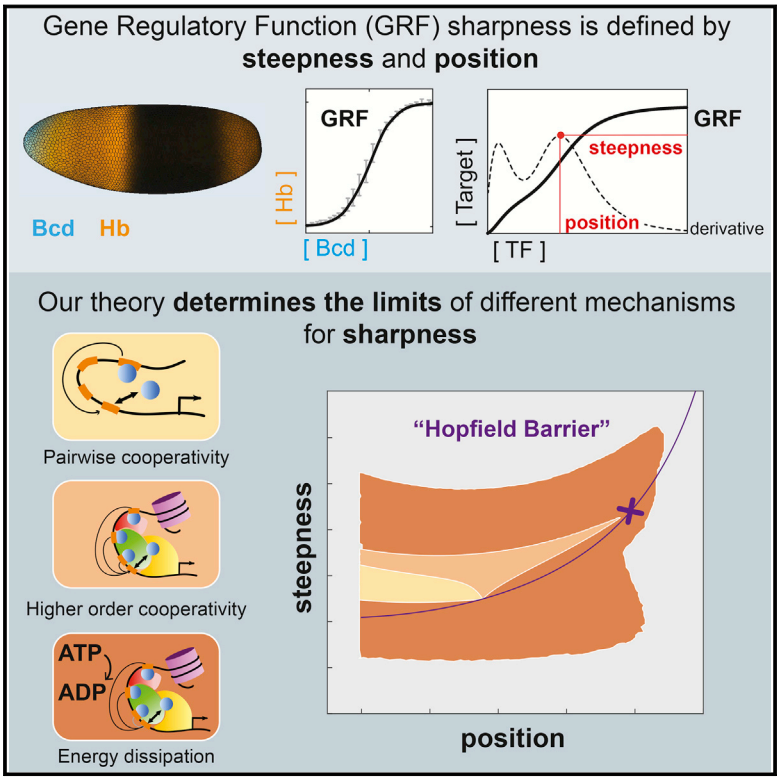
Cell

Information Integration and Energy Expenditure in Gene Regulation

Graphical Abstract



Highlights

- Gene regulation is understood quantitatively in terms of a bacterial paradigm
- This paradigm cannot account for the sharpness of gene expression in development
- Information integration or energy expenditure can explain the sharpness
- Hill functions form a "Hopfield barrier" for sharpness at thermodynamic equilibrium

Authors

Javier Estrada, Felix Wong, Angela DePace, Jeremy Gunawardena

Correspondence

jeremy@hms.harvard.edu

In Brief

The quantitative principles governing gene regulation in bacteria cannot explain the sharpness of gene expression in eukaryotes; energy use and information integration have to be taken into account as well.

Estrada et al., 2016, Cell *166*, 234–244
June 30, 2016
http://dx.doi.org/10.1016/j.cell.2016.06.012

CellPress

Cell

Information Integration and Energy Expenditure in Gene Regulation

Javier Estrada,[1] Felix Wong,[1,2] Angela DePace,[1] and Jeremy Gunawardena[1,*]
[1]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[2]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
*Correspondence: jeremy@hms.harvard.edu
http://dx.doi.org/10.1016/j.cell.2016.06.012

SUMMARY

The quantitative concepts used to reason about gene regulation largely derive from bacterial studies. We show that this bacterial paradigm cannot explain the sharp expression of a canonical developmental gene in response to a regulating transcription factor (TF). In the absence of energy expenditure, with regulatory DNA at thermodynamic equilibrium, information integration across multiple TF binding sites can generate the required sharpness, but with strong constraints on the resultant "higher-order cooperativities." Even with such integration, there is a "Hopfield barrier" to sharpness; for *n* TF binding sites, this barrier is represented by the Hill function with the Hill coefficient *n*. If, however, energy is expended to maintain regulatory DNA away from thermodynamic equilibrium, as in kinetic proofreading, this barrier can be breached and greater sharpness achieved. Our approach is grounded in fundamental physics, leads to testable experimental predictions, and suggests how a quantitative paradigm for eukaryotic gene regulation can be formulated.

INTRODUCTION

The molecular machinery which transcribes DNA into RNA is general purpose. Deciding which gene to transcribe requires regulatory DNA sequence information, which is interpreted by sequence-specific, DNA-binding transcription factors (TFs). Quantitative measurements of TF-DNA and TF-TF interactions in bacteria (Ptashne, 2004), together with analysis of the underlying physics (Ackers et al., 1982), have introduced fundamental quantitative concepts like "affinity" and "cooperativity" to explain the regulated recruitment of RNA polymerase to a gene. This bacterial paradigm has been widely used to interpret experimental results even outside the bacterial domain. However, eukaryotic transcription differs considerably from bacterial transcription; as a result, this raises the question of whether the bacterial paradigm is sufficient to explain how eukaryotic genes are regulated.

Bacterial TF sequence motifs have an average length of 16 base pairs, and those in eukaryotes are only half as long (Wunderlich and Mirny, 2009), suggesting that eukaryotes depend on combinatorial integration of many small packets of information. Such information integration might be implemented through nucleosomes or by multi-protein co-regulators, such as Mediator or CBP/p300, that make multiple contacts between TFs and the transcriptional machinery (Spitz and Furlong, 2012). Also, while bacterial gene regulation appears not to require energy from donors like ATP, making it reasonable to assume that it takes place at thermodynamic equilibrium, eukaryotic gene regulation depends on energy expenditure to reorganize chromatin, displace nucleosomes, post-translationally modify regulatory proteins, and methylate DNA. This qualitative appreciation of eukaryotic complexity has been difficult to translate into rigorous, well-defined concepts and new kinds of experiments that can explain the role of these molecular mechanisms in gene regulation.

Quantitative models grounded in physics could fill this critical gap. The physics-based "thermodynamic formalism" developed for bacteria assumes that regulation takes place at thermodynamic equilibrium. This formalism has been codified (Bintu et al., 2005) and applied to gene regulation in *Drosophila*, yeast, and human cells (Segal and Widom, 2009; Sherman and Cohen, 2012). However, the molecular complexity found in eukaryotes, especially the complexity that implements the information integration and energy expenditure described above, has not been incorporated into these models.

Questions about the sufficiency of the bacterial paradigm have been accumulating (Coulon et al., 2013), but the absence of a compelling example and the lack of appropriate concepts make it easy to fall back on what is familiar. Here, we present a compelling example of insufficiency and introduce appropriate quantitative concepts, rigorously based on the underlying physics, with which to reason about eukaryotic gene regulation.

We bring together three ingredients that exemplify a general approach to the problem. First, we focus on a property of gene regulation that can be described quantitatively; second, we identify a biological system in which that property has been measured; and, third, we exploit a mathematical framework that allows us to analyze both equilibrium and non-equilibrium systems.

The quantitative property on which we focus is the sharpness of gene expression in response to a TF, or the extent to which a small change in TF concentration can lead to a larger change in gene expression. Sharpness has been investigated in several biological systems, but is particularly evident in developmental patterning. The zygotic gap gene *hunchback* (*hb*) is expressed in an anterior region of the early *Drosophila* embryo under

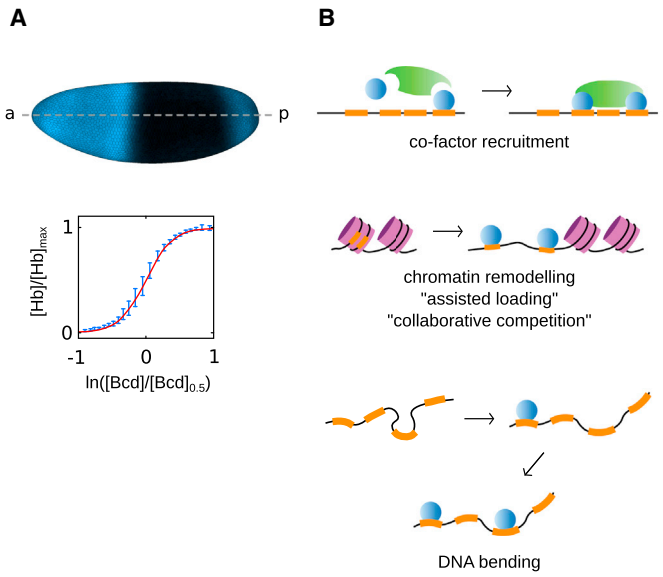


Figure 1. Sharpness in Development and Cooperativity Mechanisms

(A) Top: *Drosophila* embryo stained for Hb expression. Bottom: plot adapted from Figure 4A of Gregor et al. (2007) showing mean ± SE of Hb and Bcd from several embryos (blue) and a fit to the Hill function $\mathcal{H}_5$ (red).

(B) Examples of indirect, long-distance cooperativity, adapted from Figure 1 of Spitz and Furlong (2012).

(C) Top: pairwise cooperativity between two sites. Bottom: higher-order cooperativity of order three.

regulation by the maternal morphogen Bicoid (Bcd) (Figure 1A). The expression levels of Hb and Bcd proteins are related to each other in a way closely approximated by a simple algebraic expression (Gregor et al., 2007):

$$\frac{[\mathrm{Hb}]}{[\mathrm{Hb}]_{max}} \approx \frac{x^5}{1+x^5}. \quad (1)$$

Here, the concentration of Hb, denoted [Hb], is normalized to its maximal level, and x denotes Bcd concentration, normalized to the value at which half-maximal Hb expression is reached, so that $x = [\mathrm{Bcd}]/[\mathrm{Bcd}]_{0.5}$. Equation 1 describes the *hb* gene regulation function, which quantitatively expresses how the output of *hb* depends on [Bcd].

The expression in Equation 1 is a Hill function, $\mathcal{H}_a(x) = x^a/(1+x^a)$, for which the Hill coefficient, a, has the value $a = 5$. Increasing Hill coefficients imply increasing sharpness. In Equation 1, the sharpness represented by $a = 5$ reflects the precision with which individual nuclei use Bcd to determine their position along the anterior-posterior axis and create the tight boundary between Hb "on" and Hb "off."

Gregor et al. (2007) explain how the sharpness in Equation 1 arises by saying that it is "consistent with the idea that Hb transcription is activated by cooperative binding of effectively five Bcd molecules." This reflects the conventional bacterial paradigm, in which sharpness is accounted for at thermodynamic equilibrium by pairwise cooperativity between TFs, whereby TF binding at one site influences the affinity of TF binding at another site (Ptashne, 2004). With n binding sites and pairwise cooperativity, it is widely believed, as Gregor et al. (2007) suggest, that sharpness corresponding to a Hill coefficient of n can be achieved, without requiring any expenditure of energy.

To examine this idea, we use a recently introduced mathematical framework that generalizes the thermodynamic formalism to accommodate mechanisms that expend energy (Ahsendorf et al., 2014). We show that for regulatory DNA at thermodynamic equilibrium with only pairwise cooperativity, the experimentally measured sharpness described in Equation 1 cannot be biochemically realized, no matter how many TF binding sites are present. The widely held belief that the bacterial paradigm can be extrapolated in this way is not rigorously justified. We believe this is a compelling example of its insufficiency.

Information integration through nucleosomes or co-regulators could yield indirect, long-distance forms of cooperativity (Figure 1B), which could link multiple TF binding sites. To account for this, we introduce the concept of "higher-order cooperativity" at thermodynamic equilibrium (Figure 1C). If such cooperativities are present, greater levels of sharpness become possible. With n binding sites and higher-order cooperativities, a Hill coefficient of n or more still remains out of reach, but a Hill coefficient less than n can be achieved. Furthermore, not just the Hill coefficient but also the overall shape of the gene regulation function (GRF) can match what is found experimentally: with enough binding sites, GRFs can be found that are statistically indistinguishable in shape from the Hill functions in Equation 1. However, these GRFs lie on the edge of what can be biochemically achieved and impose stringent quantitative constraints on the mechanisms responsible for higher-order cooperativity.

Higher-order cooperativities improve sharpness but reveal fundamental barriers to what can be achieved without the expenditure of energy. The existence of such barriers was first suggested in Hopfield's work on kinetic proofreading (Hopfield, 1974; Ninio, 1975). He showed, in effect, that if a biochemical system operates at thermodynamic equilibrium, then physics imposes a barrier to how well a given information processing task can be accomplished. (In his case, the task was achieving fidelity in transcription and translation.) The only way to bypass this barrier is to expend energy and maintain the system away from equilibrium. Kinetic proofreading is one way to do this.

Here, we identify a "Hopfield barrier" for sharpness in gene regulation. With n binding sites, a Hill coefficient of n sets the Hopfield barrier; at thermodynamic equilibrium, no GRF can reach it, even with higher-order cooperativities. If, however, energy is expended to maintain regulatory DNA away from equilibrium, then much greater sharpness can be achieved.

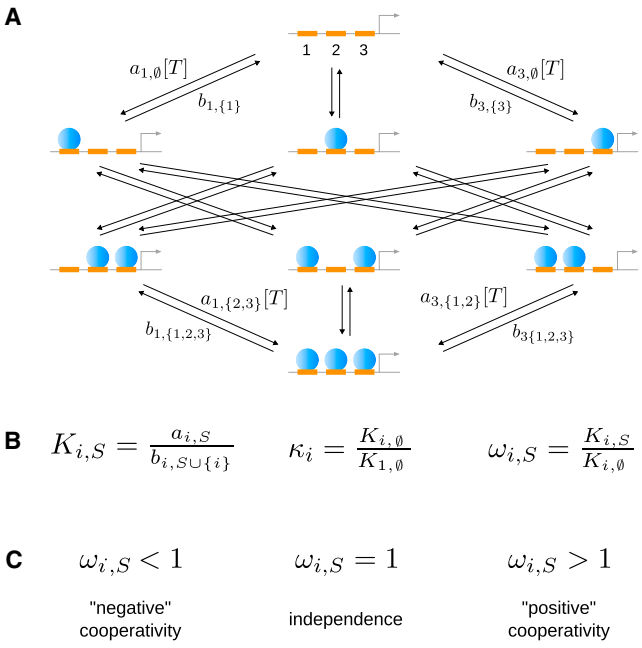


Figure 2. Linear Framework Model

(A) The graph G_3 showing the 8 microstates and the associated labeled, directed edges, with middle-layer labels omitted for clarity.

(B) The essential parameters at thermodynamic equilibrium. The association constants $K_{i,S}$ have units of $(\text{concentration})^{-1}$, and κ_i and $\omega_{i,S}$ are non-dimensional. $S \cup \{i\}$ is the set in which site i has been added to the sites in S.

(C) The higher-order cooperativity $\omega_{i,S}$ measures whether binding of T to site i, when T is already bound to the sites in S, shows reduced affinity $(\omega_{i,S} < 1)$, unchanged affinity $(\omega_{i,S} = 1)$, or enhanced affinity $(\omega_{i,S} > 1)$, as compared to binding to site i, when no other sites are bound.

RESULTS

The Rationale for the Model

We introduce a mathematical model for analyzing gene regulation. As with all models, the conclusions depend on the assumptions (Gunawardena, 2014). Our assumptions are guided by the example of *hb*, but the model is general and not restricted to this example. The anterior expression pattern of *hb* is believed to be regulated by, at least, three enhancers (Perry et al., 2011). Both the classical P2 enhancer, which is promoter proximal, and a shadow enhancer, located ~3 kb upstream, drive broad anterior patterns early in embryo development. Later, the central stripe enhancer drives expression near the middle of the embryo. Bcd is a transcriptional activator for the P2 and shadow enhancers; the stripe enhancer is also targetted by transcriptional repressors. The stripe enhancer has no effect on sharpness early in nuclear cycle 14 (Perry et al., 2012), when the data on which Equation 1 is based (Gregor et al., 2007) were acquired.

Accordingly, we focus on a single TF, binding to a specified but arbitrary number of sites and functioning solely as a transcriptional activator. TF binding sites can be anywhere on the genome and are not assumed to be confined to a single enhancer; thus, our analysis is not limited to the 5–7 Bcd binding sites thought to be present in the *hb* P2 enhancer.

Molecular mechanisms other than TF binding and unbinding, such as nucleosomes or co-regulators, are not directly represented in our model, but their influence is captured through their effects on rate constants and the dependence of these constants on the state of DNA ("microstate", see below). This permits general conclusions to be drawn without knowing the specific mechanisms at work in a particular gene but does not allow us to assign mechanisms to the effects we find. Other molecular features, such as post-transcriptional mechanisms or network effects like feedback, could influence sharpness, but these are not thought to be relevant for Bcd regulation of *hb*. Addressing such features in future work may yield further insights into sharpness.

A Graph-Based Model of Gene Regulation

We recently developed a graph-based "linear framework" for modeling gene regulation (Ahsendorf et al., 2014). We use this to formulate a general model of a gene responding to a TF, called T, binding as a monomer to a number, n, of sites. Oligomerization of a TF in solution can contribute to gene-expression sharpness, but it is not thought to be significant in Bcd regulation of *hb* (Lebrecht et al., 2005; Gregor et al., 2007), and we do not consider it here. In this section and the next, we discuss the quantitative details of how T binds and unbinds, how cooperativity is defined, and how T influences transcription.

The model consists of the labeled, directed graph, G_n (Figure 2A). The vertices of G_n represent the microstates, or patterns of T bound to DNA, with the binding sites labeled by the numbers $1, \cdots, n$. The edges represent binding or unbinding of T from the microstates. Each edge has a label describing the rate of the corresponding reaction. The label on a binding edge is the product of the concentration of T, $[T]$, and an on-rate for binding, $a_{i,S}$, where i is the binding site and S is the subset of sites at which T is already bound. Subsets are denoted $\{i_1, \cdots, i_k\}$, where the site indices, $i_1, \cdots, i_k$, are drawn from the numbers $1, \cdots, n$. The label on an unbinding edge is an off-rate, $b_{i,S'}$, where S' is the subset of sites to which T is bound and i is one of the sites in S'.

Importantly, the on-rates, $a_{i,S}$, and the off-rates, $b_{i,S'}$, can depend on the site of binding or unbinding, i, as well as on the pattern of existing binding to a subset of sites, S or S'. This reflects the potential influence of background mechanisms, such as nucleosomes or co-regulators, and allows higher-order cooperativities to be introduced below.

The linear framework describes how such a graph gives rise to a stochastic master equation for the probabilities of the microstates. As in the thermodynamic formalism, we make the basic assumption that regulatory DNA is at steady state. However, unlike the thermodynamic formalism, the linear framework allows steady-state probabilities to be calculated regardless of whether or not the system is at thermodynamic equilibrium (Ahsendorf et al., 2014); see the Experimental Procedures and the Supplemental Information.

Higher-Order Cooperativities and the Exchange Formula

Higher-order cooperativity between multiple TF binding sites may be important in gene regulation, but thermodynamic formalism models have usually been limited to pairwise cooperativity. This

is not a fundamental limitation but arises from technical difficulties with the principle of detailed balance, which imposes algebraic constraints on higher-order cooperativities (Supplemental Information) that have not been worked out within the thermodynamic formalism. Detailed balance, or "microscopic reversibility," is a fundamental requirement arising from the time-reversal symmetry of the laws of physics (Mahan, 1975). The constraints are a serious obstacle because they mean that the numerical values of higher-order cooperativities cannot be chosen independently. Thus, it is important to determine these constraints (Equation 2) and to thereby identify a subset of cooperativities for which the numerical values are independent (Equation 3).

If the regulatory system described by G_n can reach thermodynamic equilibrium, the relevant parameters are the association constants $K_{i,S}$ (Figure 2B), of which there are $n2^{n-1}$ (Supplemental Information). To define higher-order cooperativities at equilibrium, we compare the binding of T to site i when T is already bound at the sites in S ($K_{i,S}$) to the binding of T to site i when T not bound elsewhere ($K_{i,\varnothing}$, where $\varnothing$ denotes the empty set). This yields a non-dimensional higher-order cooperativity, $\omega_{i,S} = K_{i,S}/K_{i,\varnothing}$, for which the value indicates whether or not there is positive or negative cooperativity or independence (Figure 2C). To non-dimensionalize the remaining association constants, we define $\kappa_i = K_{i,\varnothing}/K_{1,\varnothing}$,

The number of sites in S is called the order of $\omega_{i,S}$ and denoted $\#S$; it specifies how many sites collaborate to influence binding. Pairwise cooperativity corresponds to order 1. Thermodynamic formalism models set $\omega_{i,S} = 1$ for $\#S > 1$. In this case, detailed balance reduces to a symmetry requirement on pairwise cooperativities, $\omega_{i,\{j\}} = \omega_{j,\{i\}}$ (see Equation 2). With only pairwise cooperativity, there are only $n(n-1)/2$ parameters, instead of $n2^{n-1}$; this greatly simplifies thermodynamic formalism calculations.

We prove that, because of detailed balance, higher-order cooperativities must satisfy the "exchange formula" (Supplemental Information),

$$\omega_{i,S\cup\{j\}}\omega_{j,S} = \omega_{j,S\cup\{i\}}\omega_{i,S}, \tag{2}$$

which summarizes the algebraic constraints among the cooperativities. Here, i, j are sites not in S, while the notation $S\cup\{v\}$, for $v = i$ or $v = j$, denotes the addition of v to the sites in S. We further prove that, if we retain only those $\omega_{i,S}$ for which i is less than all the sites in S (abbreviated $i < S$), then the parameters

$$K_{1,\varnothing}, \quad \kappa_i\ (i > 1), \quad \omega_{i,S}\ (i < S), \tag{3}$$

of which there are $2^n - 1$, are algebraically independent, and all the $K_{i,S}$ can be calculated from them using Equation 2 (Supplemental Information). We can thus vary the parameters in Equation 3 independently and be confident that detailed balance holds. These fundamental results provide the basis for the equilibrium calculations that follow.

Equilibrium Gene Regulation Functions

To calculate a gene regulation function (GRF), we make the same basic assumption as in the thermodynamic formalism and consider the overall rate of transcription to be an average over the steady-state probabilities of the microstates. For this, we must specify the rate of transcription in each microstate, about which surprisingly little is known for eukaryotic genes. As explained above, we assume that T acts as a transcriptional activator (Supplemental Information), so that the binding of T does not reduce the expression level. We consider three expression strategies which work for any number of sites (Figure 3A): all-or-nothing, in which transcription only occurs when all sites are bound; one-or-more, in which transcription occurs when at least one site is bound; and average binding, in which transcription is proportional to the number of bound sites.

It is computationally infeasible to explore all expression strategies, but these three strategies broadly sample the spectrum of possibilities (Supplemental Information). All-or-nothing and one-or-more are extreme opposites, while average binding is an intermediate strategy. All-or-nothing is widely used in thermodynamic formalism models, and average binding corresponds to the "fractional saturation" used in models of protein allostery (Monod et al., 1965; Mirny, 2010).

The level of protein expression after normalization to its asymptotic maximum is a rational function of $x = [T]$, denoted $f_n(x)$, which has the form, for the all-or-nothing strategy (Supplemental Information),

$$f_n(x) = \frac{c_n x^n}{1 + c_1 x + \cdots + c_n x^n}. \tag{4}$$

The coefficients c_k are given (Supplemental Information) by a sum of products

$$c_k = \left(\sum_{1 \le i_1 < \cdots < i_k \le n} \left(\prod_{j=1}^{k} \kappa_{i_j} \omega_{i_j,\{i_{j+1},\cdots,i_k\}} \right) \right) (K_{1,\varnothing})^k, \tag{5}$$

which involves only the independent parameters in Equation 3 and allows higher-order cooperativity of any order up to the maximum of $n - 1$. For the other strategies (Figure 3A), only the numerator of Equation 4 changes (Supplemental Information). The GRFs discussed in this paper are strictly increasing functions (Supplemental Information). For the all-or-nothing strategy at equilibrium, they also appear to be sigmoidal ("S shaped"), so that the derivative of the GRF has only a single maximum, but this is not so in general (Figure 3B).

The algebraic form of the Hill function, $\mathcal{H}_a(x) = x^a/(1 + x^a)$, is closest to that of the GRF for the all-or-nothing strategy in Equation 4 (Supplemental Information). If $a < n$, it is clear that the former cannot algebraically resemble the latter because the degrees of their respective denominator polynomials are different. If $a = n$, algebraic resemblance is only possible, and then only approximately, if the parameters in the GRF are given implausible numerical values (Supplemental Information). We see that Hill functions are not GRFs. However, the question posed by Equation 1 is whether a GRF can match the shape of a Hill function. This is a more delicate problem.

Position and Steepness as Quantitative Measures of Shape

To determine the match between a GRF and a Hill function, we introduce two quantitative measures of shape. We first normalize the concentration scale of $x = [T]$, in a similar way to Equation 1,

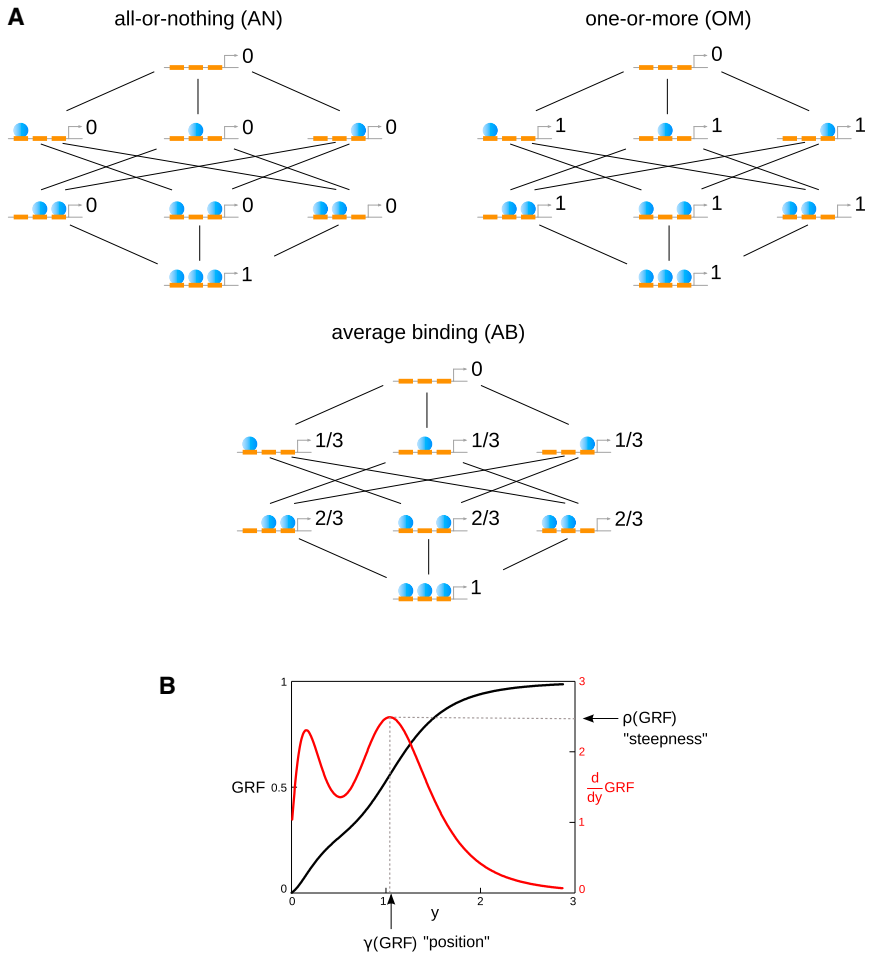


Figure 3. Gene Expression Strategies and Shape Measures

(A) The graph G_3, with pairs of reversible edges as unlabelled single lines for clarity, illustrating three expression strategies for a transcriptional activator. Each microstate is annotated with a number describing the corresponding rate of gene expression, with the maximal rate normalized to 1. (B) Plot of a hypothetical GRF (black), together with its derivative (red) showing steepness (ρ) and position (γ), as defined in Equation 6. The derivative can have multiple local maxima, and ρ and γ are defined at the global maximum.

by taking $x_{0.5}$ to be the concentration at which f_n is half-maximal, so that $f_n(x_{0.5}) = 0.5$, and setting $y = x/x_{0.5}$. The normalized GRF is $g_n(y) = f_n(yx_{0.5})$, where y and $g_n(y)$ are now both non-dimensional quantities. Note that for $\mathcal{H}_a(x)$, $x_{0.5} = 1$, so that Hill functions are already normalized.

To quantify shape, we take the maximum derivative, $\rho(g_n)$ ("steepness"), and the position of the maximum derivative $\gamma(g_n)$ ("position"),

$$\rho(g_n) = \max_{y \geq 0} \frac{dg_n}{dy}, \quad \gamma(g_n) = z \text{ such that } \left.\frac{dg_n}{dy}\right|_{y=z} = \rho(g_n), \qquad (6)$$

which are also non-dimensional quantities (Figure 3B). The advantage of γ and ρ is that they can be calculated from g_n, in contrast to a numerical fit to a Hill function, which is subject to statistical noise. Two GRFs with the same γ and ρ ("matched") are not identical but have similar sharpness. Considering only one measure of sharpness, such as ρ, can be misleading (below).

Impact of Higher-Order Cooperativity on Sharpness

We first determined the position and steepness of a GRF in the all-or-nothing strategy with $n = 5$ sites, allowing higher-order cooperativity of any order. Although γ and ρ depend only on the coefficients c_k in Equation 5, these aggregated parameters have no biochemical meaning and we seek instead to understand how γ and ρ depend on the affinities and cooperativities in Equation 3, which are defined in terms of molecular interactions. Because of normalization, γ and ρ do not depend on $K_{1,\varnothing}$ (Supplemental Information), so we set $K_{1,\varnothing} = 1$ in units of $(\text{concentration})^{-1}$ and chose κ_i and $\omega_{i,S}$ to be in the range $[10^{-3}, 10^{3}]$ by random logarithmic sampling. We believe this range is generous, but most of our results do not depend on it (below; Supplemental Information).

A sample of 10^5 GRFs chosen in this way reveals that position and steepness are not independent but are constrained within a crescent-shaped region in which the highest steepness is found at the extremes of position (Figure 4A). On the left of the region, high steepness occurs only for very low position and the resultant GRFs are highly degenerate: when these GRFs are fitted to Hill functions, they yield a Hill coefficient of $a = 1$ (inset at top and caption). The marginal distribution (top) shows that these degenerate GRFs account for nearly half of all GRFs in this parameterisation. Degeneracy underscores the importance of considering γ together with ρ.

The upper edge of the crescent-shaped region has low probability. We therefore used, for Figure 4 and those that follow, a biased sampling algorithm to identify the boundary of the region (Supplemental Information), with the same parameter range but with the GRFs filtered so that $\gamma(g_n)$ lies in the interval $[0.5\gamma(\mathcal{H}_n), 1.5\gamma(\mathcal{H}_n)]$. This focuses on the GRFs of interest and avoids the degeneracy near $\gamma(g_n) = 0$. We found the gray boundary in Figure 4A.

The right-hand edge of the gray boundary coincides with that of the randomly sampled region in a series of line segments. Strikingly, the "Hill line" on which Hill functions are located (magenta curve) lies just to the right of this boundary. We see that a GRF cannot have greater position than a Hill function of the same steepness, so that the Hill functions define a barrier. The line segments on the boundary of the region touch the Hill line at their corners and these occur, surprisingly, at

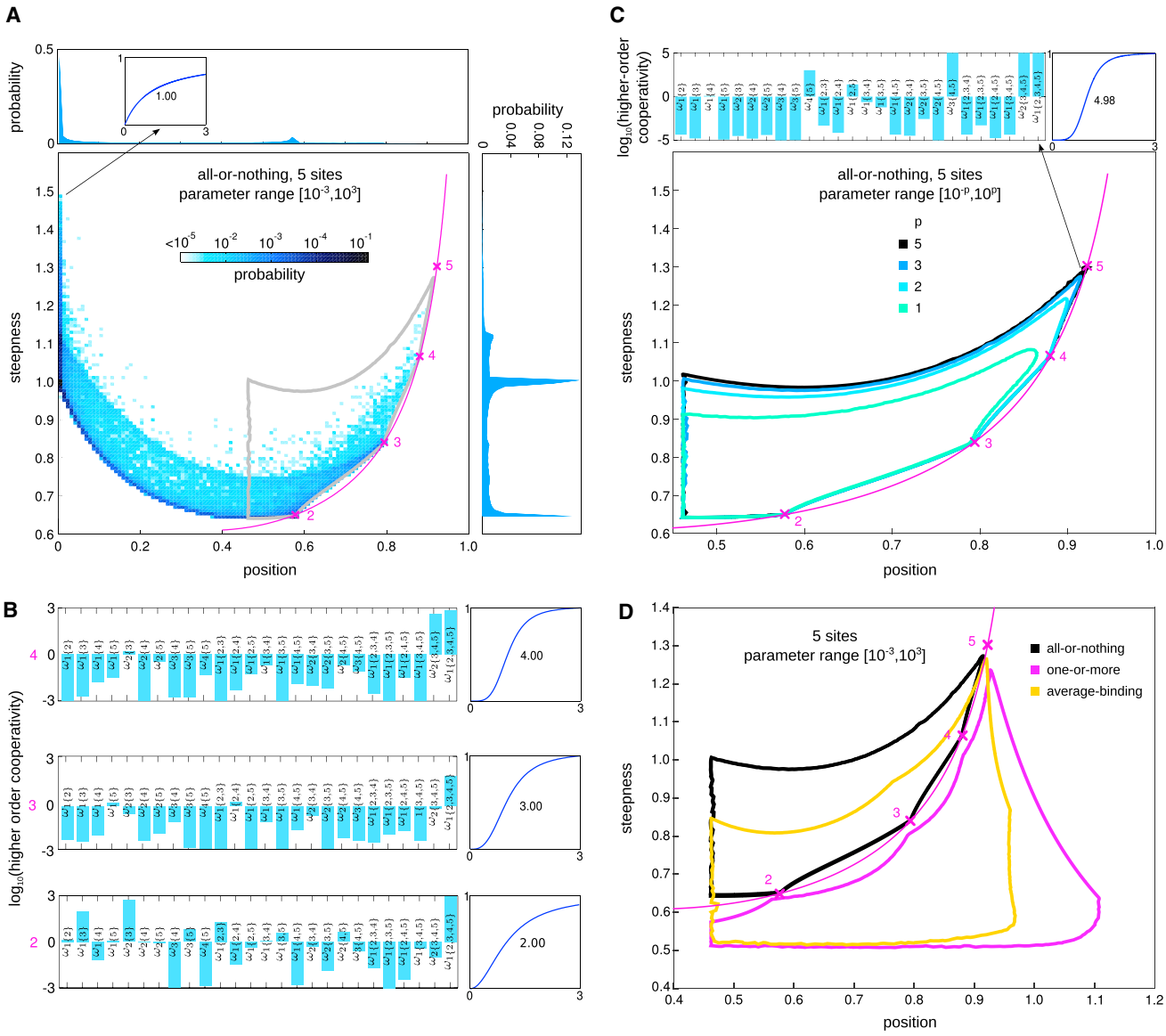


Figure 4. Position and Steepness for $n = 5$ Sites

(A) Probability density function (blue points) of $(\gamma(g_5), \rho(g_5))$, obtained by random sampling, with the respective marginal distributions (top and right). The inset (top) shows the GRF of the marked point, annotated with the value *a* obtained by fitting to $\mathcal{H}_a$. The gray line marks the boundary of the position-steepness region, obtained by a biased sampling algorithm (Supplemental Information). The magenta line is the locus of $(\gamma(\mathcal{H}_a), \rho(\mathcal{H}_a))$ for varying *a* (the "Hill line"), with the integer values of *a* marked by magenta crosses and numbers.

(B) Higher-order cooperativities (left), plotted on a logarithmic scale, for the GRFs closest in (γ, ρ) distance to the integer Hill coefficients (magenta crosses) in (A), with the Hill coefficient annotated on the left (magenta). The corresponding curve (right) is annotated with the value *a*, obtained by fitting to $\mathcal{H}_a$.

(C) Position-steepness boundaries with the parameter range $[10^{-p}, 10^{p}]$ for varying *p*. At the top are the higher-order cooperativities (left) and curve (right) for the marked GRF closest to $\mathcal{H}_5$ within the $p = 5$ region, plotted as in (B).

(D) Position-steepness regions for all three expression strategies; see also Figures 6 and S1.

See also Figure S1.

exactly the integer values, 2, 3, 4, of the Hill coefficient. When the GRFs which are closest to these points are fitted to Hill functions, the estimated Hill coefficients correspond very closely to the integer values (Figure 4B, right). Such a close correspondence in fitted shape is unexpected in view of the lack of algebraic resemblance between GRFs and Hill functions, as discussed above. The emergence of bona fide GRFs which closely match the shape of Hill functions with integer Hill coefficients is intriguing in view of the coefficient 5 found in Equation 1. However, this shape matching requires high levels of positive and negative higher-order cooperativity of all orders (Figure 4B, left).

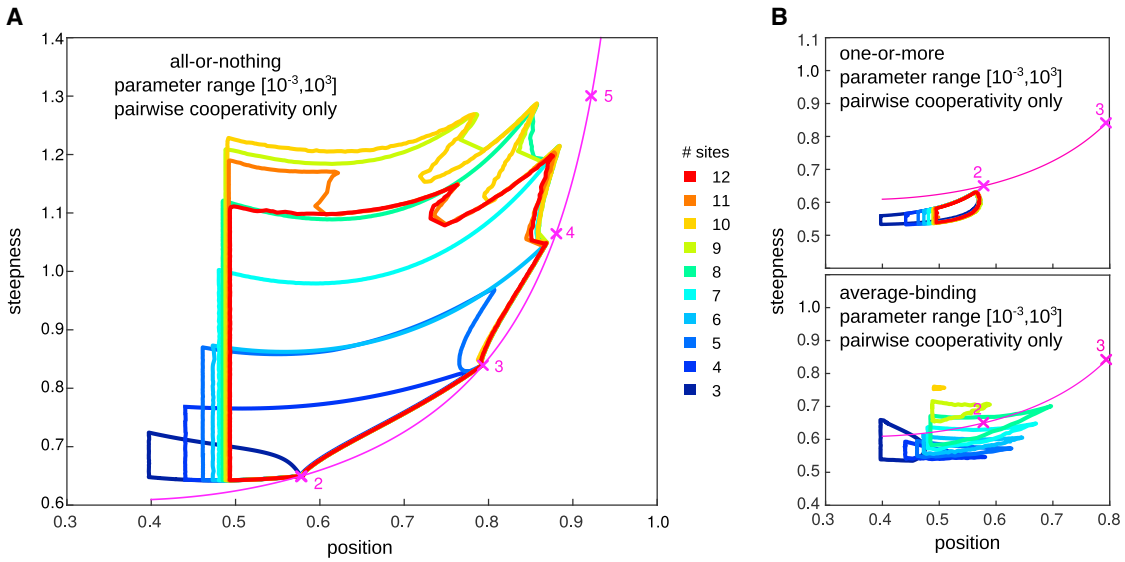


Figure 5. Pairwise Cooperativity Only

(A and B) Each panel uses the color-code in the center and shows the Hill line in magenta. (A) Position-steepness regions for the all-or-nothing strategy. (B) Position-steepness regions for the one-or-more (top) and average-binding (bottom) strategies. The biased sampling algorithm had to be modified to find the average-binding regions (Supplemental Information).

The boundary of the position-steepness region lies below $\mathcal{H}_5$ and approaches it at the tip of a cusp. Changing the parameter range does not alter the line segments in the boundary, but the tip of the cusp approaches closer to $\mathcal{H}_5$ as the range is increased (Figure 4C). The GRF closest to $\mathcal{H}_5$ has a fitted Hill coefficient close to 5 (top, right), although not as close as for integer values less than 5. This still requires high levels of positive and negative higher-order cooperativity of all orders (top, left).

Each expression strategy reveals a different trade-off between position and steepness (Figure 4D). The Hill line also presents a barrier to the one-or-more strategy but from the opposite side, while the average-binding strategy straddles the Hill line. For the one-or-more strategy, the position-steepness region approaches at cusps the Hill functions whose coefficients are integers less than 5 (Figure 4D) but, in contrast to the all-or-nothing strategy, the region does not touch the Hill line and the integer-valued Hill functions are not closely matched to the nearest GRFs unless the parameter range is increased (data not shown). The barrier presented by the Hill line seems, therefore, to act differently in the all-or-nothing and one-or-more strategies. Regardless of the expression strategy, $\mathcal{H}_5$ offers a barrier to all strategies with $n = 5$ sites: each region lies below it and only approaches it at the tip of a cusp as the parameter range is increased.

The features found above are reproduced for different numbers of sites (Figure S1 for $n = 7$ sites).

Pairwise Cooperativity Alone Permits Limited Sharpness

Thermodynamic formalism models have typically been limited to pairwise cooperativity. We restricted ourselves to pairwise cooperativity by setting $\omega_{i,S} = 1$ for $\#S > 1$. We found that the position-steepness region for the all-or-nothing strategy increases initially with increasing n but then shrinks in extent and no GRF approaches close to $\mathcal{H}_5$ (Figure 5A). The one-or-more and average-binding strategies do not even get close to $\mathcal{H}_3$ (Figure 5B).

Thus, in contrast to common assumptions, pairwise cooperativity alone is insufficient for sharp responses in eukaryotic genes. None of the expression strategies considered here can account for $\mathcal{H}_5$ in Equation 1 with only pairwise cooperativity, no matter how many sites are available.

Non-equilibrium GRFs Exceed the Equilibrium Sharpness Barriers

If the system is maintained away from thermodynamic equilibrium by energy expenditure, detailed balance no longer holds. The non-equilibrium GRF for the all-or-nothing strategy then takes the form (Supplemental Information)

$$f_n^{ne}(x) = \frac{d_n x^n + \cdots + d_{2^n-1} x^{2^n-1}}{e_0 + e_1 x + \cdots + e_{2^n-1} x^{2^n-1}}, \tag{7}$$

where the coefficients of the highest order term, x^{2^n-1}, in the numerator and the denominator are equal, so that $d_{2^n-1} = e_{2^n-1}$. GRFs for the other strategies differ only in the numerator (Supplemental Information). The denominator of Equation 7 shows a striking increase in degree, from n to $2^n - 1$, in comparison to that of the equilibrium f_n in Equation 4, despite the number of sites being the same.

The parameters in Figure 2B are no longer meaningful away from equilibrium and the coefficients d_i and e_i in Equation 7 are expressions in the rate constants $a_{i,S}$ and $b_{i,S'}$. For reasons discussed below, the largest number of sites that we can feasibly analyze is $n = 3$ (Supplemental Information).

We took non-dimensional parameters $a_{i,S}/a_{1,\varnothing}$ and $b_{i,S'}/b_{1,\{1\}}$ in the range $[10^{-2}, 10^2]$, deliberately restricting the range so that,

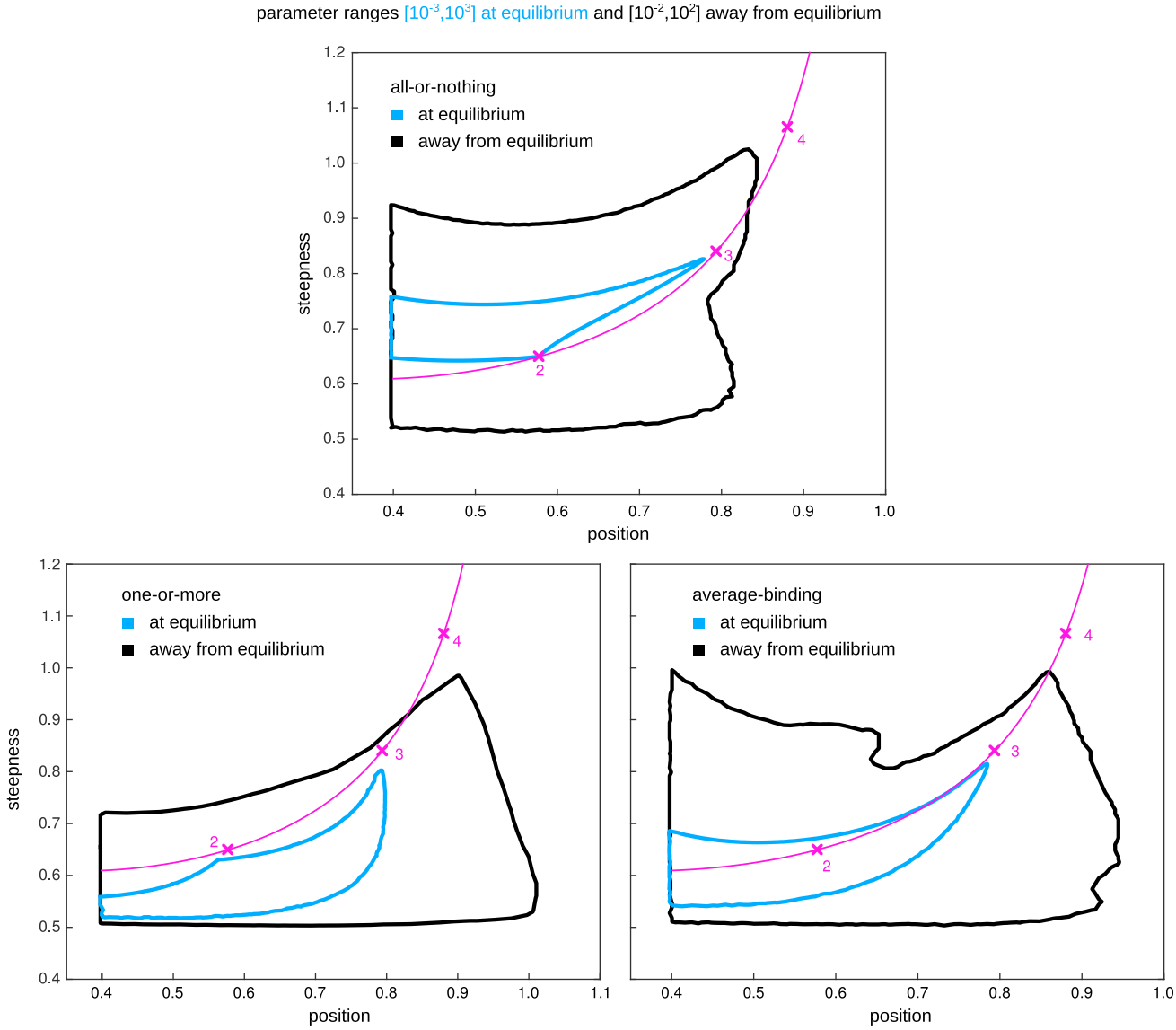


Figure 6. The Non-equilibrium Case for $n = 3$ Sites

Position-steepness regions for all expression strategies, showing the equilibrium (blue) and non-equilibrium (black) boundaries. The horizontal scale for the one-or-more strategy is extended.

See also Figure S2.

if the system were at equilibrium, it would be comparable with the previous equilibrium analysis (Supplemental Information). Because of normalization, the steepness and position of g_n^{ne} are independent of the values of $a_{1,\varnothing}$ and $b_{1,\{1\}}$ (Supplemental Information), so we set $a_{1,\varnothing} = 1$ and $b_{1,\{1\}} = 1$ in their respective units. For each expression strategy, we found (Figure 6) that the non-equilibrium position-steepness region is much enlarged (black boundary) compared to the corresponding equilibrium region for the same number of sites (blue boundary). The non-equilibrium regions now include the Hill line up to $\mathcal{H}_3$ and the all-or-nothing region can reach as far as $\mathcal{H}_5$ if the parameter range is increased (Figure S2A).

The limitation to $n = 3$ sites arises from loss of detailed balance, which leads to a dramatic increase in the complexity of the coefficients in Equation 7 (Supplemental Information; see the Discussion). This complexity is algebraic, not numerical. To compute position-steepness regions, cooperativities are treated as symbols whose numerical values are assigned by sampling. Symbolic calculation of the GRF is extremely expensive away from equilibrium but numerical calculation of individual GRFs presents no particular difficulty.

Symbolic treatment of parameters is informative because it reveals the structure of the non-equilibrium GRF (Equation 7). The denominator of this GRF increases in degree exponentially

with n but the denominator of the equilibrium GRF (Equation 4) increases only linearly. This discrepancy arises from loss of detailed balance (Supplemental Information). With $n = 3$ sites, the non-equilibrium position-steepness region comfortably exceeds the equilibrium region (Figure 6). Because of the exponential increase in the degree of the GRF denominator, when there are $n = 5$ sites, or however many sites are relevant for Bcd regulation of *hb*, the discrepancy in the position-steepness regions will be even greater and the non-equilibrium region will extend well beyond $\mathcal{H}_5$.

In confirmation of this, we used numerical parameter values to find a non-equilibrium GRF on $n = 5$ sites, with parameters in the same range, $[10^{-2}, 10^{2}]$, as in Figure 6, whose position and steepness match that of $\mathcal{H}_6$ (Figure S2B). If the parameter range is increased to $[10^{-3}, 10^{3}]$, then there is a GRF on four sites, for which position and steepness match that of $\mathcal{H}_{5.7}$ (Figure S2B). Being away from equilibrium makes it much easier to achieve the sharpness required for Equation 1.

DISCUSSION

Eukaryotic gene regulation lies at the nexus of many of the central issues in modern biology, including multi-cellular development (Davidson, 2006), the evolution of complexity (Carroll, 2008), cellular reprogramming (Takahashi and Yamanaka, 2016) and synthetic biology (Keung et al., 2015). The extraordinary molecular complexity implicated in such regulation continues to present a formidable challenge. It has made it difficult to see the wood for the trees, to discern general principles and to unravel how different molecular mechanisms contribute to specific forms of information processing.

In this paper, we have presented compelling evidence that the bacterial paradigm, upon which it has been so convenient to default, is not sufficient for reasoning about eukaryotic genes and we have introduced appropriate quantitative concepts for doing so. We have done this by taking seriously the lessons of the bacterial paradigm itself. The paradigm relied on analyzing the physics of interaction between TFs and DNA. What we have done here is to update this foundation for two of the key processes that influence gene-expression sharpness in eukaryotes: information integration and energy expenditure.

While information integration is often acknowledged, it has not been defined sufficiently clearly to know how to find it, making experimental analysis problematic. We have introduced the concept of "higher-order cooperativity," $\omega_{i,S}$ (Figure 2B), for a system at thermodynamic equilibrium, as a measure of how the affinity of TF binding at site i is influenced by the presence of TF bound at the sites in S. This is a precisely defined quantity that allows experimentally-testable hypotheses to be framed.

Different TFs often work together and the definition of higher-order cooperativities that we have given here for homotypic interactions of a single TF can be readily extended to heterotypic interactions between different TFs. Such cooperativities could arise from nucleosomes (Mirny, 2010; Voss et al., 2011) or from co-regulators like Mediator and CBP/p300 (Borggrefe and Yue, 2011; Wang et al., 2013). Mediator is especially provocative as a potential mechanism of higher-order cooperativity. Mediator has around 30 subunits and the Med1 subunit alone interacts with up to 20 different TFs (Borggrefe and Yue, 2011). Some TFs interact with multiple subunits and the overall Mediator complex exhibits different conformations, which suggests how local information may be globally integrated (Nussinov et al., 2013). Our analysis shows that if gene expression follows an all-or-nothing strategy, high levels of positive and negative higher-order cooperativity of all orders are needed to yield high levels of sharpness (Figures 4B and 4C). Experimental measurements will show whether co-regulators like Mediator or CBP/p300 can meet these requirements.

When energy is expended during gene regulation, much higher sharpness can be achieved for the same number of TF binding sites (Figure 6). The concept of a "Hopfield barrier" offers a way to articulate this rigorously. With n binding sites, the Hopfield barrier to sharpness is set by the Hill function $\mathcal{H}_n$, whose steepness cannot be exceeded by any equilibrium GRF (Figure 4D). For GRFs in the all-or-nothing strategy, the Hill line itself forms a Hopfield barrier (Figure 4C). However, both these barriers are readily breached away from equilibrium (Figures 6 and S2).

There are many routes through which energy can be expended, including chromatin reorganization, nucleosome displacement, protein post-translational modification and DNA methylation. Experiments which perturb these routes and assay the impact on sharpness can bring to light which energy expending mechanisms are particularly relevant. For the specific example of *hb* regulation by Bcd, we note that Bcd binds to *Drosophila* Mediator in a way that affects early embryonic patterning (Park et al., 2001; Bosveld et al., 2008) and that Bcd also binds to the Sin3/Rp3 histone-deacetylase complex (Singh et al., 2005). The impact of these interactions on sharpness appears not to have been previously studied. The early *Drosophila* embryo provides an unrivalled experimental context for testing the hypotheses made here and this is now work in progress.

Real-time studies have already confirmed the importance of non-equilibrium kinetics in gene regulation (Voss et al., 2011; Hammar et al., 2014). Coulon et al. (2013) have also argued for the importance of a non-equilibrium perspective. Our results strongly endorse this but an important challenge lies ahead. When a system is at thermodynamic equilibrium, detailed balance implies that any path to a microstate can be used to calculate the steady-state probability of the microstate. The history of the system is irrelevant. Away from equilibrium, detailed balance no longer holds and all possible paths to a microstate must be examined to calculate its steady-state probability. Non-equilibrium systems are history dependent. The resultant combinatorial explosion results in a profound increase in algebraic complexity, which manifests itself in the striking difference between the equilibrium GRF in Equation 4 and the non-equilibrium GRF in Equation 7. It is this which leads to the breaching of the Hopfield barrier. We suspect that further insights into non-equilibrium gene regulation are concealed within this algebraic complexity. We are only just learning how to uncover them (Ahsendorf et al., 2014).

The concepts introduced here encourage us to examine other forms of genetic information processing. If energy expenditure is important, what is it buying? What could not be achieved if regulatory DNA is at thermodynamic equilibrium? Such questions

can be answered by quantifying each information processing task, as we have done here for sharpness (Equation 6), and developing experimental systems in which it can be measured. We may look forward in this way to a quantitative classification of the kinds of information processing that genes undertake and an understanding in molecular terms of how energy expenditure breaks the corresponding Hopfield barriers.

We note two further implications of the present paper. First, Hill functions emerge in an unexpected light. When Archibald Vivian (A.V.) Hill first introduced them in 1913, he recognized that they had no biochemical justification and were only a convenient fit to the data on oxygen binding to haemoglobin (Hill, 1913). The empirical nature of Hill functions has been repeatedly pointed out (Engel, 2013; Weiss, 1997). We were all the more surprised, therefore, to find that, when the Hill coefficient is an integer, there are bona fide GRFs which are statistically indistinguishable from Hill functions (Figure 4B). In this sense, the Hill functions appear to be closer to biochemistry than Hill, or anyone else, could have imagined. We hope to clarify the mathematical reasons for this in subsequent work.

Second, we have exploited mathematics differently here to what is sometimes expected of it. We have relied on data to frame the question but we have not fitted any mathematical models to data. In recent years, experimental biologists have become more comfortable with the idea that theory can follow experiment, as a way to analyze and understand data. Our colleague Rob Phillips calls this "Figure 7 theory" (Phillips, 2015). Here, we have used mathematics to introduce concepts and to determine the limits of what can be expected, thereby providing a foundation for designing new kinds of experiments. Experiment will follow theory. This is "Figure 1 theory" (Phillips, 2015). We believe it has much to recommend it as we face the daunting molecular complexity of eukaryotic gene regulation. We need to think about how such regulation works using concepts which are not just based on intuition and induction but are also grounded in the underlying physics, which is the bedrock on which all biology rests. This is an old lesson (Gunawardena, 2013; Bialek, 2015). If we have lost sight of it in the press of mastering the molecular details, now is the time to revisit it and construct a new paradigm for eukaryotic gene regulation.

EXPERIMENTAL PROCEDURES

The mathematical model and the results presented here are based on the "linear framework" for gene regulation (Ahsendorf et al., 2014) (see the Supplemental Experimental Procedures for full details). The framework starts from a labeled, directed graph, $\mathcal{G}$, which gives rise to a stochastic master equation,

$$\frac{du}{dt} = L(\mathcal{G}).u,$$

for the vector of microstate probabilities, $u = (u_1, \cdots u_N)^t$. Here, $L(\mathcal{G})$ is the Laplacian matrix of $\mathcal{G}$ and N is the number of microstates in the graph. For the graph G_n used here, $N = 2^n$, where n is the number of TF binding sites. Provided $\mathcal{G}$ is strongly connected, which is the case for G_n, the steady state, u^*, at which $(du/dt)\,|_{u=u^*} = 0$, is unique up to a scalar multiple. A basis vector may be calculated in terms of the edge labels in one of two ways, depending on whether or not the system is at equilibrium. If the system reaches thermodynamic equilibrium, u_i^* can be calculated by choosing any path of reversible edges from the reference vertex 1 to i and taking the product, over all reversible edges in the path, of the ratio of the label on the forward edge, in the direction from 1 to i, to the label on the reverse edge. The principle of detailed balance ensures that this result is independent of the chosen path because of the cycle condition: on any cycle of reversible edges, the product of the labels going clockwise around the cycle equals the product of the labels going counterclockwise. The cycle condition leads to the exchange formula in Equation 2, which allows the algebraically independent set of parameters in Equation 3 to be chosen. For G_n, a path of reversible edges can be chosen from 1, the vertex with no sites bound, to i, such that u_i^* is expressed in terms of the independent parameters. Away from equilibrium, u_i^* has to be calculated using the matrix-tree theorem as a sum, over all directed spanning trees rooted at i, of the product of the labels on the edges of each spanning tree. Once u^* is known, the state-state probability of microstate i is given by

$$\Pr(i) = \frac{u_i^*}{u_1^* + \cdots + u_N^*}.$$

For a system that reaches thermodynamic equilibrium, the denominator in this formula is the partition function of equilibrium statistical mechanics, but the formula equally holds for a system away from thermodynamic equilibrium with u^* calculated as above. The gene regulation function for mRNA production rate as output is defined as an average over the steady-state probabilities,

$$\frac{d}{dt}[\mathrm{mRNA}] = \sum_{1 \le i \le N} r(i)\Pr(i).$$

The expression rate in microstate i, given by $r(i)$, depends on the gene expression strategy being followed, as specified in Figure 3. To obtain protein level as output, assume that mRNA is linearly degraded and that the steady-state protein level is proportional to the steady-state mRNA level. The proportionality constants are absorbed in the normalization that underlies the definitions of position and steepness in Equation 6.

In the equilibrium GRFs, the quantities u_i^* depend on paths of reversible edges from 1 to i, which can incur up to n factors of $x = [T]$, so that the degree of the denominator polynomial in $\Pr(i)$ is n (Equation 4). In contrast, in the non-equilibrium GRFs, the quantities u_i^* depend on directed spanning trees rooted at i, which each have $N - 1$ edges and can incur up to $N - 1$ factors of x, so that the degree of the denominator polynomial becomes $2^n - 1$ (Equation 7).

The numerical results presented in Figures 4, 5, and 6 are obtained by a biased sampling algorithm in which the boundary of the position-steepness region is found by successive approximation. An initial region is found by independently selecting parameter values for GRFs by logarithmic random sampling within the specified range, calculating the (γ, ρ) coordinates of these GRFs, and determining the enclosing boundary. This initial boundary is then successively improved by randomly altering GRFs on the current boundary until the area of the region ceases to increase. The details, along with the tests that were used to confirm convergence and to check the numerical accuracy of the results, are given in the Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and two figures and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.06.012.

AUTHOR CONTRIBUTIONS

A.D. and J.G. formulated and supervised the project. J.E. developed the algorithms and carried out the numerical calculations, which J.E. and F.W. separately tested. F.W. and J.G. carried out the mathematical derivations. J.E., A.D., and J.G. made the figures. J.G. wrote the paper with the assistance of all authors.

ACKNOWLEDGMENTS

We thank the editor, Robert Kruger, and the anonymous reviewers for their help. We thank Rebecca Ward for editorial assistance. We gratefully acknowledge the Department of Systems Biology for supporting J.E. F.W. was

supported by a National Science Foundation (NSF) grant (GRF DGE1144152); A.D. was supported by an NIH grant (U01 GM103804) and NSF CAREER grant (1452557); and J.G. was supported by an NSF grant (1462629).

Received: July 3, 2015
Revised: March 8, 2016
Accepted: June 1, 2016
Published: June 30, 2016

REFERENCES

Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. Proc. Natl. Acad. Sci. USA *79*, 1129–1133.

Ahsendorf, T., Wong, F., Eils, R., and Gunawardena, J. (2014). A framework for modelling gene regulation which accommodates non-equilibrium mechanisms. BMC Biol. *12*, 102.

Bialek, W. (2015). Perspectives on theory at the interface of physics and biology. arXiv, arXiv:1512.08954.

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005). Transcriptional regulation by the numbers: models. Curr. Opin. Genet. Dev. *15*, 116–124.

Borggrefe, T., and Yue, X. (2011). Interactions between subunits of the Mediator complex with gene-specific transcription factors. Semin. Cell Dev. Biol. *22*, 759–768.

Bosveld, F., van Hoek, S., and Sibon, O.C.M. (2008). Establishment of cell fate during early Drosophila embryogenesis requires transcriptional Mediator subunit dMED31. Dev. Biol. *313*, 802–813.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25–36.

Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat. Rev. Genet. *14*, 572–584.

Davidson, E.H. (2006). The Regulatory Genome: Gene Regulatory Networks in Development and Evolution (Academic Press).

Engel, P.C. (2013). A hundred years of the Hill equation. Biochem. J. *2013*, 1–4.

Gregor, T., Tank, D.W., Wieschaus, E.F., and Bialek, W. (2007). Probing the limits to positional information. Cell *130*, 153–164.

Gunawardena, J. (2013). Biology is more theoretical than physics. Mol. Biol. Cell *24*, 1827–1829.

Gunawardena, J. (2014). Models in biology: 'accurate descriptions of our pathetic thinking'. BMC Biol. *12*, 29.

Hammar, P., Walldén, M., Fange, D., Persson, F., Baltekin, O., Ullman, G., Leroy, P., and Elf, J. (2014). Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. Nat. Genet. *46*, 405–408.

Hill, A.V. (1913). The combinations of haemoglobin with oxygen and with carbon monoxide. I. Biochem. J. *7*, 471–480.

Hopfield, J.J. (1974). Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. Proc. Natl. Acad. Sci. USA *71*, 4135–4139.

Keung, A.J., Joung, J.K., Khalil, A.S., and Collins, J.J. (2015). Chromatin regulation at the frontier of synthetic biology. Nat. Rev. Genet. *16*, 159–171.

Lebrecht, D., Foehr, M., Smith, E., Lopes, F.J., Vanario-Alonso, C.E., Reinitz, J., Burz, D.S., and Hanes, S.D. (2005). Bicoid cooperative DNA binding is critical for embryonic patterning in Drosophila. Proc. Natl. Acad. Sci. USA *102*, 13176–13181.

Mahan, B.H. (1975). Microscopic reversibility and detailed balance. J. Chem. Educ. *52*, 299–302.

Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. Proc. Natl. Acad. Sci. USA *107*, 22534–22539.

Monod, J., Wyman, J., and Changeux, J.P. (1965). On the nature of allosteric transitions: a plausible model. J. Mol. Biol. *12*, 88–118.

Ninio, J. (1975). Kinetic amplification of enzyme discrimination. Biochimie *57*, 587–595.

Nussinov, R., Tsai, C.-J., and Ma, B. (2013). The underappreciated role of allostery in the cellular network. Annu. Rev. Biophys. *42*, 169–189.

Park, J.M., Gim, B.S., Kim, J.M., Yoon, J.H., Kim, H.-S., Kang, J.-G., and Kim, Y.-J. (2001). Drosophila Mediator complex is broadly utilized by diverse gene-specific transcription factors at different types of core promoters. Mol. Cell. Biol. *21*, 2312–2323.

Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc. Natl. Acad. Sci. USA *108*, 13570–13575.

Perry, M.W., Bothma, J.P., Luu, R.D., and Levine, M. (2012). Precision of hunchback expression in the Drosophila embryo. Curr. Biol. *22*, 2247–2252.

Phillips, R. (2015). Theory in biology: Figure 1 or Figure 7? Trends Cell Biol. *25*, 723–729.

Ptashne, M. (2004). A Genetic Switch: Phage Lambda Revisited, Third Edition (Cold Spring Harbor Laboratory Press).

Segal, E., and Widom, J. (2009). From DNA sequence to transcriptional behaviour: a quantitative approach. Nat. Rev. Genet. *10*, 443–456.

Sherman, M.S., and Cohen, B.A. (2012). Thermodynamic state ensemble models of cis-regulation. PLoS Comp. Biol. *8*, e1002407.

Singh, N., Zhu, W., and Hanes, S.D. (2005). Sap18 is required for the maternal gene bicoid to direct anterior patterning in Drosophila melanogaster. Dev. Biol. *278*, 242–254.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Takahashi, K., and Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. Nat. Rev. Mol. Cell Biol. *17*, 183–193.

Voss, T.C., Schiltz, R.L., Sung, M.-H., Yen, P.M., Stamatoyannopoulos, J.A., Biddie, S.C., Johnson, T.A., Miranda, T.B., John, S., and Hager, G.L. (2011). Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. Cell *146*, 544–554.

Wang, F., Marshall, C.B., and Ikura, M. (2013). Transcriptional/epigenetic regulator CBP/p300 in tumorigenesis: structural and functional versatility in target recognition. Cell. Mol. Life Sci. *70*, 3989–4008.

Weiss, J.N. (1997). The Hill equation revisited: uses and misuses. FASEB J. *11*, 835–841.

Wunderlich, Z., and Mirny, L.A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. Trends Genet. *25*, 434–440.