An allosteric theory of transcription factor induction

Manuel Razo-Mejia[1,†], **Stephanie L. Barnes**[1,†], **Nathan M. Belliveau**[1,†], **Griffin Chure**[1,†], **Tal Einav**[2,†], **Rob Phillips**[1,3,*]
[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, United States; [2]Department of Physics, California Institute of Technology, Pasadena, United States;
[3]Department of Applied Physics, California Institute of Technology, Pasadena, United States
[†] contributed equally
* phillips@pboc.caltech.edu

Abstract

Allosteric molecules serve as regulators of cellular activity across all domains of life. We present a general theory of allosteric transcriptional regulation that permits quantitative predictions for how physiological responses are tuned to environmental stimuli. To test the model's predictive power, we apply it to the specific case of the ubiquitous simple repression motif in bacteria. We measure the fold-change in gene expression at different inducer concentrations in a collection of strains that span a range of repressor copy numbers and operator binding strengths. After inferring the inducer dissociation constants using data from one of these strains, we show the broad reach of the model by predicting the induction profiles of all other strains. Finally, we derive an expression for the free energy of allosteric transcription factors which enables us to collapse the data from all of our experiments onto a single master curve, capturing the diverse phenomenology of the induction profiles.

Introduction

The interaction of organisms with their environment is a central theme in biology. At the cellular level, this interaction is mediated by a diverse collection of molecular signaling pathways. One of the most pervasive mechanisms of cell signaling is allosteric regulation, in which the binding of a ligand induces a conformational change in some target molecule subsequently triggering a signaling cascade. An important example of such signaling is offered by transcriptional regulation, in which many transcription factors are known to alter their propensity to bind DNA depending upon whether or not they are bound to a ligand. Despite the overarching importance of this mode of signaling, our understanding of the relationship between extracellular inputs and gene expression outputs is severely limited due to the lack of a rigorous dialogue between predictive theories of the induction phenomenon and experiments [1]. To that end, we present a theory of allosteric transcriptional regulation and explore it in the context of the simple repression motif – a widespread bacterial genetic regulatory architecture in which binding of a transcription factor occludes binding of an RNA polymerase and initiation of transcription. A recent survey of the frequency of usage of different regulatory architectures within the *E. coli* genome revealed that more than 100 genes are characterized by the simple repression motif, making it a common and physiologically important architecture [2].

Here we use a statistical mechanical rendering of allostery as the basis of parameter-free predictions and corresponding measurements of the induction phenomenon. Specifically, we model the allosteric response of transcriptional repressors using the Monod-Wyman-Changeux (MWC) model, which stipulates that an allosteric protein fluctuates between two distinct conformations – an active and inactive state – in thermodynamic equilibrium [3]. In the context of an inducible repressor, ligand binding increases the probability that the repressor will be in the inactive state, weakening its ability to bind to the promoter and resulting in increased expression. The theory presented here provides considerable insight beyond that of simply fitting a sigmoidal curve to inducer titration data. We aim to explain and predict the fine details of an induction profile, such as the steepness of its response as well as the limits of minimum and maximum expression. By combining this MWC framework of induction with a thermodynamic model of transcriptional regulation, we create a general quantitative model of allosteric induction that is applicable to broad classes of transcription factors.

To demonstrate the predictive power of this theory, we design a genetic circuit in which the binding probability of an allosteric repressor regulates gene expression. We first quantify the set of parameters associated with the induction of the repressor, namely, the binding affinity of the repressor to the inducer and the energy difference between the active and inactive repressor states. We then determine these parameters by fitting to measurements of the fold-change in gene expression as a function of inducer concentration for a circuit with known repressor copy number and repressor-operator binding energy. With these fitted allosteric parameters in hand, we make accurate, parameter-free predictions of the induction response for many other combinations of repressor copy number and binding energy. To this end, our genetic circuit uses components from the well-characterized *lac* system in *E. coli*.

The agreement between theory and experiment allows us to demonstrate how important features such as a circuit's leakiness and dynamic range can be understood using the MWC framework. While we show that a single model can successfully predict the behavior of circuits with a broad range of parameter values, a naive interpretation may suggest that each new set of control parameters gives rise to some significantly different input-output response. However, our model allows us to derive an expression for the free energy of the system, a nonlinear combination of physical parameters which dictates the system's mean response. We demonstrate that all of our data sets collapse onto a single master curve in which this free energy acts as the independent variable. The power of this approach is that it calls attention to the free energy as the actual tuning parameter that controls the mean gene expression response, showing how different system responses may all be understood as embodiments of a single unified theory.

Results

Characterizing Transcription Factor Induction using the Monod-Wyman-Changeux (MWC) Model

We begin by considering the induction of a simple repression genetic architecture, in which the binding of a transcriptional repressor occludes the binding of RNA polymerase (RNAP) to the DNA. When an inducer binds to the repressor, it shifts the repressor's allosteric equilibrium towards the inactive state as specified by the MWC model [3], causing the repressor to bind more weakly to the operator, thereby increasing gene expression. Simple repression motifs in the absence of inducer have been previously characterized by an equilibrium model where the probability of each state of repressor and RNAP promoter occupancy is dictated by the Boltzmann distribution [1, 3, 4, 7, 7, 8, 8, 9] (we note that non-equilibrium models of simple repression have been shown to have the same functional form that we derive below [12]). To extend these models to account for allosteric induction [1, 3, 4, 12], we account for the equilibrium state of the repressor through the MWC model as follows.

Consider a cell with copy number P of RNAP and R repressors. Our model assumes that the repressor can exist in two conformational states. R_A repressors will be in the active state (the favored state when the repressor is not bound to an inducer; in this state the repressor binds tightly to the DNA) and the remaining R_I repressors will be in the inactive state (the predominant state when repressor is bound to an inducer; in this state the repressor binds weakly to the DNA) such that $R_A + R_I = R$. Repressors fluctuate between these two conformations in thermodynamic equilibrium [13]. The key parameters of this model are listed in Table 1.

Thermodynamic models of gene expression begin by enumerating all possible states of the promoter and their corresponding statistical weights. As shown in Fig. 1, the promoter can either be empty, occupied by RNAP, or occupied by either an active or inactive repressor. We assign the repressor a different DNA binding affinity in the active and inactive state. In addition to the specific binding sites at the promoter, we assume that there are N_{NS} non-specific binding sites elsewhere (i.e. on parts of the genome outside the simple repression architecture) where RNAP or the repressor can bind. All specific binding energies are measured relative to the average non-specific binding energy. Thus, $\Delta\varepsilon_P$ represents the energy difference between the specific and non-specific binding for RNAP to the DNA. Likewise, $\Delta\varepsilon_{RA}$ and $\Delta\varepsilon_{RI}$ represent the difference in specific and non-specific binding energies for repressor in the active or inactive state, respectively.

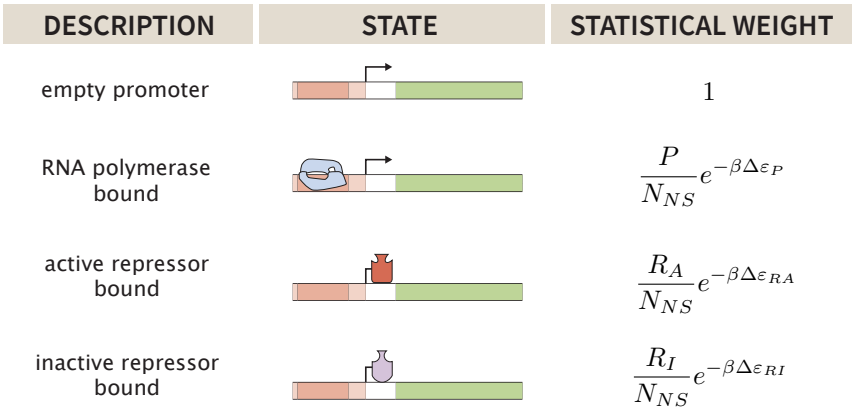

Figure 1. States and weights for the simple repression motif. RNAP (light blue) and a repressor compete for binding to the DNA. There are R_A repressors in the active state (red) and R_I repressors in the inactive state (purple). The difference in energy between a repressor bound to the operator and to another non-specific site on the DNA equals $\Delta\varepsilon_{RA}$ in the active state and $\Delta\varepsilon_{RI}$ in the inactive state; the P RNAP have a corresponding energy difference $\Delta\varepsilon_P$. N_{NS} represents the number of non-specific binding sites where RNAP and repressor can also bind.

Thermodynamic models of transcription [7–9] posit that gene expression is proportional to the probability p_{bound} that the RNAP is bound to its specific binding site, which is given by

$$p_{\text{bound}} = \frac{\frac{P}{N_{NS}} e^{-\beta \Delta\varepsilon_P}}{1 + \frac{R_A}{N_{NS}} e^{-\beta \Delta\varepsilon_{RA}} + \frac{R_I}{N_{NS}} e^{-\beta \Delta\varepsilon_{RI}} + \frac{P}{N_{NS}} e^{-\beta \Delta\varepsilon_P}}, \tag{1}$$

with $\beta = \frac{1}{k_B T}$ where k_B is the Boltzmann constant and T is the temperature of the system. Measuring p_{bound} directly is fraught with experimental difficulties [16]. Instead, we measure the fold-change in gene expression due the presence of the repressor. We define fold-change as the ratio of gene expression in the presence of repressor relative to expression in the absence of repressor, namely,

$$\text{fold-change} \equiv \frac{p_{\text{bound}}(R > 0)}{p_{\text{bound}}(R = 0)}. \tag{2}$$

We can simplify this expression using two well-justified approximations: (1) $\frac{P}{N_{NS}} e^{-\beta \Delta\varepsilon_P} \ll 1$ implying that the RNAP binds weakly to the promoter ($N_{NS} = 4.6 \times 10^6$, $P \approx 10^3$ [15] , $\Delta\varepsilon_P \approx -2$ to $-5\ k_B T$ [16], so that $\frac{P}{N_{NS}} e^{-\beta \Delta\varepsilon_P} \approx 10^{-5}$) and (2) $\frac{R_I}{N_{NS}} e^{-\beta \Delta\varepsilon_{RI}} \ll 1 + \frac{R_A}{N_{NS}} e^{-\beta \Delta\varepsilon_{RA}}$ which reflects our assumption that the inactive repressor binds weakly to the DNA. Using these approximations, the fold-change reduces to the form

$$\text{fold-change} \approx \left(1 + \frac{R_A}{N_{NS}} e^{-\beta \Delta\varepsilon_{RA}}\right)^{-1} \equiv \left(1 + p_A(c) \frac{R}{N_{NS}} e^{-\beta \Delta\varepsilon_{RA}}\right)^{-1}, \tag{3}$$

where in the last step we have introduced the fraction $p_A(c)$ of repressors in the active state given a concentration c of inducer, which is defined as $R_A(c) = p_A(c) R$. Since inducer binding shifts the repressors from the active to the inactive state, $p_A(c)$ is a decreasing function of c [15].

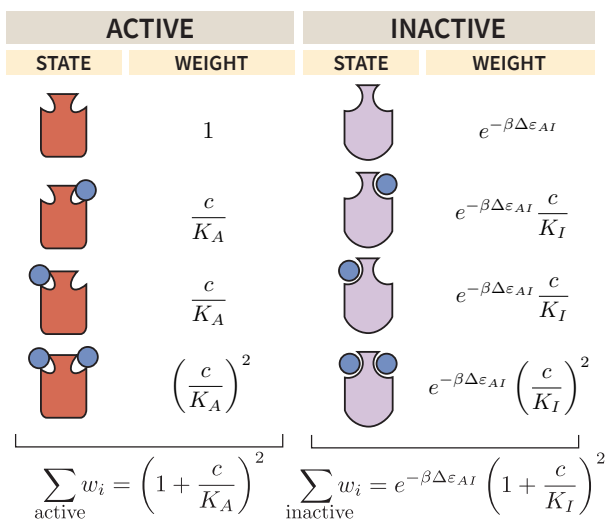

Figure 2. States and weights of an allosteric repressor with $n = 2$ inducer binding sites. A repressor has an active conformation (red, left column) and inactive conformation (purple, right column), with the energy difference between these two states given by $\Delta\varepsilon_{AI}$. The inducer (blue circle) at concentration c is capable of binding to the repressor with dissociation constants K_A in the active state and K_I in the inactive state. The eight states for a dimer with $n = 2$ inducer binding sites are shown.

We compute the probability $p_A(c)$ that a repressor with n inducer binding sites will be active using the MWC model [3]. After first enumerating all possible configurations of a repressor bound to inducer (see Fig. 2), $p_A(c)$ is given by the sum of the weights of the active states divided by the sum of the

Table 1. Key model parameters for induction of an allosteric repressor.

Parameter	Description
c	Concentration of the inducer
K_A, K_I	Dissociation constant between an inducer and the repressor in the active/inactive state
$\Delta\varepsilon_{AI}$	The difference between the free energy of repressor in the inactive and active states
$\Delta\varepsilon_P$	Binding energy between the RNAP and its specific binding site
$\Delta\varepsilon_{RA}, \Delta\epsilon_{RI}$	Binding energy between the operator and the active/inactive repressor
n	Number of inducer binding sites per repressor
P	Number of RNAP
R_A, R_I, R	Number of active/inactive/total repressors
$p_A = \frac{R_A}{R}$	Probability that a repressor will be in the active state
p_{bound}	Probability that an RNAP is bound to the promoter of interest, assumed to be proportional to gene expression
fold-change	Ratio of gene expression in the presence of repressor to that in the absence of repressor
F	Free energy of the system
N_{NS}	The number of non-specific binding sites for the repressor in the genome
$\beta = \frac{1}{k_B T}$	The inverse product of the Boltzmann constant k_B and the temperature T of the system

weights of all possible states, namely,

$$p_A(c) = \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{-\beta\Delta\varepsilon_{AI}}\left(1 + \frac{c}{K_I}\right)^n}, \tag{4}$$

where K_A and K_I represent the dissociation constant between the inducer and repressor in the active and inactive states, respectively, and $\Delta\varepsilon_{AI} = \varepsilon_I - \varepsilon_A$ denotes the free energy difference between a repressor in the active and inactive state. The specific case of a repressor dimer with $n = 2$ inducer binding sites is shown in Fig. 2. This assumes that the repressor is comprised of two allosterically independent dimers (see Appendix A for an alternate analysis which assumes the repressor is an allosteric tetramer). A repressor which favors the active state in the absence of inducer ($\Delta\varepsilon_{AI} > 0$) will be driven towards the inactive state upon inducer binding when $K_I < K_A$. The case $K_A < K_I$ represents a corepressor which binds to the repressor and stabilizes its active state.

Substituting $p_A(c)$ from Eq. (4) into Eq. (3) yields the general formula for induction of a simple repression regulatory architecture, namely,

$$\text{fold-change} = \left(1 + \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{-\beta\Delta\varepsilon_{AI}}\left(1 + \frac{c}{K_I}\right)^n}\frac{R}{N_{NS}}e^{-\beta\Delta\varepsilon_{RA}}\right)^{-1}. \tag{5}$$

This key formula enables us to make precise quantitative statements about induction profiles. Once the MWC parameters (K_A, K_I, and $\Delta\varepsilon_{AI}$) for the repressor and inducer have been determined, the response of the system is fully characterized and the input/output response under different conditions can be predicted. For example, this framework can predict the response of the system at different repressor copy numbers R, repressor-operator affinities $\Delta\varepsilon_{RA}$, or inducer concentrations c. Motivated by the broad range of predictions implied by Eq. (5), we designed a series of experiments to tune the control parameters for a simple repression genetic circuit in living cells.

Experimental Design

To exploit the expression for fold-change in Eq. (5) and make contact between theory and experiment, we used a set of simple repression *E. coli* constructs based on components of the *lac* system, which we have used in previous studies and were originally inspired by Oehler *et al.* [18]. We tune the parameters shown in Eq. (5), namely, the Lac repressor (LacI) copy number R and operator binding energy $\Delta\varepsilon_{RA}$. Here the repressor copy number R refers to the number of repressor dimers in the cell, rather than tetramers as reported in [8]. Through modification of the *lacI* ribosomal binding site, we generated strains with mean repressor copy numbers of $R = 22$, 60, 124, 260, 1220, and 1740. Gene expression was measured using a Yellow Fluorescent Protein (YFP) gene, driven by a *lacUV5* promoter. Each repressor copy number variant was paired with the native O1, O2, or O3 LacI operator placed at the YFP transcription start site, thereby generating eighteen unique strains [18]. The repressor-operator binding energies (O1 $\Delta\varepsilon_{RA} = -15.3\ k_BT$, O2 $\Delta\varepsilon_{RA} = -13.9\ k_BT$, and O3 $\Delta\varepsilon_{RA} = -9.7\ k_BT$) were previously inferred by measuring the fold-change of the *lac* system at different repressor copy numbers [8].

Each experimental strain thus uses a known repressor copy number R and binding energy $\Delta\varepsilon_{RA}$. We measure fold-change over a range of known IPTG concentrations c, using $n = 2$ inducer binding sites per LacI dimer and approximating the number of nonspecific binding sites as the length in base-pairs of the *E. coli* genome, $N_{NS} = 4.6 \times 10^6$. After accounting for each of these known parameter values, the fold-change in gene expression (Eq. (5)) has three unknown parameters: the free energy difference between the active and inactive states of the repressor ($\Delta\varepsilon_{AI}$) and the inducer dissociation constants for the repressor in the active state (K_A) and inactive state (K_I).

Our experimental pipeline for determining fold-change using flow cytometry is shown in Fig. 3. Briefly, cells were grown to exponential phase under concentrations of the inducer IPTG ranging between 0 and 5mM. We measure YFP fluorescence using flow cytometry and automatically gate the data to include only single-cell measurements (see Appendix B). The fold-change (Eq. (2)) is determined as the ratio of fluorescence in the presence and absence of repressor. In addition to using flow cytometry, we wanted to confirm that distinct methods for measuring gene expression yielded the same basic results and hence we undertook these same measurements for a subset of strains using single-cell microscopy (see Appendix C).

Determination of the *in vivo* MWC Parameters

The three parameters that we tune experimentally are shown in Fig. 4A, leaving the three allosteric parameters ($\Delta\varepsilon_{AI}$, K_A, and K_I) to be determined by fitting. Using previous *lac* repressor fold-change data, we infer that $\Delta\varepsilon_{AI} = 4.5\ k_BT$ (see Appendix D). Rather than fitting K_A and K_I to our entire data set of eighteen unique constructs, we performed Bayesian parameter estimation to data from a single strain with $R = 260$ and an O2 operator ($\Delta\varepsilon_{RA} = -13.9\ k_BT$ [8]) shown in Fig. 4C (white circles). Using Markov Chain Monte Carlo, we determine the most likely parameter values as $K_A = 139^{+29}_{-22} \times 10^{-6}$ M and $K_I = 0.53^{+0.04}_{-0.04} \times 10^{-6}$ M, which are the modes and the superscript and subscript values represent the upper and lower bounds of the 95th percentile of the parameter value distributions. Using these parameters, we then predicted the fold-change and the associated credible region for the remaining seventeen strains with no further fitting. These predictions are plotted in Fig. 4 for the different operators.

Comparison of Experimental Measurements with Theoretical Predictions

We tested the predictions shown in Fig. 4 by measuring the fold-change induction profiles for a broad range of different repressor copy numbers, repressor binding energies, and inducer concentrations. The results, shown in Fig. 5, demonstrate very good agreement between theory and experiment across all of our strains. We note that the apparently systematic shifts in all of the O3 strains and all of the $R = 1220$ and $R = 1740$ strains may be partially due to imprecise previous measurements of their $\Delta\varepsilon_{RA}$ and R values (see Appendix E, where we perform a global fit to all parameters including the repressor copy number R and the binding energy $\Delta\varepsilon_{RA}$, although a possible discrepancy in the steepness of the response for all O3 strains remains in spite of the global fit). We also repeated this experiment using the

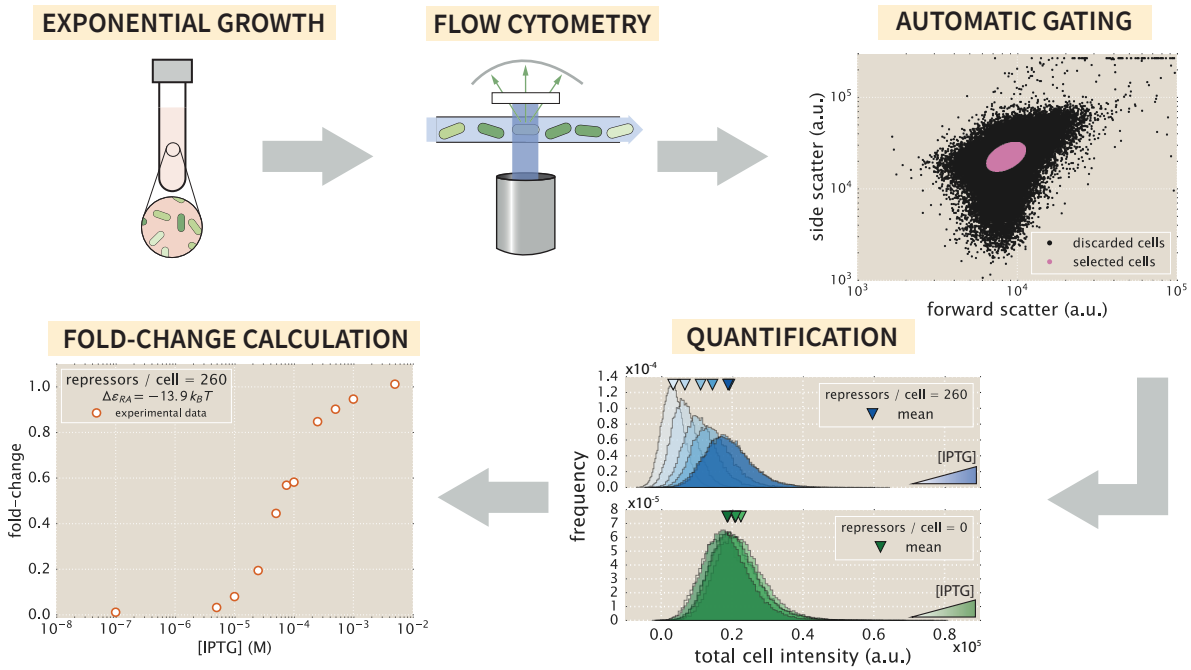

Figure 3. An experimental pipeline for high-throughput fold-change measurements. Cells are grown to exponential steady state and their fluorescence is measured using flow cytometry. Automatic gating methods using forward and side scattering are used to ensure that all measurements come from single cells (see Methods). Mean expression is then quantified at different IPTG concentrations (top, blue histograms) and for a strain without repressor (bottom, green histograms), which shows no response to IPTG as expected. Fold-change is computed by dividing the mean fluorescence in the presence of repressor by the mean fluorescence in the absence of repressor.

synthetic Oid operator which exhibits stronger repression than the O1, O2, and O3 operators. We note that there is some subtlety regarding measurements of fold-change for systems that exhibit such high repression, but our theory is consistent to within 1 k_BT to the binding energy previously reported (see Appendix F for more details).

One of the remarkable features of the predicted curves in Fig. 4 is that strains vary significantly in their values of fold-change in the absence of inducer, which we define as the leakiness, and the difference between the maximum and the minimum fold-change, which we define as the dynamic range. The MWC framework captured in Eq. (5) makes it possible to see how different MWC parameters give rise to these distinct responses (see Appendix G). We begin by analyzing the minimum fold-change which is observed in the absence of ligand,

$$\text{fold-change}(c=0) = \left(1 + \frac{1}{1+e^{-\beta\Delta\varepsilon_{AI}}}\frac{R}{N_{NS}}e^{-\beta\Delta\varepsilon_{RA}}\right)^{-1}, \tag{6}$$

and the maximum fold change which is observed in the presence of saturating ligand,

$$\text{fold-change}(c\to\infty) = \left(1 + \frac{1}{1+e^{-\beta\Delta\varepsilon_{AI}}\left(\frac{K_A}{K_I}\right)^n}\frac{R}{N_{NS}}e^{-\beta\Delta\varepsilon_{RA}}\right)^{-1}. \tag{7}$$

Eq. (6) describes the fold-change in gene expression when the system is maximally repressed. The difference between the maximum fold-change in Eq. (7) and minimal fold-change represents the contrast between the system's output in the presence and absence of ligand. In Appendix G, we discuss these properties in more detail, along with the midpoint and steepness of the response which are also critical physiological parameters that can be understood intuitively within the MWC framework.

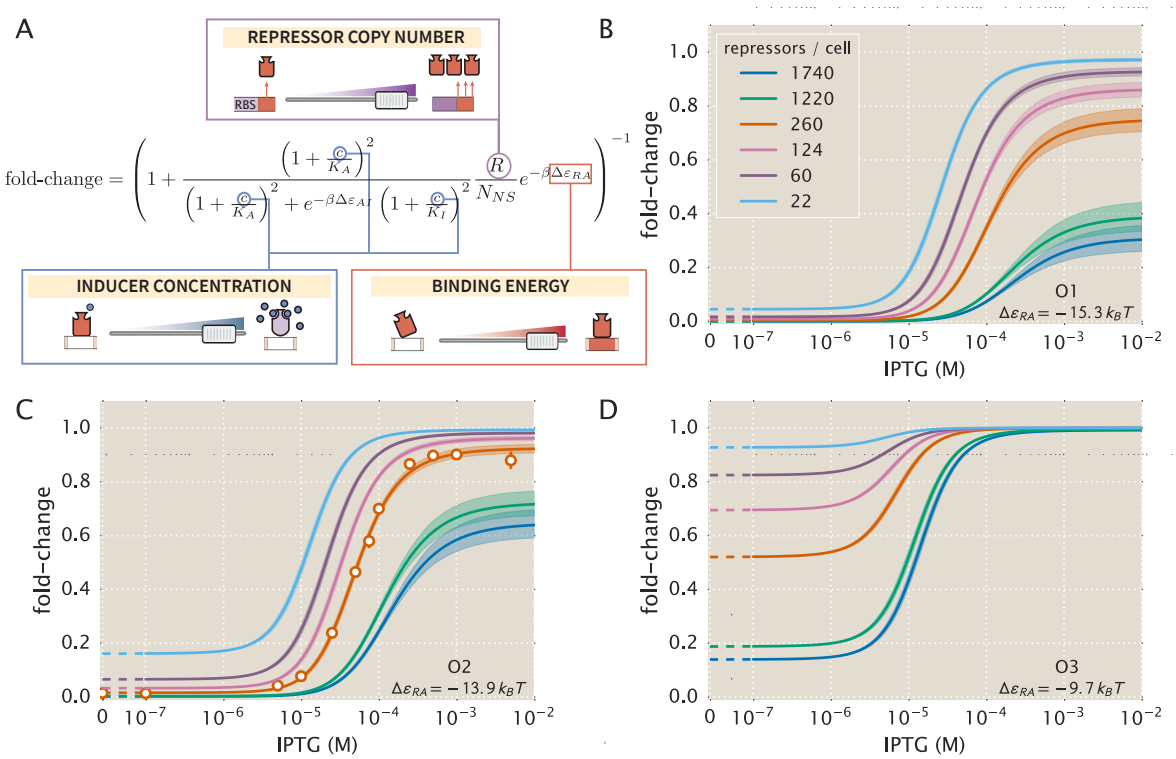

Figure 4. Predicting induction profiles for different biological control parameters. (A) We can quantitatively tune R via ribosomal binding site (RBS) modifications, $\Delta\varepsilon_{RA}$ by mutating the operator sequence, and c by adding different amounts of IPTG to the growth medium. (B-D) Predicted IPTG titration curves for different repressor copy numbers and operator strengths. Titration data for the O2 strain (white circles in C) with $R = 260$, $\Delta\varepsilon_{RA} = -13.9\ k_BT$, $n = 2$, and $\Delta\varepsilon_{AI} = 4.5\ k_BT$ can be used to determine the thermodynamic parameters $K_A = 139^{+29}_{-22} \times 10^{-6}$ M and $K_I = 0.53^{+0.04}_{-0.04} \times 10^{-6}$ M (orange line). The remaining solid lines predict the fold-change (Eq. (5)) for other repressor copy numbers. Error bars of experimental data show the standard error of the mean (eight or more replicates) and shaded regions denote the 95% credible region. Where the credible region is not visible, this indicates that the interval is in fact narrower than the line itself. To display the measured fold-change in the absence of inducer, we alter the scaling of the x-axis between 0 and 10^{-7} M to linear rather than logarithmic, as indicated by a dashed line.

To show how this model can be used to predict system behavior, we examine variations in system responses at the limits of $c = 0$ and $c \to \infty$. At $c = 0$ for strains with the weak O3 operator ($\Delta\varepsilon_{RA} = -9.7\ k_BT$), Eq. (6) can be approximated as fold-change $\approx (1 + 0.004R)^{-1}$, which predicts a broad range in leakiness values for the range of repressor copy numbers R we consider. For instance, a system with the O3 operator and $R = 22$ is predicted to have a leakiness value of ≈ 0.9, while increasing the repressor copy number to $R = 1740$ results in a leakiness value of ≈ 0.1. In contrast, for systems with the strong operator O1 ($\Delta\varepsilon_{RA} = -15.3\ k_BT$), Eq. (6) reduces to fold-change $\approx (1 + 0.95R)^{-1}$, which predicts a much narrower range in leakiness given by ≈ 0.05 for $R = 22$ and ≈ 0.001 for $R = 1740$.

We can similarly explore the fold-change at saturating inducer concentrations ($c \to \infty$). In this limit, nearly all of the repressors are bound to inducer and are thus inactivated. However, even when bound to inducer some fraction of the repressors will remain in the active state, and will thus be capable of repressing the genetic circuit. For the case of the weak operator O3, the saturation predicted by Eq. (7) is approximated by fold-change $\approx \left(1 + 5 \times 10^{-6} R\right)^{-1}$. Hence we predict a fold-change of ≈ 1 for both $R = 22$ and $R = 1740$, reflecting the extremely low probability of repressor binding when the number of

active repressors is very small and the operator binding energy is weak. In contrast, saturated systems with the strong O1 operator have fold-change $\approx \left(1 + 10^{-3}R\right)^{-1}$, which ranges from values of ≈ 0.98 at $R = 22$ to ≈ 0.36 at $R = 1740$. This demonstrates that systems which incorporate strong operator binding energies may still experience considerable repression levels at high repressor copy numbers. In both the $c = 0$ and $c \to \infty$ limits, the predictions discussed here are strongly supported by the data shown in Fig. 5.

We define the 95% credible region as the interval that contains 95% of the probability density of the fold-change predictions given our inferred parameter probability distributions (see Methods). An interesting aspect of our predictions is the width of the credible regions increases with repressor copy number and inducer concentration but decreases with the repressor-operator binding energy. Note that the fold-change (Eq. (5)) depends on the product of $\frac{R}{N_{NS}} e^{-\beta \Delta\varepsilon_{RA}}$ with the MWC parameters K_A, K_I, and $\Delta\varepsilon_{AI}$. Therefore, strains with small repressor copy numbers as well as the O3 strains with weak binding energies will necessarily suppress variation in the MWC parameters (see Appendix H).

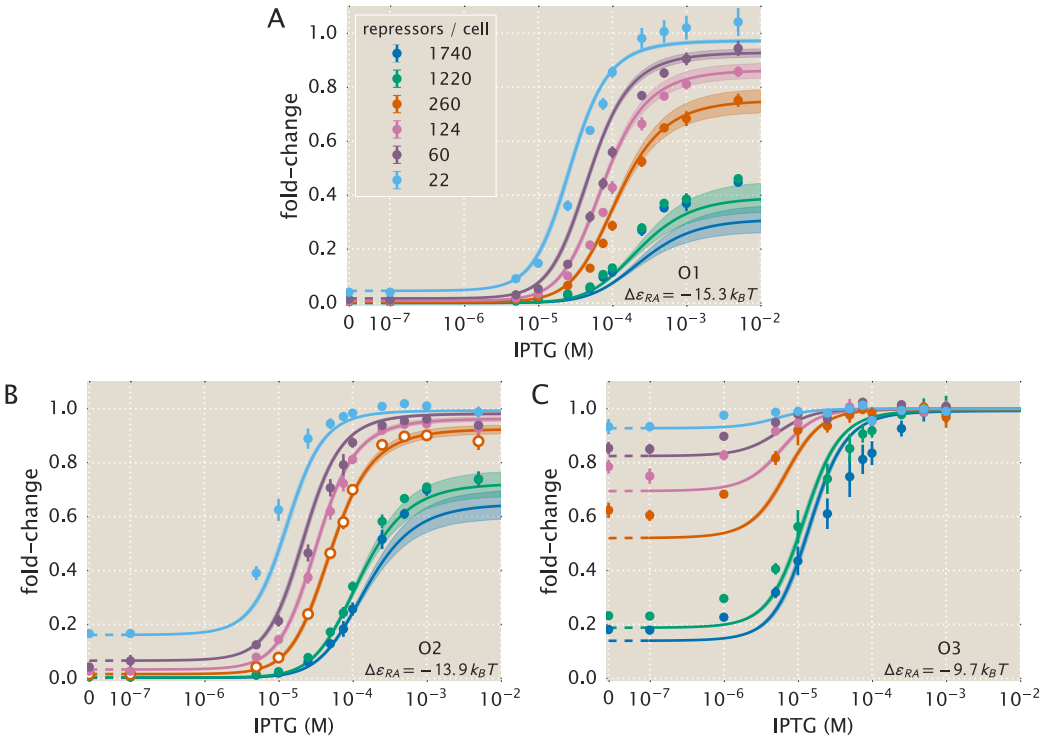

Figure 5. Parameter-free fold-change predictions versus experimental measurements using different operator binding energies and repressor copy numbers. The O2 strain with $R = 260$ (white circles) is used to predict the IPTG titration data for all other combinations of operator and repressor copy number. By changing the operator binding energy $\Delta\varepsilon_{RA}$, we can predict the titration curves for all (A) O1, (B) O2, and (C) O3 strains for a range of repressor copy numbers spanning two orders of magnitude. Error bars of experimental data show the standard error of the mean (eight or more replicates) and shaded regions denote the 95% credible region. Where the credible region is not visible, this indicates that the interval is in fact narrower than the line itself. The dashed lines ranging from 0 to 10^{-7} M IPTG indicates a linear scale, whereas solid lines represent a log scale.

We note that rather than using data from the strain with $R = 260$ repressors and an O2 operator, we could have used any of the strains to estimate K_A and K_I, and we consider all such possibilities in Appendix I. We also performed a global fit using the data from all eighteen strains for the following parameters: the LacI dissociation constants K_A and K_I, the repressor copy numbers R, and the repressor DNA binding energy $\Delta\varepsilon_{RA}$ (see Appendix E). These different approaches led to very similar

results, lending strong support for our quantitative understanding of induction in the simple repression architecture.

Data Collapse of Induction Profiles

An appealing aspect of our induction model is that it offers a unifying perspective with which to understand multiple seemingly different induction profiles. To demonstrate this, we rewrite Eq. (5) in the following form,

$$\text{fold-change} = \frac{1}{1 + e^{-\beta F(c)}}, \tag{8}$$

where $F(c)$ is the free energy of the system [12, 19, 20], which is given by

$$F(c) = -k_B T \left(\log \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{-\beta \Delta\varepsilon_{AI}} \left(1 + \frac{c}{K_I}\right)^n} + \log \frac{R}{N_{\text{NS}}} - \frac{\Delta\varepsilon_{RA}}{k_B T} \right). \tag{9}$$

The first term in parenthesis denotes the contribution from the inducer concentration, the second the effect of the repressor copy number, and the last the repressor-operator binding energy. Any combination of these parameters that yield the same free energy will, according to the theory described here, have the same fold-change, and therefore the same mean cellular response. Note that elsewhere, this free energy has been dubbed the Bohr parameter since such families of curves are analogous to the shifts in hemoglobin binding curves at different pHs known as the Bohr effect [12, 21, 22].

Fig. 6A demonstrates how the same mean physiological responses can arise from different parameter values. For example, suppose a system originally has a fold-change of 0.2 and operator mutations increase the $\Delta\varepsilon_{RA}$ binding energy. While this serves to initially increase both the free energy and the fold-change, a corresponding increase in the repressor copy number could bring the cell back to the original fold-change level. As the natural variable for induction, the free energy quantifies these trade-offs between the biological tuning parameters of the system. The 216 fold-change measurements from the eighteen strains collapse onto a single master curve as a function of the free energy as shown in Fig. 6B. The agreement across this diverse set of genetic constructs demonstrates how the free energy captures the underlying mechanism governing induction of the simple repression motif.

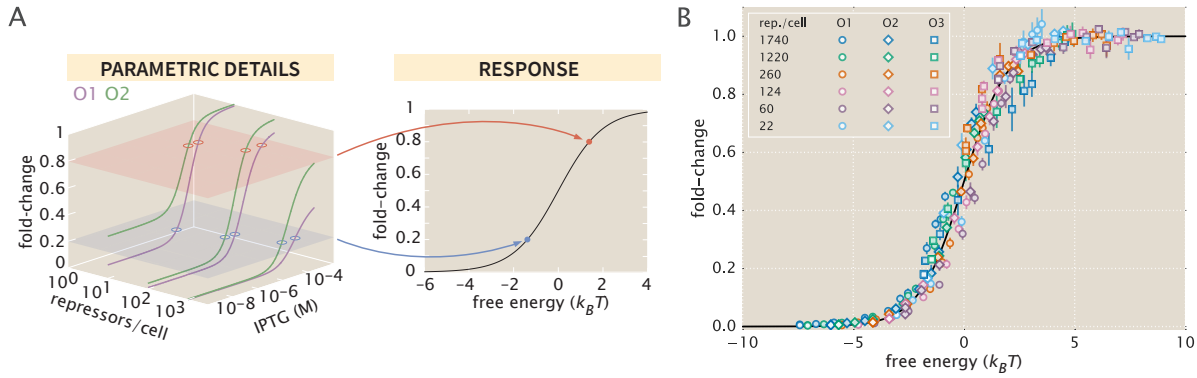

Figure 6. Fold-change data from a broad collection of different strains collapse onto a single master curve. (A) Any combination of parameters can be mapped to a single physiological response (i.e. fold-change) via the free energy, which encompasses the parametric details of the model. (B) Experimental data from Fig. 5 collapse onto a single master curve as a function of the free energy Eq. (9). All data points represent the mean and error bars are the standard error of the mean with eight or more replicates each.

Discussion

Since the early work by Monod, Wyman, and Changeux [3], a broad list of different biological phenomena have been tied to the existence of macromolecules that switch between inactive and active states. Examples can be found in a wide variety of cellular processes and include ligand-gated ion channels [23], enzymatic reactions [22, 24], chemotaxis [20], G-protein coupled receptors [25], transport proteins [15, 26, 27], and beyond. One of the most ubiquitous examples of allostery is in the context of gene expression, where an array of molecular players bind to transcription factors to either aid or deter their ability to regulate gene activity.

Others have developed quantitative models describing different aspects of allosteric regulatory systems. Martins and Swain analytically derived fundamental properties of the MWC model, including the leakiness and dynamic range described in this work, noting the inherent trade-offs in these properties when tuning the microscopic parameters of the model [14]. Recently, Daber *et al.* investigated how mutations alter the induction profiles of transcription factors [1]. Vilar and Saiz considered the broad range of interactions in inducible *lac*-based systems including the effects of oligomerization and DNA folding on transcription factor induction [29].

In this paper, we have built upon this previous work to show how the MWC model provides a clear statistical mechanical description of allosteric induction of transcriptional repressors. The MWC framework predicts gene expression as a function of inducer concentration which implicitly includes predictions about the leakiness, dynamic range, and both the concentration and slope at half-maximal induction (see Appendix G).

We tested our model in the context of a *lac*-based simple repression system. We first determined the allosteric dissociation constants K_A and K_I from a single data set (O2 operator with binding energy $\Delta\varepsilon_{RA}$ = -13.9 k_BT, and repressor copy number $R = 260$) and used these values to make parameter-free predictions of the induction profiles for seventeen other strains which we then measured using flow cytometry (see Fig. 4 and Fig. 5). These predictions consistently and accurately capture the primary features for each induction data set, most notably the leakiness and dynamic range of the response. For example, each of the data sets for O1, which has a strong binding energy, is very tightly repressed at low IPTG concentrations, but the fold-change values at saturation vary significantly with R. This trend is reversed in data sets for O3, which has a weak binding energy. Regardless of R, constructs using O3 approach fold-change values of 1 at high IPTG concentrations. O2, with its intermediate binding energy, shows neither strong repression nor full induction, but O2 constructs consistently have a large dynamic range. For both the O1 and O2 data sets, our model also accurately predicts the slopes of the curves at half-maximum, though the slope predictions for O3 are somewhat less accurate. This may be due to inaccuracies in previous determinations of the binding energies and repressor copy numbers, and in Appendix E we perform a global fit for the data together with the repressor copy numbers and DNA binding energies yielding better O3 agreement, in support of that hypothesis.

One of the demanding criteria of our approach is that a single small set of parameters must self consistently describe data from a very broad collection of regulatory situations and experimental methods. Specifically, we require self-consistency between a very diverse collection of data sets taken using distinct methods such as LacZ assays and bulk and single-cell fluorescence experiments to measure fold-change (see Appendices B and F), as well as quantitative Western blots [8] and binomial partitioning methods to count repressors [7, 30]. Furthermore, we demand that these same parameters account for architectures that are integrated into the chromosome, plasmid-borne, and even for cases where there are competing binding sites to take repressors out of circulation [7, 8] or where there are multiple operators to allow DNA looping [31]. Finally, these parameters are expected to hold for the entire range of inducer conditions described here. We are unaware of any comparable study that demands one predictive framework cover such a broad array of regulatory situations.

Despite the diversity observed in the induction profiles of each of our strains, our data are unified by their reliance on fundamental biophysical parameters. In particular, we have shown that our model for fold-change can be rewritten in terms of the free energy difference, which encompasses all of the physical parameters of the system. When our fold-change data are plotted against the respective free energies for each construct, they collapse cleanly onto a single curve (see Fig. 6). This enables us to analyze how

parameters can compensate each other; for example, given that typical repressor copy numbers found in bacteria range between 1 and 1000 [32], changing the repressor copy number from one extreme to the other could offset a repressor-operator binding energy difference $\Delta\varepsilon_{RA}$ of as much as 7 k_BT and still yield the same level of fold-change, suggesting potential different evolutionary paths to the same quantitative response.

While our experiments validated the theoretical predictions in the case of simple repression, we expect the framework presented here to apply much more generally to different biological instances of allosteric regulation. For example, we can use this model to explore different regulatory configurations such as corepression, activation, and coactivation, each of which are found in *E. coli*, and in Appendix J we state the fold-change equations for each of these cases. Additionally, by characterizing not just the mean but the full gene expression distribution, we can quantify the impact of noise on this system [33]. Another extension of this approach would be to theoretically predict and experimentally verify whether the repressor-inducer dissociation constants K_A and K_I or the energy difference $\Delta\varepsilon_{AI}$ between the allosteric states can be tuned by making single amino acid substitutions in the transcription factor [1, 12]. Finally, we expect that the kind of rigorous quantitative description of the allosteric phenomenon provided here will make it possible to construct biophysical models of fitness for allosteric proteins similar to those already invoked to explore the fitness effects of transcription factor binding site strengths and protein stability [34–36].

Methods

Bacterial Strains and DNA Constructs

All strains used in these experiments were derived from *E. coli* K12 MG1655 with the *lac* operon removed, adapted from those created and described in [6, 8]. Briefly, the operator variants and YFP reporter gene were cloned into a pZS25 background which contains a *lacUV5* promoter that drives expression as is shown in Fig. 1. These constructs carried a kanamycin resistance gene and were integrated into the *galK* locus of the chromosome using λ Red recombineering [17]. The *lacI* gene was constitutively expressed via a $P_{\text{LtetO-1}}$ promoter [18], with ribosomal binding site mutations made to vary the LacI copy number as described in [40] using site-directed mutagenesis (Quickchange II; Stratagene), with further details in [8]. These *lacI* constructs carried a chloramphenicol resistance gene and were integrated into the *ybcN* locus of the chromosome. Final strain construction was achieved by performing repeated P1 transduction [41] of the different operator and *lacI* constructs to generate each combination used in this work. Integration was confirmed by PCR amplification of the replaced chromosomal region and by sequencing. Primers and final strain genotypes are listed in Appendix K.

It is important to note that the rest of the *lac* operon (*lacZYA*) was never expressed. The LacY protein is a transmembrane protein which actively transports lactose as well as IPTG into the cell. As LacY was never produced in our strains, we assume that the extracellular and intracellular IPTG concentration was approximately equal due to diffusion across the membrane into the cell as is suggested by previous work [42].

To make this theory applicable to transcription factors with any number of DNA binding domains, we used a different definition for repressor copy number than has been used previously. We define the LacI copy number as the average number of repressor dimers per cell whereas in [8], the copy number is defined as the average number of repressor tetramers in each cell. To motivate this decision, we consider the fact that the LacI repressor molecule exists as a tetramer in *E. coli* [43] in which a single DNA binding domain is formed from dimerization of LacI proteins, so that wild-type LacI might be described as dimer of dimers. We have assumed that each DNA binding domain of the LacI repressor complex is allosterically independent (see Appendix A). This means that a single LacI tetramer can be treated as two functional repressors. Therefore, we have simply multiplied the number of repressors reported in [8] by a factor of two. This factor is included as a keyword argument in the numerous Python functions used to perform this analysis, as discussed in the code documentation.

A subset of strains in these experiments were measured using fluorescence microscopy for validation of the flow cytometry data and results. To aid in the high-fidelity segmentation of individual cells, the strains were modified to constitutively express an mCherry fluorophore. This reporter was cloned into a pZS4*1 backbone [18] in which mCherry is driven by the *lacUV5* promoter. All microscopy and flow cytometry experiments were performed using these strains.

Growth Conditions for Flow Cytometry Measurements

All measurements were performed with *E. coli* cells grown to mid-exponential phase in standard M9 minimal media (M9 5X Salts, Sigma-Aldrich M6030; 2 mM magnesium sulfate, Mallinckrodt Chemicals 6066-04; 100 μM calcium chloride, Fisher Chemicals C79-500) supplemented with 0.5% (w/v) glucose. Briefly, 500 μL cultures of *E. coli* were inoculated into Lysogeny Broth (LB Miller Powder, BD Medical) from a 50% glycerol frozen stock (-80°C) and were grown overnight in a 2 mL 96-deep-well plate sealed with a breathable nylon cover (Lab Pak - Nitex Nylon, Sefar America Inc. Cat. No. 241205) with rapid agitation for proper aeration. After approximately 12 to 15 hours, the cultures had reached saturation and were diluted 1000-fold into a second 2mL 96-deep-well plate where each well contained 500μL of M9 minimal media supplemented with 0.5% w/v glucose (anhydrous D-Glucose, Macron Chemicals) and the appropriate concentration of IPTG (Isopropyl β-D-1 thiogalactopyranoside Dioxane Free, Research Products International). These were sealed with a breathable cover and were allowed to grow for approximately eight hours. Cells were then diluted ten-fold into a round-bottom 96-well plate (Corning Cat. No. 3365) containing 90 μL of M9 minimal media supplemented with 0.5% w/v glucose along with

the corresponding IPTG concentrations. For each IPTG concentration, a stock of 100-fold concentrated IPTG in double distilled water was prepared and partitioned into 100 μL aliquots. The same parent stock was used for all experiments described in this work.

Flow Cytometry

Unless explicitly mentioned, all fold-change measurements were collected on a Miltenyi Biotec MACSquant Analyzer 10 Flow Cytometer graciously provided by the Pamela Björkman lab at Caltech. Detailed information regarding the voltage settings of the photo-multiplier detectors can be found in Appendix Table S1. Prior to each days experiments, the analyzer was calibrated using MACSQuant Calibration Beads (Cat. No. 130-093-607) such that day-to-day experiments would be comparable. All YFP fluorescence measurements were collected via 488 nm laser excitation coupled with a 525/50 nm emission filter. Unless otherwise specified, all measurements were taken over the course of two to three hours using automated sampling from a 96-well plate kept at approximately 4° - 10°C on a MACS Chill 96 Rack (Cat. No. 130-094-459). Cells were diluted to a final concentration of approximately 4×10^4 cells per μL which corresponded to a flow rate of 2,000 - 6,000 measurements per second. Once completed, the data were extracted and immediately processed using the following methods.

Unsupervised Gating of Flow Cytometry Data

Flow cytometry data will frequently include a number of spurious events or other undesirable data points such as cell doublets and debris. The process of restricting the collected data set to those data determined to be "real" is commonly referred to as gating. These gates are typically drawn manually [44] and restrict the data set to those points which display a high degree of linear correlation between their forward-scatter (FSC) and side-scatter (SSC). The development of unbiased and unsupervised methods of drawing these gates is an active area of research [45, 46]. For our purposes, we assume that the fluorescence level of the population should be normally distributed about some mean value. With this assumption in place, we developed a method that allows us to restrict the data used to compute the mean fluorescence intensity of the population to the smallest two-dimensional region of the log(FSC) vs. log(SSC) space in which 40% of the data is found. This was performed by fitting a bivariate Gaussian distribution and restricting the data used for calculation to those that reside within the 40th percentile. This procedure is described in more detail in the supplementary information as well as in a Jupyter notebook located in this paper's Github repository.

Experimental Determination of Fold-Change

For each strain and IPTG concentration, the fold-change in gene expression was calculated by taking the ratio of the population mean YFP expression in the presence of LacI repressor to that of the population mean in the absence of LacI repressor. However, the measured fluorescence intensity of each cell also includes the autofluorescence contributed by the weak excitation of the myriad protein and small molecules within the cell. To correct for this background, we computed the fold change as

$$\text{fold-change} = \frac{\langle I_{R>0} \rangle - \langle I_{\text{auto}} \rangle}{\langle I_{R=0} \rangle - \langle I_{\text{auto}} \rangle}, \tag{10}$$

where $\langle I_{R>0} \rangle$ is the average cell YFP intensity in the presence of repressor, $\langle I_{R=0} \rangle$ is the average cell YFP intensity in the absence of repressor, and $\langle I_{\text{auto}} \rangle$ is the average cell autofluorescence intensity, as measured from cells that lack the *lac*-YFP construct.

Bayesian Parameter Estimation

In this work, we determine the the most-likely parameter values for the inducer dissociation constants K_A and K_I of the active and inactive state, respectively, using Bayesian methods. We compute the

probability distribution of the value of each parameter given the data D, which by Bayes' theorem is given by

$$P(K_A, K_I \mid D) = \frac{P(D \mid K_A, K_I) P(K_A, K_I)}{P(D)}, \tag{11}$$

where D is all the data composed of independent variables (repressor copy number R, repressor-DNA binding energy $\Delta\varepsilon_{RA}$, and inducer concentration c) and one dependent variable (experimental fold-change). $P(D \mid K_A, K_I)$ is the likelihood of having observed the data given the parameter values for the dissociation constants, $P(K_A, K_I)$ contains all the prior information on these parameters, and $P(D)$ serves as a normalization constant, which we can ignore in our parameter estimation. Eq. (5) assumes a deterministic relationship between the parameters and the data, so in order to construct a probabilistic relationship as required by Eq. (11), we assume that the experimental fold-change for the ith datum given the parameters is of the form

$$\text{fold-change}^{(i)}_{\text{exp}} = \left(1 + \frac{\left(1 + \frac{c^{(i)}}{K_A}\right)^2}{\left(1 + \frac{c^{(i)}}{K_A}\right)^2 + e^{-\beta \Delta\varepsilon_{AI}} \left(1 + \frac{c^{(i)}}{K_I}\right)^2} \frac{R^{(i)}}{N_{NS}} e^{-\beta \Delta\varepsilon^{(i)}_{RA}} \right)^{-1} + \epsilon^{(i)}, \tag{12}$$

where $\epsilon^{(i)}$ represents the departure from the deterministic theoretical prediction for the ith data point. If we assume that these $\epsilon^{(i)}$ errors are normally distributed with mean zero and standard deviation σ, the likelihood of the data given the parameters is of the form

$$P(D|K_A, K_I, \sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \prod_{i=1}^{n} \exp\left[-\frac{(\text{fold-change}^{(i)}_{\text{exp}} - \text{fold-change}(K_A, K_I, R^{(i)}, \Delta\varepsilon^{(i)}_{RA}, c^{(i)}))^2}{2\sigma^2} \right], \tag{13}$$

where $\text{fold-change}^{(i)}_{\text{exp}}$ is the experimental fold-change and $\text{fold-change}(\cdots)$ is the theoretical prediction. The product $\prod_{i=1}^{n}$ captures the assumption that the n data points are independent. Note that the likelihood and prior terms now includes the extra unknown parameter σ. In applying Eq. (13), a choice of K_A and K_I that provides better agreement between theoretical fold-change predictions and experimental measurements will result in a more probable likelihood.

Both mathematically and numerically, it is convenient to define $\tilde{k}_A = -\log \frac{K_A}{1\,\text{M}}$ and $\tilde{k}_I = -\log \frac{K_I}{1\,\text{M}}$ and fit for these parameters on a log scale. Dissociation constants are scale invariant, so that a change from $10\,\mu$M to $1\,\mu$M leads to an equivalent increase in affinity as a change from $1\,\mu$M to $0.1\,\mu$M. With these definitions we assume for the prior $P(\tilde{k}_A, \tilde{k}_I, \sigma)$ that all three parameters are independent. In addition, we assume a uniform distribution for $\tilde{k}_A$ and $\tilde{k}_I$ and a Jeffreys prior [13] for the scale parameter σ. This yields the complete prior

$$P(\tilde{k}_A, \tilde{k}_I, \sigma) \equiv \frac{1}{(\tilde{k}_A^{\max} - \tilde{k}_A^{\min})} \frac{1}{(\tilde{k}_I^{\max} - \tilde{k}_I^{\min})} \frac{1}{\sigma}. \tag{14}$$

These priors are maximally uninformative meaning that they imply no prior knowledge of the parameter values. We defined the $\tilde{k}_A$ and $\tilde{k}_A$ ranges uniform on the range of -7 to 7, although we note that this particular choice does not affect the outcome provided the chosen range is sufficiently wide.

Putting all these terms together we can now sample from $P(\tilde{k}_A, \tilde{k}_I, \sigma \mid D)$ using Markov chain Monte Carlo (see GitHub repository) to compute the most likely parameter as well as the error bars (given by the 95% credible region) for K_A and K_I.

Data Curation

All of the data used in this work as well as all relevant code can be found at this dedicated website. Data was collected, stored, and preserved using the Git version control software in combination with off-site storage and hosting website GitHub. Code used to generate all figures and complete all processing step as and analyses are available on the GitHub repository. Many analysis files are stored as instructive Jupyter Notebooks. The scientific community is invited to fork our repositories and open constructive issues on the GitHub repository.

Acknowledgements

This work has been a wonderful exercise in scientific collaboration. We thank Hernan Garcia for information and advice for working with these bacterial strains, Pamela Björkman and Rachel Galimidi for access and training for use of the Miltenyi Biotec MACSQuant flow cytometer, and Colin deBakker of Milteny Biotec for useful advice and instruction in flow cytometry. The experimental front of this work began at the Physiology summer course at the Marine Biological Laboratory in Woods Hole, MA operated by the University of Chicago. We thank Simon Alamos, Nalin Ratnayeke, and Shane McInally for their work on the project during the course. We also thank Suzannah Beeler, Justin Bois, Robert Brewster, Soichi Hirokawa, Jané Kondev, Heun Jin Lee, Mitch Lewis, Muir Morrison, and Julie Theriot for useful advice and discussion. This work was supported by La Fondation Pierre-Gilles de Gennes, the Rosen Center at Caltech, and the National Institutes of Health DP1 OD000217 (Director's Pioneer Award), R01 GM085286, and 1R35 GM118043-01 (MIRA). Nathan Belliveau is a Howard Hughes Medical Institute International Student Research fellow.

Competing interests

The authors have declared that no competing interests exist.

References

1. Lindsley JE, Rutter J. 2006. Whence cometh the allosterome? *Proceedings of the National Academy of Sciences of USA* **103**: 10533–5.

2. Rydenfelt M, Garcia HG, Cox RS, Phillips R. 2014. The influence of promoter architectures and regulatory motifs on gene expression in *Escherichia coli*. *PLoS ONE* **9**: 1–31.

3. Monod J, Wyman J, Changeux JP. 1965. On the Nature of Allosteric Transitions: A Plausible Model. *Journal of molecular biology* **12**: 88–118.

4. Daber R, Sochor MA, Lewis M. 2011. Thermodynamic analysis of mutant lac repressors. *Journal of Molecular Biology* **409**: 76–87.

5. Kuhlman T, Zhang Z, Saier MH, Hwa T. 2007. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences of USA* **104**: 6043–8.

6. Setty Y, Mayo AE, Surette MG, Alon U. 2003. Detailed map of a cis-regulatory input function. *Proceedings of the National Academy of Sciences* **100**: 7702–7707.

7. Buchler NE, Gerland U, Hwa T. 2003. On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences of USA* **100**: 5136–41.

8. Vilar JM, Leibler S. 2003. DNA Looping and Physical Constraints on Transcription Regulation. *Journal of Molecular Biology* **331**: 981–989.

9. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Kuhlman T, Phillips R. 2005. Transcriptional regulation by the numbers: applications. *Current Opinion in Genetics & Development* **15**: 125–135.

10. Garcia HG, Phillips R. 2011. Quantitative dissection of the simple repression input-output function. *Proceedings of the National Academy of Sciences of USA* **108**: 12173–8.

11. Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, Phillips R. 2014. The transcription factor titration effect dictates level of gene expression. *Cell* **156**: 1312–1323.

12. Phillips R. 2015. Napoleon Is in Equilibrium. *Annual Review of Condensed Matter Physics* **6**: 85–111.

13. Kern D, Zuiderweg ER. 2003. The role of dynamics in allosteric regulation. *Current Opinion in Structural Biology* **13**: 748–757.

14. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Phillips R. 2005. Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development* **15**: 116–124.

15. Klumpp S, Hwa T. 2008. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proceedings of the National Academy of Sciences of USA* **105**: 20245–50.

16. Brewster RC, Jones DL, Phillips R. 2012. Tuning Promoter Strength through RNA Polymerase Binding Site Design in *Escherichia coli*. *PLoS Computational Biology* 8.

17. Marzen S, Garcia HG, Phillips R. 2013. Statistical mechanics of Monod-Wyman-Changeux (MWC) models. *Journal of Molecular Biology* **425**: 1433–1460.

18. Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, Müller-Hill B. 1994. Quality and position of the three lac operators of *E. coli* define efficiency of repression. *The EMBO journal* **13**: 3348–3355.

19. Keymer JE, Endres RG, Skoge M, Meir Y, Wingreen NS. 2006. Chemosensing in *Escherichia coli*: two regimes of two-state receptors. *Proceedings of the National Academy of Sciences of USA* **103**: 1786–91.

20. Swem LR, Swem DL, Wingreen NS, Bassler BL. 2008. Deducing Receptor Signaling Parameters from In Vivo Analysis: LuxN/AI-1 Quorum Sensing in Vibrio harveyi. *Cell* **134**: 461–473.

21. Mirny LA. 2010. Nucleosome-mediated cooperativity between transcription factors. *Proceedings of the National Academy of Sciences of USA* **107**: 22534–9.

22. Einav T, Mazutis L, Phillips R. 2016. Statistical Mechanics of Allosteric Enzymes. *The Journal of Physical Chemistry B* .

23. Auerbach A. 2012. Thinking in Cycles: MWC is a Good Model for Acetylcholine Receptor-Channels. *The Journal of Physiology* **590**: 93–8.

24. Velyvis A, Yang YR, Schachman HK, Kay LE. 2007. A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase. *Proceedings of the National Academy of Sciences of USA* **104**: 8815–20.

25. Canals M, Lane JR, Wen A, Scammells PJ, Sexton PM, Christopoulos A. 2012. A Monod-Wyman-Changeux mechanism can explain G protein-coupled receptor (GPCR) allosteric modulation. *Journal of Biological Chemistry* **287**: 650–659.

26. Levantino M, Spilotros A, Cammarata M, Schirò G, Ardiccioni C, Vallone B, Brunori M, Cupane A. 2012. The Monod-Wyman-Changeux allosteric model accounts for the quaternary transition dynamics in wild type and a recombinant mutant human hemoglobin. *Proceedings of the National Academy of Sciences of USA* **109**: 14894–9.

27. Mohapatra L, Goode BL, Jelenkovic P, Phillips R, Kondev J. 2015. Design Principles of Length Control of Cytoskeletal Structures. *Annual Review of Biophysics* **45**: 85–116.

28. Martins BMC, Swain PS. 2011. Trade-Offs and constraints in allosteric sensing. *PLoS Computational Biology* **7**: 1–13.

29. Vilar JMG, Saiz L. 2013. Reliable Prediction of Complex Phenotypes from a Modular Design in Free Energy Space: An Extensive Exploration of the *lac* Operon. *ACS Synthetic Biology* **2**: 576–586.

30. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. 2005. Gene regulation at the single-cell level. *Science* **307**: 1962–1965.

31. Boedicker JQ, Garcia HG, Phillips R. 2013. Theoretical and Experimental Dissection of DNA Loop-Mediated Repression. *Physical Review Letters* **110**: 018101.

32. Li GW, Burkhardt D, Gross C, Weissman JS. 2014. Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell* **157**: 624–635.

33. Eldar A, Elowitz M. 2010. Functional roles for noise in genetic circuits. *Nature* **467**: 167–173.

34. Gerland U, Hwa T. 2002. On the Selection and Evolution of Regulatory DNA Motifs. *Journal of Molecular Evolution* **55**: 386–400.

35. Berg J, Willmann S, Lässig M. 2004. Adaptive evolution of transcription factor binding sites. *BMC Evolutionary Biology* **4**: 42.

36. Zeldovich KB, Shakhnovich EI. 2008. Understanding Protein Evolution: From Protein Physics to Darwinian Selection. *Annual Review of Physical Chemistry* **59**: 105–127.

37. Garcia HG, Lee HJ, Boedicker JQ, Phillips R. 2011. Comparison and Calibration of Different Reporters for Quantitative Analysis of Gene Expression. *Biophysical Journal* **101**: 535–544.

38. Sharan SK, Thomason LC, Kuznetsov SG, Court DL. 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols* **4**: 206—223.

39. Lutz R, Bujard H. 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic acids research* **25**: 1203–10.

40. Salis HM, Mirsky EA, Voigt CA. 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology* **27**: 946–950.

41. Thomason LC, Costantino N, Court DL. 2007. *E. coli* genome manipulation by P1 transduction. *Current protocols in molecular biology* **Chapter 1**: Unit 1.17–1.17.8.

42. Fernández-Castané A, Vine CE, Caminal G, López-Santín J. 2012. Evidencing the role of lactose permease in IPTG uptake by *Escherichia coli* in fed-batch high cell density cultures. *Journal of Biotechnology* **157**: 391–398.

43. Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, Brennan RG, Lu P. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**: 1247–54.

44. Maecker HT, Rinfret A, D'Souza P, Darden J, Roig E, Landry C, Hayes P, Birungi J, Anzala O, Garcia M, et al.. 2005. Standardization of cytokine flow cytometry assays. *BMC Immunology* **6**: 13.

45. Aghaeepour N, Finak G, The FlowCAP Consortium, The DREAM Consortium, Hoos H, Mosmann TR, Brinkman R, Gottardo R, Scheuermann RH. 2013. Critical assessment of automated flow cytometry data analysis techniques. *Nature Methods* **10**: 228–238.

46. Lo K, Brinkman RR, Gottardo R. 2008. Automated gating of flow cytometry data via robust model-based clustering. *Cytometry Part A* **73A**: 321–332.

47. Sivia D, Skilling J. 2006. Data analysis: a Bayesian tutorial. OUP Oxford.