by RNA interference). The degradation rate of a species could be altered by introducing a molecule that denatures, degrades, or otherwise regulates the protein, for example a targeted ubiquitin-like protein [45, 46]. Other opportunities for scaling could exist for specific systems. It is worth noting that any method used to determine cooperativity (including non-linear fitting) would require some scaling technique to relate changes of ϕ_A to changes in ϕ_B, and would be subject to the effects of extrinsic noise. The practical issues discussed are not unique to our method.

The present study extends the emerging paradigm that statistically analyzing the propagation of intrinsic noise through a network can yield useful information about regulatory structures [21, 23, 25, 43]. In the simplest example, a stochastic upward fluctuation in the copy number of an activating transcription factor increases the probability of transcription, and a positive correlation is expected between the transcription factor and its target gene product. The relations derived in this paper exploit combinations of additional moments (correlation functions) for more complex inferences that can be hard to elucidate by other means. While here we have focused on the steady-state, the analytical methods used to derive Eq. 9 can be used to treat time-dependent systems as well. We believe that this would be a fruitful direction for further study, and it would aid in the development of new approaches for probing the dynamics of molecular systems [47].

Acknowledgments

We thank Lawrence Uricchio, S. M. Ali Tabei and Shannon Stewman for helpful discussions. This work was supported by the National Science Foundation and the Department of Energy Computational Science Graduate Fellowship program.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

References

1. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. J Comput Biol. 2000; 7:601. [PubMed: 11108481]
2. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. Science. 2005; 308:523. [PubMed: 15845847]
3. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, Califano A. ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006; 7(Suppl 1):S7. [PubMed: 16723010]
4. Wall, ME.; Rechsteiner, A.; Rocha, LM. *A Practical Approach to Microarray Data Analysis*, chapter 5: Singular value decomposition and principal component analysis. Kluwer; Norwell, MA: 2003. p. 91-109.
5. Morelli MJ, ten Wolde PR, Allen RJ. DNA looping provides stability and robustness to the bacteriophage λ switch. Proc Natl Acad Sci USA. 2009; 106:8101. [PubMed: 19416825]
6. Thattai M, van Oudenaarden A. Attenuation of noise in ultrasensitive signaling cascades. Biophys J. 2002; 82:2943. [PubMed: 12023217]
7. Gomez-Uribe C, Verghese GC, Mirny LA. Operating regimes of signaling cycles: Statics, dynamics, and noise filtering. PLoS Comp Biol. 2007; 3:2487.
8. Shibata T, Fujimoto K. Noisy signal amplification in ultrasensitive signal transduction. Proc Natl Acad Sci USA. 2005; 102:331. [PubMed: 15625116]
9. Hochschild A, Ptashne M. Cooperative binding of lambda-repressors to sites separated by integral turns of the DNA helix. Cell. 1986; 44:681. [PubMed: 3948245]
10. Werner M, Zhu L, Aurell E. Cooperative action in eukaryotic gene regulation: Physical properties of a viral example. Phys Rev E. 2007; 76:061909.
11. Stein B, Baldwin AS. Distinct mechanisms for regulation of the interleukin-8 gene involve synergism and cooperativity between C/EBP and NF-kappa-B. Mol Cell Biol. 1993; 13:7191. [PubMed: 8413306]
12. Lipshtat A, Loinger A, Balaban NQ, Biham O. Genetic toggle switch without cooperative binding. Phys Rev Lett. 2006; 96:188101. [PubMed: 16712399]
13. Mirny, Leonid A. Nucleosome-mediated cooperativity between transcription factors. 2009.
14. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach. J R Statist Soc B. 2007; 69:741.
15. Goldbeter A, Koshland DE Jr. An amplified sensitivity arising from covalent modification in biological systems. Proc Natl Acad Sci USA. 1981; 78:6840. [PubMed: 6947258]
16. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002; 297:1183. [PubMed: 12183631]
17. Vilar JM, Kueh HY, Barkai N, Leibler S. Mechanisms of noise-resistance in genetic oscillators. Proc Natl Acad Sci USA. 2002; 99:5988. [PubMed: 11972055]
18. Pedraza J, van Oudenaarden A. Noise propagation in gene networks. Science. 2005; 307:1965. [PubMed: 15790857]
19. Paulsson J, Berg OG, Ehrenberg M. Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation. Proc Natl Acad Sci USA. 2000; 97:7148. [PubMed: 10852944]
20. Paulsson J. Summing up the noise in gene networks. Nature. 2004; 427:415. [PubMed: 14749823]
21. Munsky B, Trinh B, Khammash M. Listening to the noise: random fluctuations reveal gene network parameters. Mol Syst Biol. 5(318):2009.
22. Walczak AM, Mugler A, Wiggins CH. A stochastic spectral analysis of transcriptional regulatory cascades. Proc Natl Acad Sci USA. 2009; 106:6529. [PubMed: 19351901]
23. Cox CD, McCollum JM, Allen MS, Dar RD, Simpson ML. Using noise to probe and characterize gene circuits. Proc Natl Acad Sci USA. 2008; 105:10809. [PubMed: 18669661]
24. Gerstung M, Timmer J, Fleck C. Noisy signaling through promoter logic gates. Phys Rev E. 2009; 79:011923.
25. Warmflash A, Dinner AR. Signatures of combinatorial regulation in intrinsic biological noise. Proc Natl Acad Sci USA. 2008; 105:17262. [PubMed: 18981421]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

26. van Kampen, NG. *Stochastic Processes in Physics and Chemistry*, chapter 10: The Expansion of the Master Equation.. North-Holland: 1992.
27. Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. 2007; 8:450. [PubMed: 17510665]
28. Setty Y, Mayo AE, Surette MG, Alon U. Detailed map of a cis-regulatory input function. Proc Natl Acad Sci USA. 2003; 100:7702. [PubMed: 12805558]
29. Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U. Plasticity of the cis-regulatory input function of a gene. PLoS Biol. 2006; 4:555.
30. Kuhlman T, Zhang Z, Saier MH, Hwa T. Combinatorial transcriptional control of the lactose operon of escherichia coli. Proc Natl Acad Sci USA. 2007; 104:6043. [PubMed: 17376875]
31. Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ. A bottom-up approach to gene regulation. Nature. 2006; 439:856. [PubMed: 16482159]
32. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N. Noise in protein expression scales with natural protein abundance. Nat Genet. 2006; 38:636. [PubMed: 16715097]
33. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci USA. 2001; 98:8614. [PubMed: 11438714]
34. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977; 81:2340.
35. Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. Chapman & Hall/CRC; 1998.
36. Marquardt DW. An algorithm for the least-squares estimation of nonlinear parameters. J Soc Indust Appl Math. 1963; 11:431.
37. Anderson LM, Yang H. DNA looping can enhance lysogenic CI transcription in phage lambda. Proc Natl Acad Sci USA. 2008; 105:5827. [PubMed: 18391225]
38. Revet B, von Wilcken-Bergmann B, Bessert H, Barker A, Muller-Hill B. Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. Curr Biol. 1999; 9:344.
39. Meyer BJ, Ptashne M. Gene-regulation at the right operator (OR) of bacteriophage-lambda. 3. lambda-repressor directly activates gene-transcription. J Mol Biol. 1980; 139:195. [PubMed: 6447796]
40. Dodd IB, Perkins AJ, Tsemitsidis D, Egan JB. Octamerization of lambda CI repressor is needed for effective repression of P-RM and efficient switching from lysogeny. Gene Dev. 2001; 15:3013. [PubMed: 11711436]
41. Hermsen R, Tans S, ten Wolde PR. Transcriptional regulation by competing transcription factor modules. PLoS Comp Biol. 2006; 2:1552.
42. Hu J, Sealfon SC, Hayot F, Jayaprakash C, Kumar M, Pendleton AC, Ganee A, Fernandez-Sesma A, Moran TM, Wetmur JG. Chromosome-specific and noisy IFNB1 transcription in individual virus-infected human primary dendritic cells. Nuc Acids Res. 2007; 35:5232.
43. Dunlop MJ, Cox RS III, Levine JH, Murray RM, Elowitz MB. Regulatory activity revealed by dynamics correlations in gene expression noise. Nature Genetics. 2008; 40:1493. [PubMed: 19029898]
44. Rausenberger J, Kollmann M. Quantifying origins of cell-to-cell variations in gene expression. Biophys J. 2008; 95:4523. [PubMed: 18689455]
45. Welchman RL, Gordon C, Mayer RJ. Ubiquitin and ubiquitin-like proteins as multifunctional signals. Nature Rev Mol Cell Biol. 2005; 6:599. [PubMed: 16064136]
46. Grabbe C, Dikic I. Functional roles of ubiquitin-like domain (ULD) and ubiquitin-binding domain (UBD) containing proteins. Chem Rev. 2009; 109:1481. [PubMed: 19253967]
47. Li Y, Qu XH, Ma A, Smith GJ, Scherer NF, Dinner AR. Models of single-molecule experiments with periodic perturbations reveal hidden dynamics in RNA folding. J Phys Chem B. 2009; 113:7579. [PubMed: 19415919]

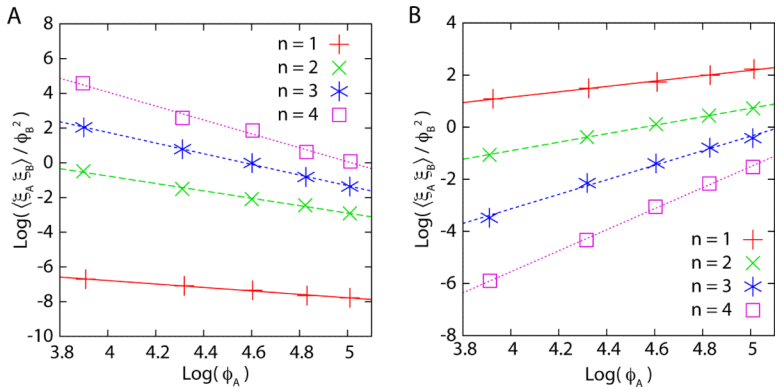

Figure 1.

Comparison of theory and simulation for cooperativity arising from oligomerization. Plots of (A) Eqs. 9 and (B) 14 for an activator and a repressor, respectively, each scaled by g_A for $n = 1$, 2, 3, and 4 with best fit lines. Averages were computed after removing the first 2000 points from the trajectory, to ensure that sampling was performed on the steady-state distribution. Averages are in units of copy number (i.e. $\phi_A = \langle N_A \rangle$) rather than concentration as there is no explicit volume.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

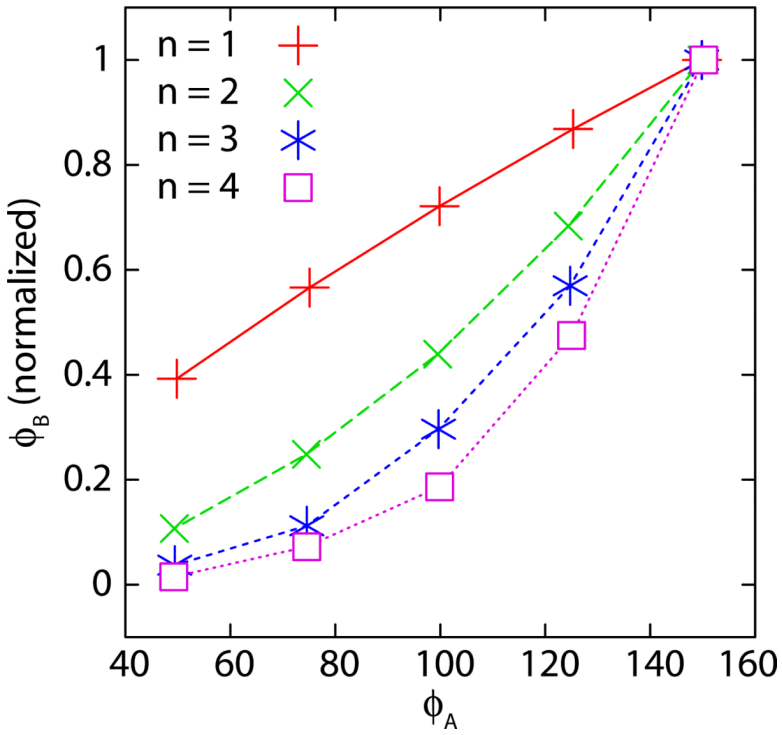

Figure 2.

Conventional nonlinear fitting. Plot of ϕ_B versus ϕ_A for $n = 1$, 2, 3, and 4, for the simulations of an activator scaled by g_A only (so that changes in ϕ_B are solely due to changes in ϕ_A); recall $\phi_B \propto f(\phi_A)$. Values of ϕ_B for a given n were normalized by the largest copy number observed in any of the five simulations; averages for ϕ_A are in units of copy number.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

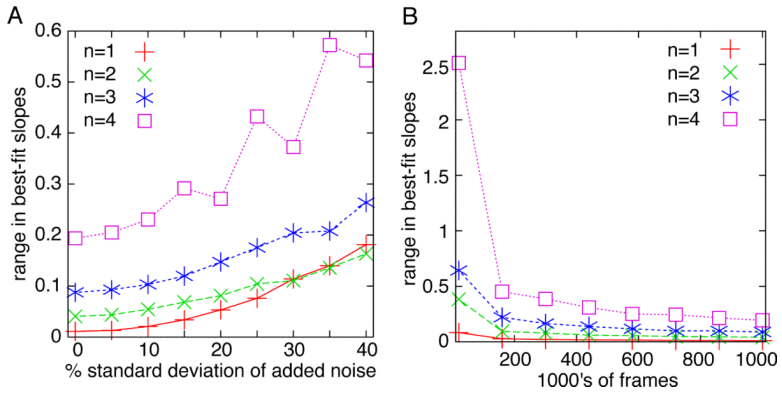

Figure 3.

Estimation of the effect of (A) measurement noise and (B) number of data points on the determination of cooperativity from the confidence interval of best-fit slopes. In each case, the first 2000 frames were removed to ensure that the simulation was around its steady-state. See text for details.

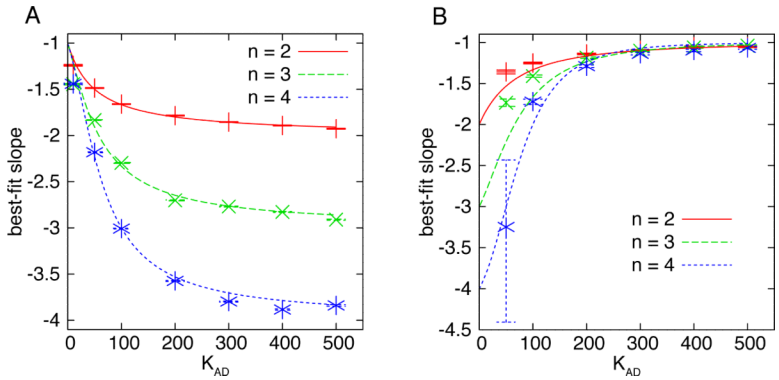

Figure 4.

Comparison of theory and simulation for cooperativity arising from independent binding. Two modes of regulation by A are shown: (A) AND-like and (B) OR-like activator. Solid lines are analytical calculations using the CRIF and points are numerical averages. In each case, rate constants were the same as those in Table I and simulations were performed as discussed previously with the following exceptions: the g_{AD} value listed was used for all binding events (and the k_{AD}/g_{AD} ratio was scaled for each point), and g_R was divided by the number of species contributing to R (1 for AND-activator, n for the OR-activator); the rate constants for these reactions were equivalent. Scaling in all cases was by g_A, and error bars reflect a 95% confidence interval from the bootstrap.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

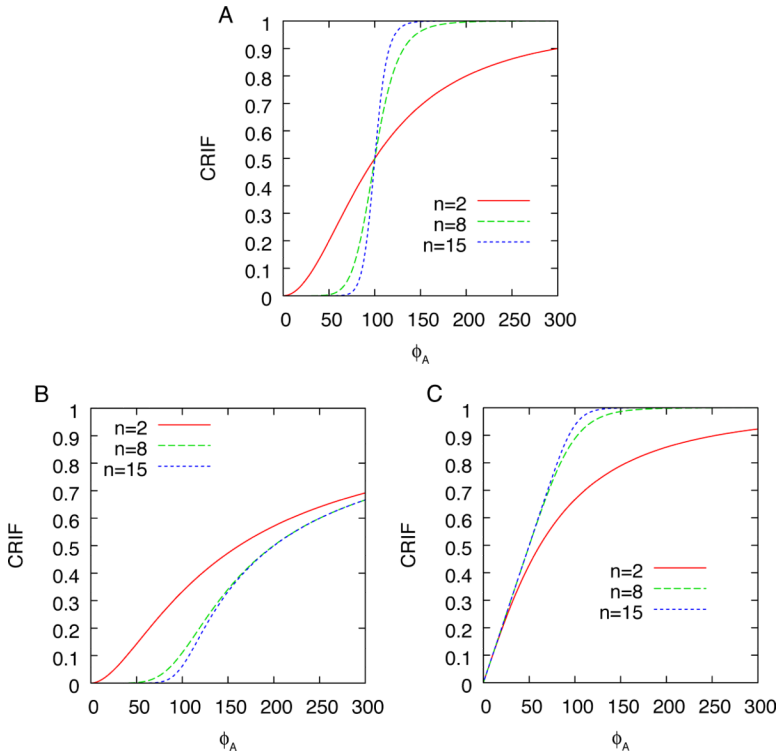

Figure 5.

Plots of the CRIFs for the oligomerization network (A), and independent binding networks with AND-like activators (B) and OR-like activators (C); each has $v = 1$ and $K_{AD} = 100$ ($\mathcal{K}_n = 100^n$ for the oligomerization network). The repressor CRIFs are not shown, as $f_{repressor} = 1 - f_{activator}$, so conclusions for the activator hold for the repressor. CRIFs are plotted for $n = 2$, 8, and 15 to exaggerate the trends in line shape with increasing n. The oligomerization case sharpens both before and after 100 with increasing n, but the independent binding networks only sharpen on one side.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

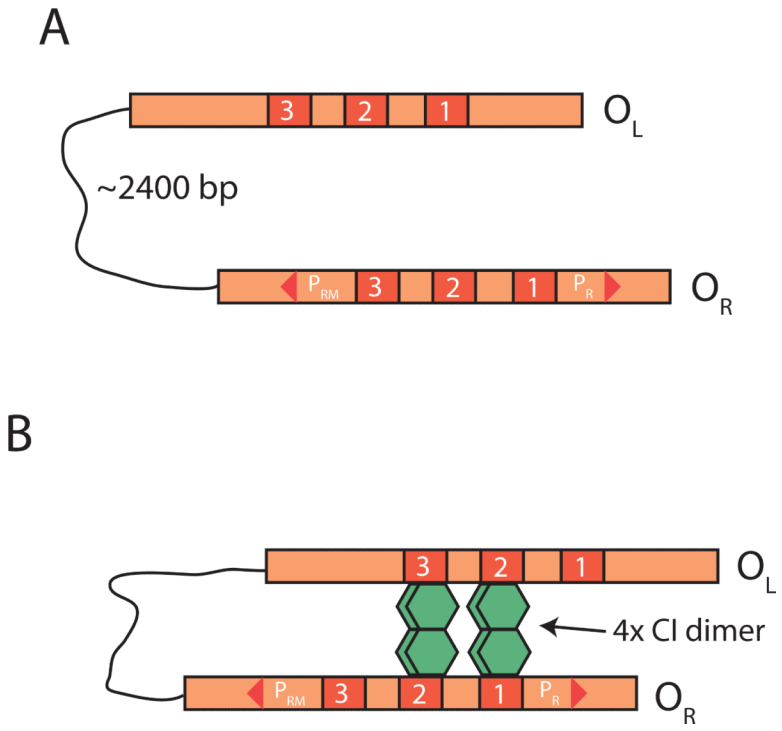

Figure 6.

Schematic of the phage-λ operators O_R and O_L and promoters P_R and P_{RM} (A), and one possible loop formation between the operators, facilitated by a CI octamer (B). Both O_R and O_L have 3 binding sites for CI dimers; CI is transcribed from the P_{RM} promoter, and the reporter protein B from the P_R promoter.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE I

Reaction rate constants for the idealized gene regulatory networks. Parameters were chosen such that fluctuations in the copy number of intermediates were small, ensuring close adherence to the steady-state approximation. In each case the total amount of gene species present ($N_{D,dot}$) was 4.

Rate Constant	Activator	Repressor
g_A	1	1
g_B	1	1
k_A	0.01	0.01
k_B	0.01	1
g_{A2}	0.0005	0.01
g_{A_i} ($i > 2$)	0.01	0.01
k_{A_i}	5	5
g_{A_nD}	0.01	2
k_{A_nD}	5	1
g_R	10	10
k_R	1	1

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE II

Best-fit slopes for the plots in Fig. 1. Numbers in parentheses represent 95% confidence intervals from the bootstrap method. Analytical calculations from Eq. 9 give slopes of $-n$ and $-(n+1)$ for an activator and n and $n-1$ for a repressor.

	Activator, scaled by		
n	g_A	g_A and g_B	k_A and k_B
1	–0.99 (–1.00, –0.98)	–1.98 (–1.98, –1.97)	–2.04 (–2.05, –2.03)
2	–2.13 (–2.15, –2.11)	–3.03 (–3.05, –3.01)	–3.19 (–2.30, –3.17)
3	–3.07 (–3.11, –3.03)	–4.04 (–4.08, –4.00)	–4.20 (–4.23, –4.17)
4	–4.01 (–4.07, –3.92)	–5.22 (–5.25, –5.06)	–4.95 (–5.04, –4.87)

	Repressor, scaled by		
n	g_A	g_A and g_B	k_A and k_B
1	1.03 (0.98, 1.08)	–0.02 (–0.08, 0.03)	0.00 (–0.06, 0.05)
2	1.63 (1.59, 1.67)	0.65 (0.62, 0.68)	0.69 (0.66, 0.73)
3	2.79 (2.77, 2.81)	1.78 (1.76, 1.80)	1.90 (1.88, 1.92)
4	4.04 (4.01, 4.07)	2.97 (2.94, 3.00)	3.23 (3.22, 3.25)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE III

Hill coefficients obtained from fitting $f(\phi_A) = v\phi_A^n / \left(K_n + \phi_A^n\right)$ to the plots in Fig. 2 (unnormalized). Initial guess sets were $guess_n = (v_0, n_0, K_{n,0}) = (1000, n, 100^n)$, chosen as such because K_n should scale roughly on the order of ϕ_A^n. Parameter values were determined by least squares fitting and the standard errors of the asymptotic values are given (only values for n are shown).

initial guess set, fit	1	2	3	4
$guess_1$, n =	1.00 ± 0.01	1.49 ± 2.02	1.34 ± 5.41	1.20 ± 8.13
$guess_2$, n =	1.00 ± 0.01	1.57 ± 1.94	1.57 ± 4.31	1.56 ± 5.93
$guess_3$, n =	1.00 ± 0.01	2.01 ± 0.05	2.39 ± 2.04	2.39 ± 3.83
$guess_4$, n =	1.00 ± 0.01	2.01 ± 0.05	3.15 ± 0.20	3.24 ± 2.31

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE IV

Cooperativities of CI and CI_2 and their sum (computed as CI + $2CI_2$, as each dimer represents two monomers) on the transcription of B for phage-λ, for non-looping and looping systems. Numbers in parenthesis represent 95% confidence intervals from the bootstrap method. In regards to the large confidence intervals, it should be noted that simulations performed with higher transcription rates of B had better statistics but similar trends: decreased effective cooperativity in the looping system (data not shown).

System type	CI	CI_2	CI + $2CI_2$
Non-looping	5.62 (2.45, 6.57)	3.50 (1.81, 3.74)	3.66 (1.95, 3.89)
Looping	4.32 (2.64, 6.67)	2.64 (2.43, 2.85)	2.81 (2.59, 3.03)