

Supporting Information for Tuning transcriptional regulation through signaling: A predictive theory of allosteric induction

Manuel Razo-Mejia^{1,†}, Stephanie L. Barnes^{1,†}, Nathan M. Belliveau^{1,†}, Griffin Chure^{1,†}, Tal Einav^{2,†}, Mitchell Lewis³, Rob Phillips^{1,2,4,*}

¹Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, 91125, USA

²Department of Physics, California Institute of Technology, Pasadena, CA, 91125, USA

³Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104, USA

⁴Department of Applied Physics, California Institute of Technology, Pasadena, CA, 91125, USA

[†]Contributed equally

*Correspondence: phillips@pboc.caltech.edu

Contents

A	Induction of Simple Repression with Multiple Promoters or Competitor Sites	A3
A.1	Chemical Potential Formulation to Calculate Fold-Change	A3
A.2	Variable Repressor Copy Number (R) with Multiple Specific Binding Sites ($N_S > 1$) .	A4
A.3	Variable Number of Specific Binding Sites N_S with Fixed Repressor Copy Number (R)	A4
A.4	Competitor Binding Sites	A4
A.5	Properties of the Induction Response	A6
B	Single-Cell Microscopy	A9
B.1	Strains and Growth Conditions	A9
B.2	Imaging Procedure	A9
B.3	Image Processing	A10
B.3.1	Correcting Uneven Illumination	A10
B.3.2	Cell Segmentation	A10
B.3.3	Calculation of Fold-Change	A11
B.4	Parameter Estimation and Comparison	A12
C	Fold-Change Sensitivity Analysis	A14
D	Applicability of Theory to the Oid Operator Sequence	A16
E	Applications to Other Regulatory Architectures	A18
E.1	Corepression	A18
E.2	Activation	A18

A Induction of Simple Repression with Multiple Promoters or Competitor Sites

We made the choice to perform all of our experiments using strains in which a single copy of our simple repression construct had been integrated into the chromosome. This stands in contrast to the methods used by a number of other studies (Oehler et al. 1994; Setty et al. 2003; Oehler et al. 2006; Daber et al. 2009; Daber et al. 2011; Vilar and Saiz 2013; Shis et al. 2014; Sochor 2014), in which reporter constructs are placed on plasmid, meaning that the number of constructs in the cell is not precisely known. It is also common to express repressor on plasmid to boost its copy number, which results in an uncertain value for repressor copy number. Here we show that our treatment of the MWC model has broad predictive power beyond the single-promoter scenario we explore experimentally, and indeed can account for systems in which multiple promoters compete for the repressor of interest. Additionally, we demonstrate the importance of having precise control over these parameters, as they can have a significant effect on the induction profile.

A.1 Chemical Potential Formulation to Calculate Fold-Change

In this section, we discuss a simple repression construct which we generalize in two ways from the scenario discussed in the text. First, we will allow the repressor to bind to N_S identical specific promoters whose fold-change we are interested in measuring, with each promoter containing a single repressor binding site ($N_S = 1$ in the main text). Second, we consider N_C identical competitor sites which do not regulate the promoter of interest, but whose binding energies are substantially stronger than non-specific binding ($N_C = 0$ in the main text). As in the main text, we assume that the rest of the genome contains N_{NS} non-specific binding sites for the repressor. As in Appendix , we can write the fold-change Eq. (2) in the grand canonical ensemble as

$$\text{fold-change} = \frac{1}{1 + \lambda_r e^{-\beta \Delta \varepsilon_{RA}}}, \quad (\text{A1})$$

where λ_r is the fugacity of the repressor and $\Delta \varepsilon_{RA}$ represents the energy difference between the repressor's binding affinity to the specific operator of interest relative to the repressor's non-specific binding affinity to the rest of the genome.

We now expand our definition of the total number of repressors in the system, R_{tot} , so that it is given by

$$R_{\text{tot}} = R_S + R_{NS} + R_C, \quad (\text{A2})$$

where R_S , R_{NS} , and R_C represent the number of repressors bound to the specific promoter, a non-specific binding site, or to a competitor binding site, respectively. The value of R_S is given by

$$R_S = N_S \frac{\lambda_r e^{-\beta \Delta \varepsilon_{RA}}}{1 + \lambda_r e^{-\beta \Delta \varepsilon_{RA}}}, \quad (\text{A3})$$

where N_S is the number of specific binding sites in the cell. The value of R_{NS} is similarly give by

$$R_{NS} = N_{NS} \frac{\lambda_r}{1 + \lambda_r}, \quad (\text{A4})$$

where N_{NS} is the number of non-specific sites in the cell (recall that we use $N_{NS} = 4.6 \times 10^6$ for *E. coli*), and R_C is given by

$$R_C = N_C \frac{\lambda_r e^{-\beta \Delta \varepsilon_C}}{1 + \lambda_r e^{-\beta \Delta \varepsilon_C}}, \quad (\text{A5})$$

where N_C is the number of competitor sites in the cell and $\Delta \varepsilon_C$ is the binding energy of the repressor to the competitor site relative to its non-specific binding energy to the rest of the genome.

To account for the induction of the repressor, we replace the total number of repressors R_{tot} in Eq. (A2) by the number of active repressors in the cell, $p_A(c)R_{\text{tot}}$. Here, p_A denotes the probability that

the repressor is in the active state (Eq. (4)),

$$p_A(c) = \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{-\beta\Delta\epsilon_{AI}} \left(1 + \frac{c}{K_I}\right)^n}. \quad (\text{A6})$$

Substituting in Eqs. (A3)-(A5) into the modified Eq. (A2) yields the form

$$p_A(c)R_{\text{tot}} = N_S \frac{\lambda_r e^{-\beta\Delta\epsilon_{RA}}}{1 + \lambda_r e^{-\beta\Delta\epsilon_{RA}}} + N_{NS} \frac{\lambda_r}{1 + \lambda_r} + N_C \frac{\lambda_r e^{-\beta\Delta\epsilon_C}}{1 + \lambda_r e^{-\beta\Delta\epsilon_C}}. \quad (\text{A7})$$

For systems where the number of binding sites N_S , N_{NS} , and N_C are known, together with the binding affinities $\Delta\epsilon_{RA}$ and $\Delta\epsilon_C$, we can solve numerically for λ_r and then substitute it into Eq. (A1) to obtain a fold-change at any concentration of inducer c . In the following sections, we will theoretically explore the induction curves given by Eq. (A7) for a number of different combinations of simple repression binding sites, thereby predicting how the system would behave if additional specific or competitor binding sites were introduced.

A.2 Variable Repressor Copy Number (R) with Multiple Specific Binding Sites ($N_S > 1$)

In the the main text, we consider the induction profiles of strains with varying R but a single, specific binding site $N_S = 1$ (see Fig. 5). Here we predict the induction profiles for similar strains in which R is varied, but $N_S > 1$, as shown in Fig. A1. The top row shows induction profiles in which $N_S = 10$ and the bottom row shows profiles in which $N_S = 100$, assuming three different choices for the specific operator binding sites given by the O1, O2, and O3 operators. These values of N_S were chosen to mimic the common scenario in which a promoter construct is placed on either a low or high copy number plasmid. A few features stand out in these profiles. First, as the magnitude of N_S surpasses the number of repressors R , the leakiness begins to increase significantly, since there are no longer enough repressors to regulate all copies of the promoter of interest. Second, in the cases where $\Delta\epsilon_{RA} = -15.3 k_B T$ for the O1 operator or $\Delta\epsilon_{RA} = -13.9 k_B T$ for the O2 operator, the profiles where $N_S = 100$ are notably sharper than the profiles where $N_S = 10$, and it is possible to achieve dynamic ranges approaching 1. Finally, it is interesting to note that the profiles for the O3 operator where $\Delta\epsilon_{RA} = -9.7 k_B T$ are nearly indifferent to the value of N_S .

A.3 Variable Number of Specific Binding Sites N_S with Fixed Repressor Copy Number (R)

The second set of scenarios we consider is the case in which the repressor copy number $R = 260$ is held constant while the number of specific promoters N_S is varied (see Fig. A2). Again we see that leakiness is increased significantly when $N_S > R$, though all profiles for $\Delta\epsilon_{RA} = -9.7 k_B T$ exhibit high leakiness, making the effect less dramatic for this operator. Additionally, we find again that adjusting the number of specific sites can produce induction profiles with maximal dynamic ranges. In particular, the O1 and O2 profiles with $\Delta\epsilon_{RA} = -15.3$ and $-13.9 k_B T$, respectively, have dynamic ranges approaching 1 for $N_S = 50$ and 100.

A.4 Competitor Binding Sites

An intriguing scenario is presented by the possibility of competitor sites elsewhere in the genome. This serves as a model for situations in which a promoter of interest is regulated by a transcription factor that has multiple targets. This is highly relevant, as the majority of transcription factors in *E. coli* have at least two known binding sites, with approximately 50 transcription factors having more than ten known binding sites (Rydenfelt et al. 2014b; Schmidt et al. 2015). If the number of competitor sites

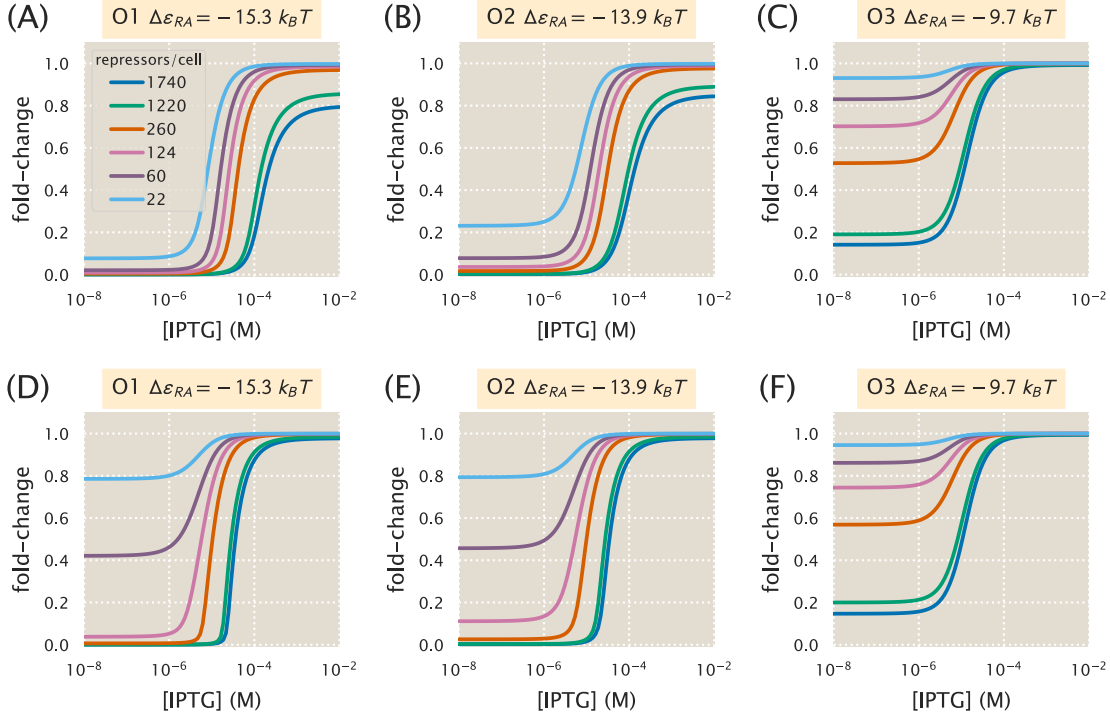


Figure A1. Induction with variable R and multiple specific binding sites. Induction profiles are shown for strains with variable R and $\Delta\varepsilon_{RA} = -15.3, -13.9, \text{ or } -9.7 k_B T$. (A-C) The number of specific sites, N_S , is held constant at 10 as R and $\Delta\varepsilon_{RA}$ are varied. (D-F) N_S is held constant at 100 as R and $\Delta\varepsilon_{RA}$ are varied. These situations mimic the common scenario in which a promoter construct is placed on either a low or high copy number plasmid.

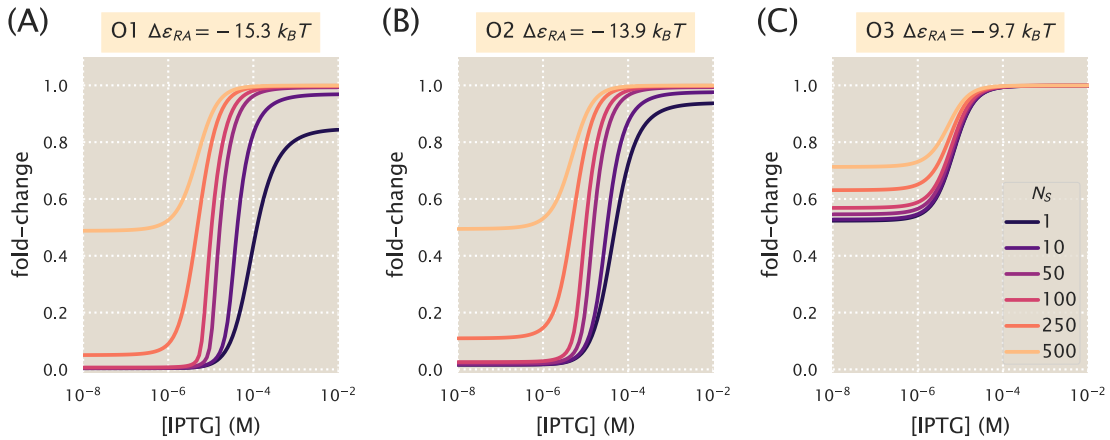


Figure A2. Induction with variable specific sites and fixed R . Induction profiles are shown for strains with $R = 260$ and (A) $\Delta\varepsilon_{RA} = -15.3 k_B T$, (B) $\Delta\varepsilon_{RA} = -13.9 k_B T$, or (C) $\Delta\varepsilon_{RA} = -9.7 k_B T$. The number of specific sites N_S is varied from 1 to 500.

and their average binding energy is known, however, they can be accounted for in the model. Here, we predict the induction profiles for strains in which $R = 260$ and $N_S = 1$, but there is a variable number of competitor sites N_C with a strong binding energy $\Delta\varepsilon_C = -17.0 k_B T$. In the presence of such a

strong competitor, when $N_C > R$ the leakiness is greatly increased, as many repressors are siphoned into the pool of competitor sites. This is most dramatic for the case where $\Delta\varepsilon_{RA} = -9.7 k_B T$, in which it appears that no repression occurs at all when $N_C = 500$. Interestingly, when $N_C < R$ the effects of the competitor are not especially notable.

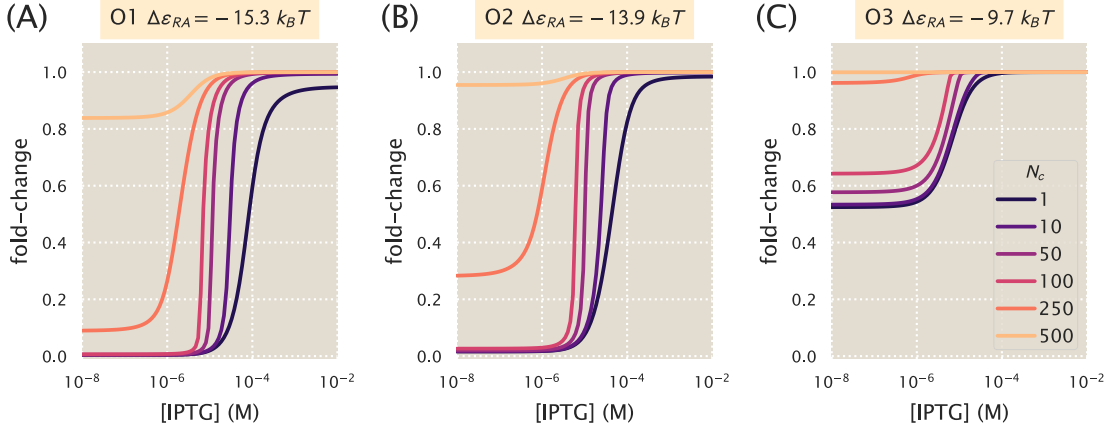


Figure A3. Induction with variable competitor sites, a single specific site, and fixed R . Induction profiles are shown for strains with $R = 260$, $N_s = 1$, and (A) $\Delta\varepsilon_{RA} = -15.3 k_B T$ for the O1 operator, (B) $\Delta\varepsilon_{RA} = -13.9 k_B T$ for the O2 operator, or (C) $\Delta\varepsilon_{RA} = -9.7 k_B T$ for the O3 operator. The number of specific sites, N_C , is varied from 1 to 500. This mimics the common scenario in which a transcription factor has multiple binding sites in the genome.

A.5 Properties of the Induction Response

As discussed in the main body of the paper, our treatment of the MWC model allows us to predict key properties of induction responses. Here, we consider the leakiness, saturation, and dynamic range (see Fig. 1) by numerically solving Eq. (A7) in the absence of inducer, $c = 0$, and in the presence of saturating inducer $c \rightarrow \infty$. Using Eq. (A6), the former case is given by

$$R_{\text{tot}} \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}}} = N_S \frac{\lambda_r e^{-\beta\Delta\varepsilon_{RA}}}{1 + \lambda_r e^{-\beta\Delta\varepsilon_{RA}}} + N_{NS} \frac{\lambda_r}{1 + \lambda_r} + N_C \frac{\lambda_r e^{-\beta\Delta\varepsilon_C}}{1 + \lambda_r e^{-\beta\Delta\varepsilon_C}}, \quad (\text{A8})$$

whereupon substituting in the value of λ_r into Eq. (A1) will yield the leakiness. Similarly, the limit of saturating inducer is found by determining λ_r from the form

$$R_{\text{tot}} \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^2} = N_S \frac{\lambda_r e^{-\beta\Delta\varepsilon_{RA}}}{1 + \lambda_r e^{-\beta\Delta\varepsilon_{RA}}} + N_{NS} \frac{\lambda_r}{1 + \lambda_r} + N_C \frac{\lambda_r e^{-\beta\Delta\varepsilon_C}}{1 + \lambda_r e^{-\beta\Delta\varepsilon_C}}. \quad (\text{A9})$$

In Fig. A4 we show how the leakiness, saturation, and dynamic range vary with R and $\Delta\varepsilon_{RA}$ in systems with $N_S = 10$ or $N_S = 100$. An inflection point occurs where $N_S = R$, with leakiness and dynamic range behaving differently when $R < N_S$ than when $R > N_S$. This transition is more dramatic for $N_S = 100$ than for $N_S = 10$. Interestingly, the saturation values consistently approach 1, indicating that full induction is easier to achieve when multiple specific sites are present. Moreover, dynamic range values for O1 and O2 strains with $\Delta\varepsilon_{RA} = -15.3$ and $-13.9 k_B T$ approach 1 when $R > N_S$, although when $N_S = 10$ there is a slight downward dip owing to saturation values of less than 1 at high repressor copy numbers.

In Fig. A5 we similarly show how the leakiness, saturation, and dynamic range vary with R and $\Delta\varepsilon_{RA}$ in systems with $N_S = 1$ and multiple competitor sites $N_C = 10$ or $N_C = 100$. Each of the competitor sites has a binding energy of $\Delta\varepsilon_C = -17.0 k_B T$. The phenotypic profiles are very similar to those for

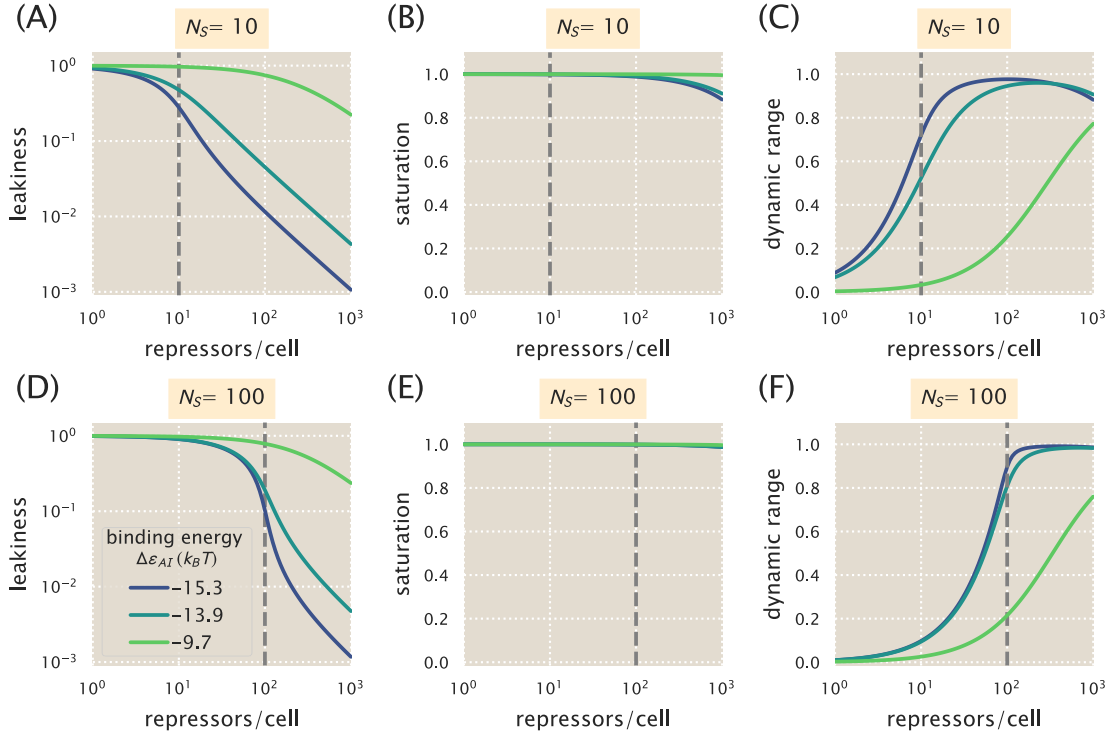


Figure A4. Phenotypic properties of induction with multiple specific binding sites. The leakiness (A, D), saturation (B, E), and dynamic range (C, F) are shown for systems with number of specific binding sites $N_S = 10$ (A-C) or $N_S = 100$ (D-F). The dashed vertical line indicates the point at which $N_S = R$.

multiple specific sites shown in Fig. A4, with sharper transitions at $R = N_C$ due to the greater binding strength of the competitor site. This indicates that introducing competitors has much the same effect on the induction phenotypes as introducing additional specific sites, as in either case the influence of the repressors is dampened when there are insufficient repressors to interact with all of the specific binding sites.

This section of the appendix gives a quantitative analysis of the nuances imposed on induction response in the case of systems involving multiple gene copies as are found in the vast majority of studies on induction. In these cases, the intrinsic parameters of the MWC model get entangled with the parameters describing gene copy number.

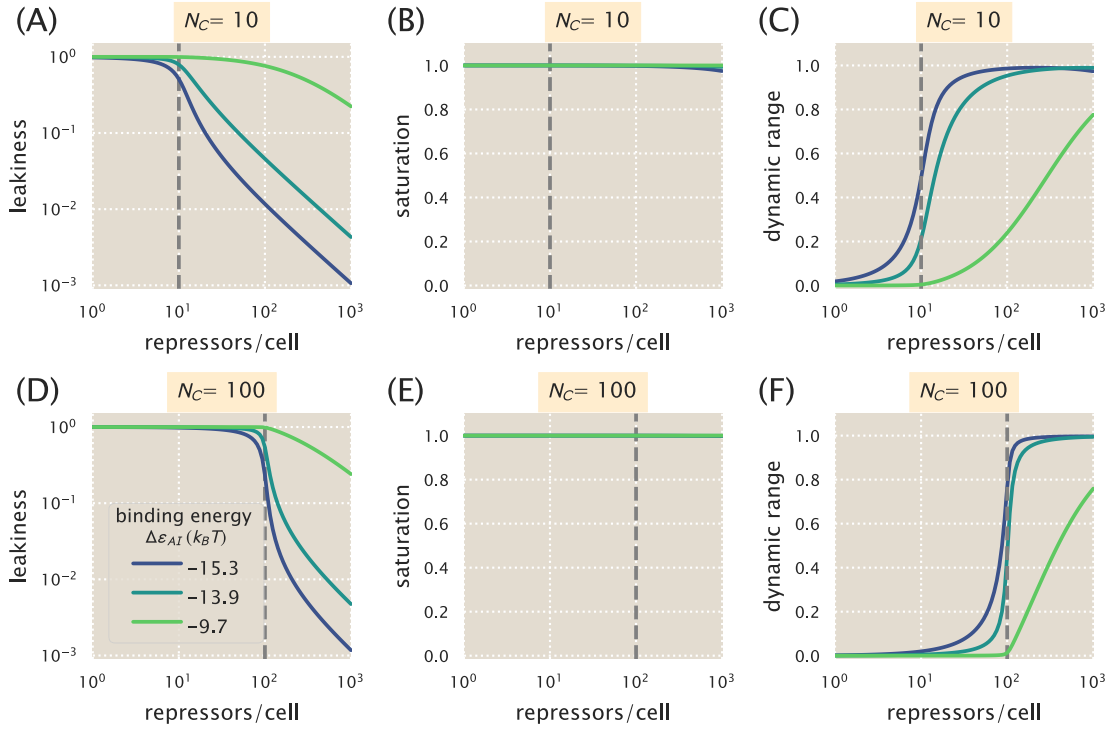


Figure A5. Phenotypic properties of induction with a single specific site and multiple competitor sites. The leakiness (A, D), saturation (B, E), and dynamic range (C, F) are shown for systems with a single specific binding site $N_S = 1$ and a number of competitor sites $N_C = 10$ (A-C) or $N_C = 100$ (D-F). All competitor sites have a binding energy of $\Delta\epsilon_C = -17.0 k_B T$. The dashed vertical line indicates the point at which $N_C = R$.

B Single-Cell Microscopy

In this section, we detail the procedures and results from single-cell microscopy verification of our flow cytometry measurements. Our previous measurements of fold-change in gene expression have been measured using bulk-scale Miller assays (Garcia and Phillips 2011a) or through single-cell microscopy (Brewster et al. 2014). In this work, flow cytometry was an attractive method due to the ability to screen through many different strains at different concentrations of inducer in a short amount of time. To verify our results from flow cytometry, we examined two bacterial strains with different repressor-DNA binding energies ($\Delta\varepsilon_{RA}$) of $-13.9 k_B T$ and $-15.3 k_B T$ with $R = 260$ repressors per cell using fluorescence microscopy and estimated the values of the parameters K_A and K_I for direct comparison between the two methods. For a detailed explanation of the Python code implementation of the processing steps described below, please see this paper’s [GitHub repository](#). An outline of our microscopy workflow can be seen in Fig. A6.

B.1 Strains and Growth Conditions

Cells were grown in an identical manner to those used for measurement via flow cytometry (see Methods). Briefly, cells were grown overnight (between 10 and 13 hours) to saturation in rich media broth (LB) with $100 \mu\text{g} \cdot \text{mL}^{-1}$ spectinomycin in a deep-well 96 well plate at 37°C . These cultures were then diluted 1000-fold into $500 \mu\text{L}$ of M9 minimal medium supplemented with 0.5% glucose and the appropriate concentration of the inducer IPTG. Strains were allowed to grow at 37°C with vigorous aeration for approximately 8 hours. Prior to mounting for microscopy, the cultures were diluted 10-fold into M9 glucose minimal medium in the absence of IPTG. Each construct was measured using the same range of inducer concentration values as was performed in the flow cytometry measurements (between 100 nM and 5 mM IPTG). Each condition was measured in triplicate in microscopy whereas approximately ten measurements were made using flow cytometry.

B.2 Imaging Procedure

During the last hour of cell growth, an agarose mounting substrate was prepared containing the appropriate concentration of the IPTG inducer. This mounting substrate was composed of M9 minimal medium supplemented with 0.5% glucose and 2% agarose (Life Technologies UltraPure Agarose, Cat. No. 16500100). This solution was heated in a microwave until molten followed by addition of the IPTG to the appropriate final concentration. This solution was then thoroughly mixed and a $500 \mu\text{L}$ aliquot was sandwiched between two glass coverslips and was allowed to solidify.

Once solid, the agarose substrates were cut into approximately $10 \text{ mm} \times 10 \text{ mm}$ squares. An aliquot of one to two microliters of the diluted cell suspension was then added to each pad. For each concentration of inducer, a sample of the autofluorescence control, the $\Delta lacI$ constitutive expression control, and the experimental strain was prepared yielding a total of thirty-six agarose mounts per experiment. These samples were then mounted onto two glass-bottom dishes (Ted Pella Wilco Dish, Cat. No. 14027-20) and sealed with parafilm.

All imaging was performed on a Nikon Ti-Eclipse inverted fluorescent microscope outfitted with a custom-built laser illumination system and operated by the open-source MicroManager control software (Edelstein et al. 2014). The YFP fluorescence was imaged using a CrystaLaser 514 nm excitation laser coupled with a laser-optimized (Semrock Cat. No. LF514-C-000) emission filter.

For each sample, between fifteen and twenty positions were imaged allowing for measurement of several hundred cells. At each position, a phase contrast image, an mCherry image, and a YFP image were collected in that order with exposures on a time scale of ten to twenty milliseconds. For each channel, the same exposure time was used across all samples in a given experiment. All images were collected and stored in `ome.tiff` format. All microscopy images are available on the CaltechDATA online repository under DOI: 10.22002/D1.229.

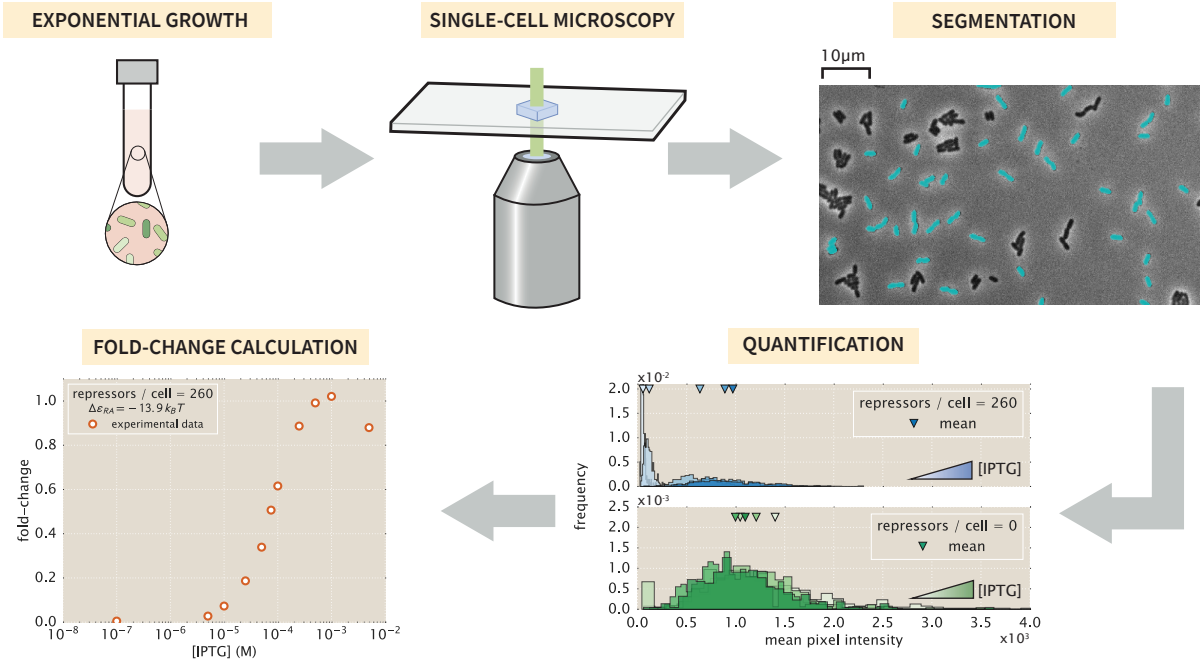


Figure A6. Experimental workflow for single-cell microscopy. For comparison with the flow cytometry results, the cells were grown in an identical manner to those described in the main text. Once cells had reached mid to late exponential growth, the cultures were diluted and placed on agarose substrates and imaged under $100\times$ magnification. Regions of interest representing cellular mass were segmented and average single-cell intensities were computed. The means of the distributions were used to compute the fold-change in gene expression.

B.3 Image Processing

B.3.1 Correcting Uneven Illumination

The excitation laser has a two-dimensional gaussian profile. To minimize non-uniform illumination of a single field of view, the excitation beam was expanded to illuminate an area larger than that of the camera sensor. While this allowed for an entire field of view to be illuminated, there was still approximately a 10% difference in illumination across both dimensions. This nonuniformity was corrected for in post-processing by capturing twenty images of a homogeneously fluorescent plastic slide (Autofluorescent Plastic Slides, Chroma Cat. No. 920001) and averaging to generate a map of illumination intensity at any pixel I_{YFP} . To correct for shot noise in the camera (Andor iXon+ 897 EMCCD), twenty images were captured in the absence of illumination using the exposure time used for the experimental data. Averaging over these images produced a map of background noise at any pixel I_{dark} . To perform the correction, each fluorescent image in the experimental acquisition was renormalized with respect to these average maps as

$$I_{\text{flat}} = \frac{I - I_{\text{dark}}}{I_{\text{YFP}} - I_{\text{dark}}} \langle I_{\text{YFP}} - I_{\text{dark}} \rangle, \quad (\text{A10})$$

where I_{flat} is the renormalized image and I is the original fluorescence image. An example of this correction can be seen in Fig. A7.

B.3.2 Cell Segmentation

Each bacterial strain constitutively expressed an mCherry fluorophore from a low copy-number plasmid. This served as a volume marker of cell mass allowing us to segment individual cells through edge detection in fluorescence. We used the Marr-Hildreth edge detector (Marr and Hildreth 1980) which identifies edges

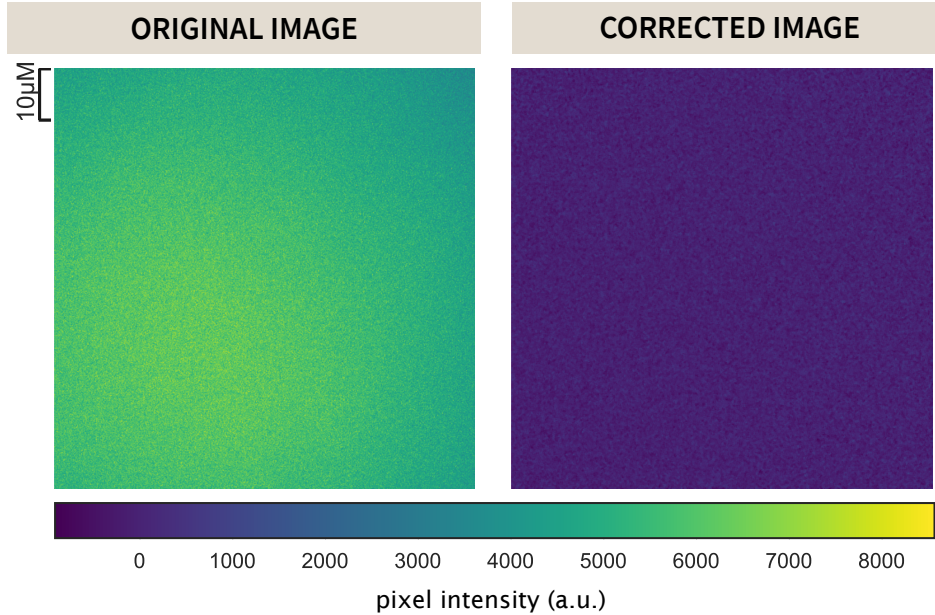


Figure A7. Correction for uneven illumination. A representative image of the illumination profile of the 512 nm excitation beam on a homogeneously fluorescent slide is shown in the left panel. This is corrected for using equation Eq. (A10) and is shown in the right panel.

by taking the second derivative of a lightly Gaussian blurred image. Edges are identified as those regions which cross from highly negative to highly positive values or vice-versa within a specified neighborhood. Bacterial cells were defined as regions within an intact and closed identified edge. All segmented objects were then labeled and passed through a series of filtering steps.

To ensure that primarily single cells were segmented, we imposed area and eccentricity bounds. We assumed that single cells projected into two dimensions are roughly $2\ \mu\text{m}$ long and $1\ \mu\text{m}$ wide, so that cells are likely to have an area between $0.5\ \mu\text{m}^2$ and $6\ \mu\text{m}^2$. To determine the eccentricity bounds, we assumed that a single cell can be approximated by an ellipse with semi-major (a) and semi-minor (b) axis lengths of $0.5\ \mu\text{m}$ and $0.25\ \mu\text{m}$, respectively. The eccentricity of this hypothetical cell can be computed as

$$\text{eccentricity} = \sqrt{1 - \left(\frac{b}{a}\right)^2}, \quad (\text{A11})$$

yielding a value of approximately 0.8. Any objects with an eccentricity below this value were not considered to be single cells. After imposing both an area (Fig. A8(A)) and eccentricity filter (Fig. A8(B)), the remaining objects were considered cells of interest (Fig. A8(C)) and the mean fluorescence intensity of each cell was extracted.

B.3.3 Calculation of Fold-Change

Cells exhibited background fluorescence even in the absence of an expressed fluorophore. We corrected for this autofluorescence contribution to the fold-change calculation by subtracting the mean YFP fluorescence of cells expressing only the mCherry volume marker from each experimental measurement. The fold-change in gene expression was therefore calculated as

$$\text{fold-change} = \frac{\langle I_{R>0} \rangle - \langle I_{\text{auto}} \rangle}{\langle I_{R=0} \rangle - \langle I_{\text{auto}} \rangle}, \quad (\text{A12})$$

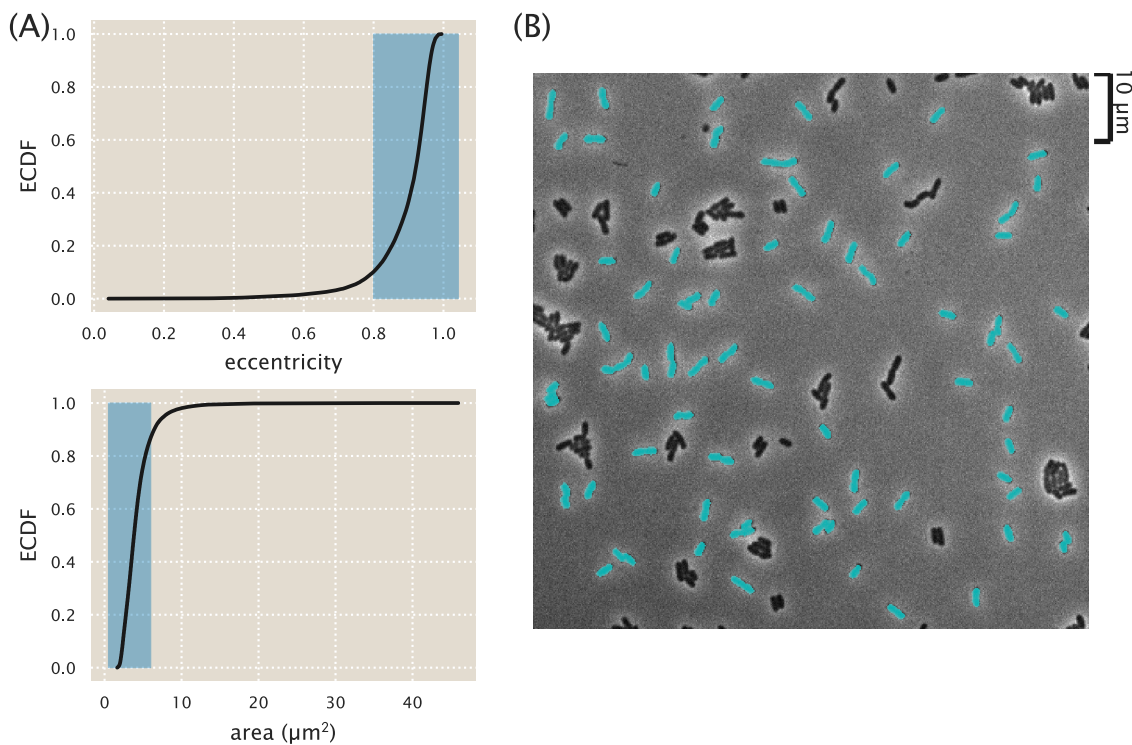


Figure A8. Segmentation of single bacterial cells. (A) Objects were selected if they had an eccentricity greater than 0.8 and an area between $0.5 \mu\text{m}^2$ and $6 \mu\text{m}^2$. Highlighted in blue are the regions considered to be representative of single cells. The black lines correspond to the empirical cumulative distribution functions for the parameter of interest. (B) A representative final segmentation mask is shown in which segmented cells are depicted in cyan over the phase contrast image.

where $\langle I_{R>0} \rangle$ is the mean fluorescence intensity of cells expressing LacI repressors, $\langle I_{\text{auto}} \rangle$ is the mean intensity of cells expressing only the mCherry volume marker, and $\langle I_{R=0} \rangle$ is the mean fluorescence intensity of cells in the absence of LacI. These fold-change values were very similar to those obtained through flow cytometry and were well described using the thermodynamic parameters used in the main text. With these experimentally measured fold-change values, the best-fit parameter values of the model were inferred and compared to those obtained from flow cytometry.

B.4 Parameter Estimation and Comparison

To confirm quantitative consistency between flow cytometry and microscopy, the parameter values of K_A and K_I were also estimated from three biological replicates of IPTG titration curves obtained by microscopy for strains with $R = 260$ and operators O1 and O2. Fig. A9(A) shows the data from these measurements (orange circles) and the ten biological replicates from our flow cytometry measurements (blue circles), along with the fold-change predictions from each inference. In comparison with the values obtained by flow cytometry, each parameter estimate overlapped with the 95% credible region of our flow cytometry estimates, as shown in Fig. A9(B). Specifically, these values were $K_A = 142^{+40}_{-34} \mu\text{M}$ and $K_I = 0.6^{+0.1}_{-0.1} \mu\text{M}$ from microscopy and $K_A = 149^{+14}_{-12} \mu\text{M}$ and $K_I = 0.57^{+0.03}_{-0.02} \mu\text{M}$ from the flow cytometry data. We note that the credible regions from the microscopy data shown in Fig. A9(B) are much broader than those from flow cytometry due to the fewer number of replicates performed.

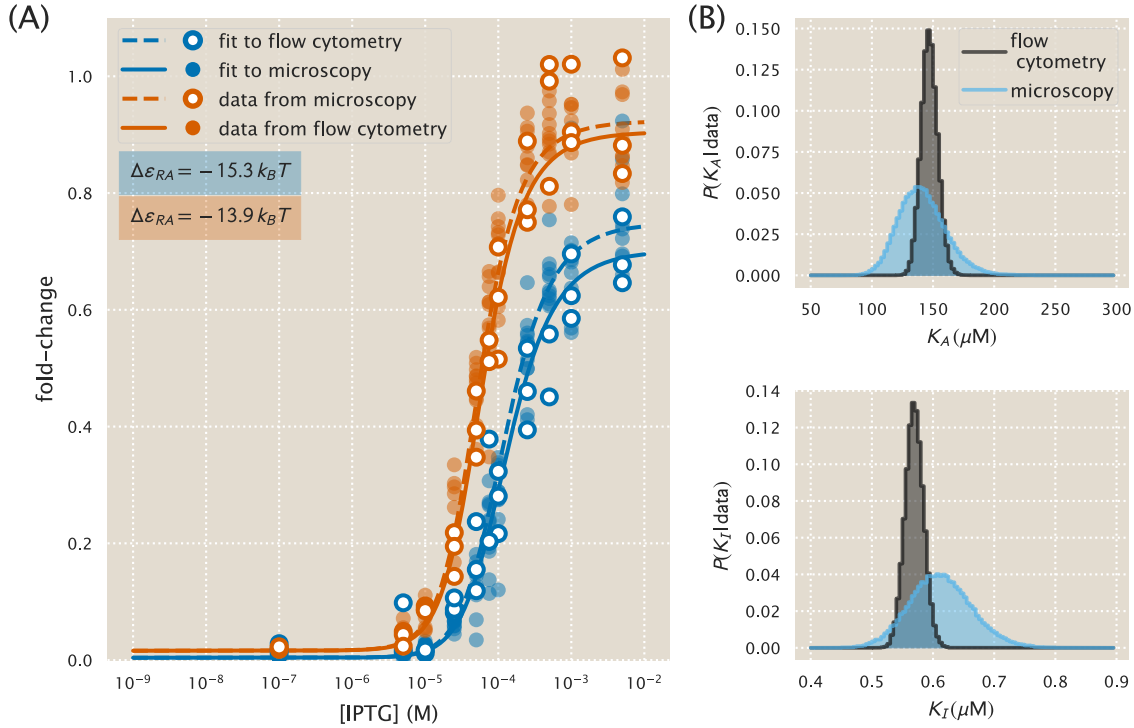


Figure A9. Comparison of measured fold-change between flow cytometry and single-cell microscopy. (A) Experimentally measured fold-change values obtained through single-cell microscopy and flow cytometry are shown as white filled and solid colored circles, respectively. Solid and dashed lines indicate the predicted behavior using the most likely parameter values of K_A and K_I inferred from flow cytometry data and microscopy data, respectively. The red and blue plotting elements correspond to the different operators O1 and O2 with binding energies $\Delta\epsilon_{RA}$ of $-13.9 k_B T$ and $-15.3 k_B T$, respectively (Garcia and Phillips 2011a). (B) The marginalized posterior distributions for K_A and K_I are shown in the top and bottom panel, respectively. The posterior distribution determined using the microscopy data is wider than that computed using the flow cytometry data due to a smaller collection of data sets (three for microscopy and ten for flow cytometry).

C Fold-Change Sensitivity Analysis

In Fig. 5 we found that the width of the credible regions varied widely depending on the repressor copy number R and repressor operator binding energy $\Delta\varepsilon_{RA}$. More precisely, the credible regions were much narrower for low repressor copy numbers R and weak binding energy $\Delta\varepsilon_{RA}$. In this section, we explain how this behavior comes about. We focus our attention on the maximum fold-change in the presence of saturating inducer given by Eq. (7). While it is straightforward to consider the width of the credible regions at any other inducer concentration, Fig. 5 shows that the credible region are widest at saturation.

The width of the credible regions corresponds to how sensitive the fold-change is to the fit values of the dissociation constants K_A and K_I . To be quantitative, we define

$$\Delta \text{fold-change}_{K_A} \equiv \text{fold-change}(K_A, K_I^{\text{fit}}) - \text{fold-change}(K_A^{\text{fit}}, K_I^{\text{fit}}), \quad (\text{A13})$$

the difference between the fold-change at a particular K_A value relative to the best-fit dissociation constant $K_A^{\text{fit}} = 139 \times 10^{-6}$ M. For simplicity, we keep the inactive state dissociation constant fixed at its best-fit value $K_I^{\text{fit}} = 0.53 \times 10^{-6}$ M. A larger difference $\Delta \text{fold-change}_{K_A}$ implies a wider credible region. Similarly, we define the analogous quantity

$$\Delta \text{fold-change}_{K_I} = \text{fold-change}(K_A^{\text{fit}}, K_I) - \text{fold-change}(K_A^{\text{fit}}, K_I^{\text{fit}}) \quad (\text{A14})$$

to measure the sensitivity of the fold-change to K_I at a fixed K_A^{fit} . Fig. A10 shows both of these quantities in the limit $c \rightarrow \infty$ for different repressor-DNA binding energies $\Delta\varepsilon_{RA}$ and repressor copy numbers R . See our [GitHub repository](#) for the code that reproduces these plots.

To understand how the width of the credible region scales with $\Delta\varepsilon_{RA}$ and R , we can Taylor expand the difference in fold-change to first order, $\Delta \text{fold-change}_{K_A} \approx \frac{\partial \text{fold-change}}{\partial K_A} (K_A - K_A^{\text{fit}})$, where the partial derivative has the form

$$\frac{\partial \text{fold-change}}{\partial K_A} = \frac{e^{-\beta\Delta\varepsilon_{AI}} \frac{n}{K_I} \left(\frac{K_A}{K_I}\right)^{n-1}}{\left(1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^n\right)^2} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}} \left(1 + \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^n} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}}\right)^{-2}. \quad (\text{A15})$$

Similarly, the Taylor expansion $\Delta \text{fold-change}_{K_I} \approx \frac{\partial \text{fold-change}}{\partial K_I} (K_I - K_I^{\text{fit}})$ features the partial derivative

$$\frac{\partial \text{fold-change}}{\partial K_I} = -\frac{e^{-\beta\Delta\varepsilon_{AI}} \frac{n}{K_I} \left(\frac{K_A}{K_I}\right)^n}{\left(1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^n\right)^2} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}} \left(1 + \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^n} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}}\right)^{-2}. \quad (\text{A16})$$

From Eqs. (A15) and (A16), we find that both $\Delta \text{fold-change}_{K_A}$ and $\Delta \text{fold-change}_{K_I}$ increase in magnitude with R and decrease in magnitude with $\Delta\varepsilon_{RA}$. Accordingly, we expect that the O3 strains (with the least negative $\Delta\varepsilon_{RA}$) and the strains with the smallest repressor copy number will lead to partial derivatives with smaller magnitude and hence to tighter credible regions. Indeed, this prediction is carried out in Fig. A10.

Lastly, we note that Eqs. (A15) and (A16) enable us to quantify the scaling relationship between the width of the credible region and the two quantities R and $\Delta\varepsilon_{RA}$. For example, for the O3 strains, where the fold-change at saturating inducer concentration is ≈ 1 , the right-most term in both equations which equals the fold-change squared is roughly 1. Therefore, we find that both $\frac{\partial \text{fold-change}}{\partial K_A}$ and $\frac{\partial \text{fold-change}}{\partial K_I}$ scale linearly with R and $e^{-\beta\Delta\varepsilon_{RA}}$. Thus the width of the $R = 22$ strain will be roughly 1/1000 as large as that of the $R = 1740$ strain; similarly, the width of the O3 curves will be roughly 1/1000 the width of the O1 curves.

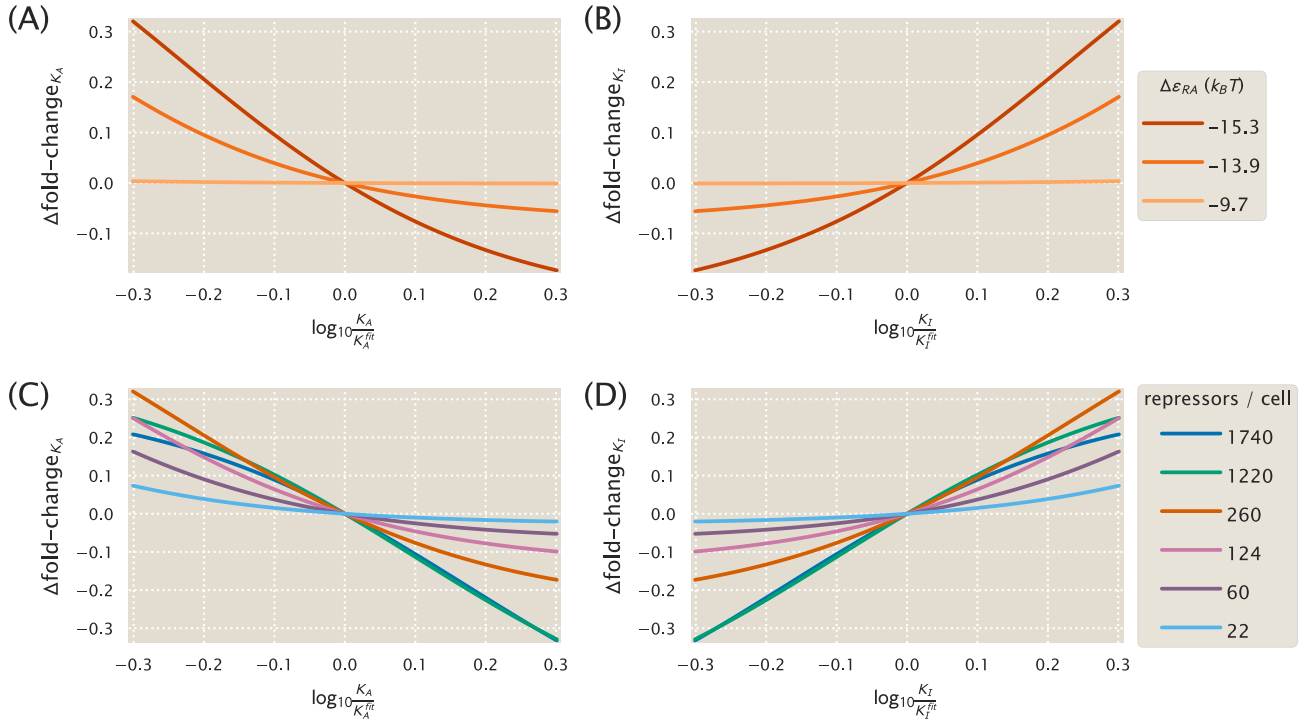


Figure A10. Determining how sensitive the fold-change values are to the fit values of the dissociation constants. (A) The difference $\Delta \text{fold-change}_{K_A}$ in fold change when the dissociation constant K_A is slightly offset from its best-fit value $K_A = 139_{-22}^{+29} \times 10^{-6}$ M, as given by Eq. (A13). Fold-change is computed in the limit of saturating inducer concentration ($c \rightarrow \infty$, see Eq. (7)) where the credible regions in Fig. 5 are widest. The O3 strain ($\Delta \varepsilon_{RA} = -9.7 k_B T$) is about 1/1000 as sensitive as the O1 operator to perturbations in the parameter values, and hence its credible region is roughly 1/1000 as wide. All curves were made using $R = 260$. (B) As in Panel (A), but plotting the sensitivity of fold-change to the K_I parameter relative to the best-fit value $K_I = 0.53_{-0.04}^{+0.04} \times 10^{-6}$ M. Note that only the magnitude, and not the sign, of this difference describes the sensitivity of each parameter. Hence, the O3 strain is again less sensitive than the O1 and O2 strains. (C) As in Panel (A), but showing how the fold-change sensitivity for different repressor copy numbers. The strains with lower repressor copy number are less sensitive to changes in the dissociation constants, and hence their corresponding curves in Fig. 5 have tighter credible regions. All curves were made using $\Delta \varepsilon_{RA} = -13.9 k_B T$. (D) As in Panel (C), the sensitivity of fold-change with respect to K_I is again smallest (in magnitude) for the low repressor copy number strains.

D Applicability of Theory to the Oid Operator Sequence

In addition to the native operator sequences (O1, O2, and O3) considered in the main text, we were also interested in testing our model predictions against the synthetic Oid operator. In contrast to the other operators, Oid is one base pair shorter in length (20 bp), is fully symmetric, and is known to provide stronger repression than the native operator sequences considered so far. While the theory should be similarly applicable, measuring the lower fold-changes associated with this YFP construct was expected to be near the sensitivity limit for our flow cytometer, due to the especially strong binding energy of Oid ($\Delta\varepsilon_{RA} = -17.0 k_B T$) (Garcia et al. 2011b). Accordingly, fluorescence data for Oid were obtained using microscopy, which is more sensitive than flow cytometry. Appendix B gives a detailed explanation of how microscopy measurements were used to obtain induction curves.

We follow the approach of the main text and make fold-change predictions based on the parameter estimates from our strain with $R = 260$ and an O2 operator. These predictions are shown in Fig. A11(A), where we also plot data taken in triplicate for strains containing $R = 22, 60,$ and 124 , obtained by single-cell microscopy. We find that the data are systematically below the theoretical predictions. We also considered our global fitting approach (see Appendix) to see whether we might find better agreement with the observed data. Interestingly, we find that the majority of the parameters remain largely unchanged, but our estimate for the Oid binding energy $\Delta\varepsilon_{RA}$ is shifted to $-17.7 k_B T$ instead of the value $-17.0 k_B T$ found by Garcia and Phillips (2011a). In Fig. A11(B) we again plot the Oid fold-change data but with theoretical predictions using the new estimate for the Oid binding energy from our global fit and find substantially better agreement.

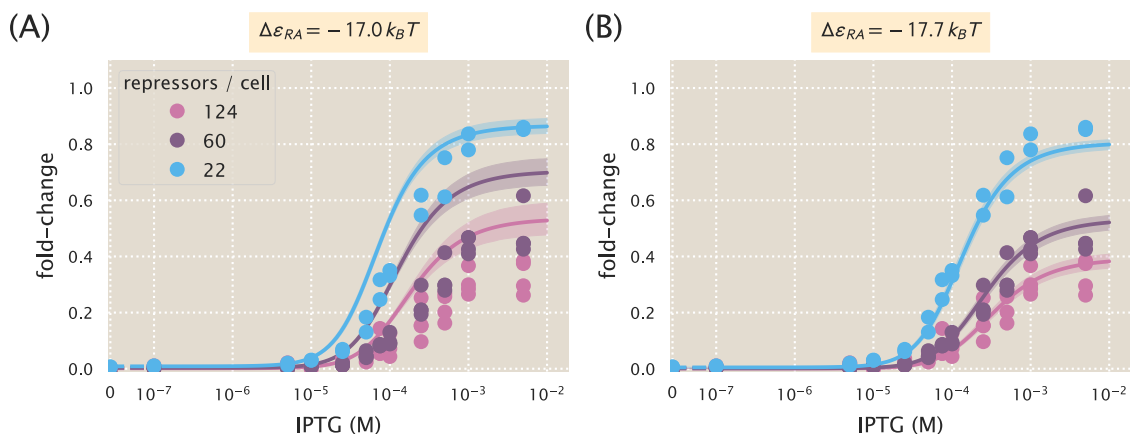


Figure A11. Predictions of fold-change for strains with an Oid binding sequence versus experimental measurements with different repressor copy numbers. (A) Experimental data is plotted against the parameter-free predictions that are based on our fit to the O2 strain with $R = 260$. Here we use the previously measured binding energy $\Delta\varepsilon_{RA} = -17.0 k_B T$ (Garcia and Phillips 2011a). (B) The same experimental data is plotted against the best-fit parameters using the complete O1, O2, O3, and Oid data sets to infer K_A , K_I , repressor copy numbers, and the binding energies of all operators (see Appendix). Here the major difference in the inferred parameters is a shift in the binding energy for Oid from $\Delta\varepsilon_{RA} = -17.0 k_B T$ to $\Delta\varepsilon_{RA} = -17.7 k_B T$, which now shows agreement between the theoretical predictions and experimental data. Shaded regions from the theoretical curves denote the 95% credible region. These are narrower in Panel (B) because the inference of parameters was performed with much more data, and hence the best-fit values are more tightly constrained. Individual data points are shown due to the small number of replicates. The dashed lines at 0 IPTG indicate a linear scale, whereas solid lines represent a log scale.

Fig. A12 shows the cumulative data from Garcia and Phillips (2011a) and Brewster et al. (2014), as well as our data with $c = 0 \mu M$, which all measured fold-change for the same simple repression architecture

utilizing different reporters and measurement techniques. We find that the binding energies from the global fit, including $\Delta\epsilon_{RA} = -17.7 k_B T$, compare reasonably well with all previous measurements.

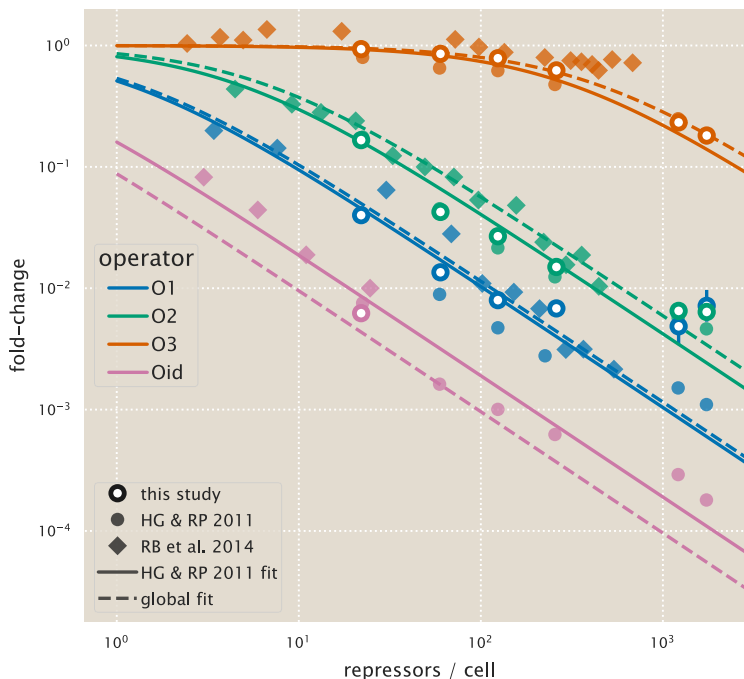


Figure A12. Comparison of fold-change predictions based on binding energies from Garcia and Phillips and those inferred from this work. Fold-change curves for the different repressor-DNA binding energies $\Delta\epsilon_{RA}$ are plotted as a function of repressor copy number when IPTG concentration $c = 0$. Solid curves use the binding energies determined from Garcia and Phillips (2011a), while the dashed curves use the inferred binding energies we obtained when performing a global fit of K_A , K_I , repressor copy numbers, and the binding energies using all available data from our work. Fold-change measurements from our experiments (outlined circles) Garcia and Phillips (2011a) (solid circles), and Brewster et al. (2014) (diamonds) show that the small shifts in binding energy that we infer are still in agreement with prior data. Note that only a single flow cytometry data point is shown for Oid from this study, since the $R = 60$ and $R = 124$ curves from Fig. A11 had extremely low fold-change in the absence of inducer ($c = 0$) so as to be indistinguishable from autofluorescence, and in fact their fold-change values in this limit were negative and hence do not appear on this plot.

E Applications to Other Regulatory Architectures

In this section, we discuss how the theoretical framework presented in this work is sufficiently general to include a variety of regulatory architectures outside of simple repression by LacI. We begin by noting that the exact same formula for fold-change given in Eq. (5) can also describe corepression. We then demonstrate how our model can be generalized to include other architectures, such as a coactivator binding to an activator to promote gene expression. In each case, we briefly describe the system and describe its corresponding theoretical description. For further details, we invite the interested reader to read Bintu et al. (2005b) and Marzen et al. (2013).

E.1 Corepression

Consider a regulatory architecture where binding of a transcriptional repressor occludes the binding of RNAP to the DNA. A corepressor molecule binds to the repressor and shifts its allosteric equilibrium towards the active state in which it binds more tightly to the DNA, thereby decreasing gene expression (in contrast, an inducer shifts the allosteric equilibrium towards the inactive state where the repressor binds more weakly to the DNA). As in the main text, we can enumerate the states and statistical weights of the promoter and the allosteric states of the repressor. We note that these states and weights exactly match Fig. 2 and yield the same fold-change equation as Eq. (5),

$$\text{fold-change} \approx \left(1 + \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{\beta\Delta\epsilon_{AI}} \left(1 + \frac{c}{K_I}\right)^n} \frac{R}{N_{NS}} e^{-\beta\Delta\epsilon_{RA}} \right)^{-1}, \quad (\text{A17})$$

where c now represents the concentration of the corepressor molecule. Mathematically, the difference between these two architectures can be seen in the relative sizes of the dissociation constants K_A and K_I between the inducer and repressor in the active and inactive states, respectively. The corepressor is defined by $K_A < K_I$, since the corepressor favors binding to the repressor's active state; an inducer must satisfy $K_I < K_A$, as was found in the main text from the induction data (see Fig. 4). Much as was performed in the main text, we can make some predictions about the how the response of a corepressor. In Fig. A13(A), we show how varying the repressor copy number R and the repressor-DNA binding energy $\Delta\epsilon_{RA}$ influence the response. We draw the reader's attention to the decrease in fold-change as the concentration of effector is increased.

E.2 Activation

We now turn to the case of activation. While this architecture was not studied in this work, we wish to demonstrate how the framework presented here can be extended to include transcription factors other than repressors. To that end, we consider a transcriptional activator which binds to DNA and aids in the binding of RNAP through energetic interaction term ϵ_{AP} . Note that in this architecture, binding of the activator does not occlude binding of the polymerase. Binding of a coactivator molecule shifts its allosteric equilibrium towards the active state ($K_A < K_I$), where the activator is more likely to be bound to the DNA and promote expression. Enumerating all of the states and statistical weights of this architecture and making the approximation that the promoter is weak generates a fold-change equation of the form

$$\text{fold-change} = \frac{1 + \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{\beta\Delta\epsilon_{AI}} \left(1 + \frac{c}{K_I}\right)^n} \frac{A}{N_{NS}} e^{-\beta\Delta\epsilon_{AA}} e^{-\beta\epsilon_{AP}}}{1 + \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{\beta\Delta\epsilon_{AI}} \left(1 + \frac{c}{K_I}\right)^n} \frac{A}{N_{NS}} e^{-\beta\Delta\epsilon_{AA}}}, \quad (\text{A18})$$

where A is the total number of activators per cell, c is the concentration of a coactivator molecule, $\Delta\epsilon_{AA}$ is the binding energy of the activator to the DNA in the active allosteric state, and ϵ_{AP} is the interaction energy between the activator and the RNAP. Unlike in the cases of induction and corepression, the

fold-change formula for activation includes terms from when the RNAP is bound by itself on the DNA as well as when both RNAP and the activator are simultaneously bound to the DNA. Fig. A13(B) explores predictions of the fold-change in gene expression by manipulating the activator copy number, DNA binding energy, and the polymerase-activator interaction energy. Note that with this activation scheme, the fold-change must necessarily be greater than one. An interesting feature of these predictions is the observation that even small changes in the interaction energy ($< 0.5 k_B T$) can result in dramatic increase in fold-change.

As in the case of induction, the Eq. (A18) is straightforward to generalize. For example, the relative values of K_I and K_A can be switched such that $K_I < K_A$ in which the secondary molecule drives the activator to assume the inactive state represents induction of an activator. While these cases might be viewed as separate biological phenomena, mathematically they can all be described by the same underlying formalism.

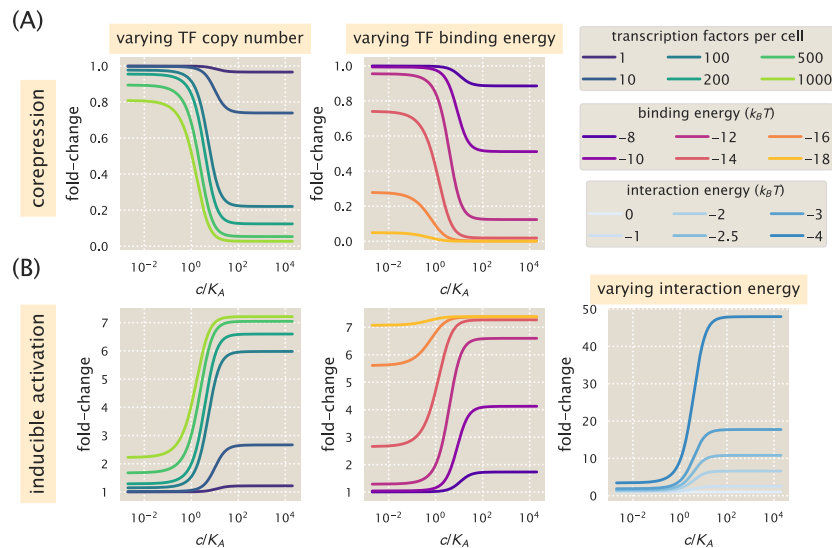


Figure A13. Representative fold-change predictions for allosteric corepression and activation. (A) Contrary to the case of induction described in the main text, addition of a corepressor decreases fold-change in gene expression. The left and right panels demonstrate how varying the values of the repressor copy number R and repressor-DNA binding energy $\Delta\epsilon_{RA}$, respectively, change the predicted response profiles. (B) In the case of inducible activation, binding of an effector molecule to an activator transcription factor increases the fold-change in gene expression. Note that for activation, the fold-change is greater than 1. The left and center panels show how changing the activator copy number A and activator-DNA binding energy $\Delta\epsilon_{AA}$ alter response, respectively. The right panel shows how varying the polymerase-activator interaction energy ϵ_{AP} alters the fold-change. Relatively small perturbations to this energetic parameter drastically changes the level of activation and plays a major role in dictating the dynamic range of the system.

Supplemental References

- Edelstein, A. D., Tsuchida, M. A., Amodaj, N., Pinkard, H., Vale, R. D., and Stuurman, N. (2014). Advanced methods of microscope control using μ Manager software. *Journal of Biological Methods* 1.2, pp. 10–10.
- Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society B: Biological Sciences* 207.1167, pp. 187–217.
- Oehler, S., Alberti, S., and Müller-Hill, B. (2006). Induction of the *lac* promoter in the absence of DNA loops and the stoichiometry of induction. *Nucleic Acids Research* 34.2, pp. 606–612.
- Rydenfelt, M., Garcia, H. G., Cox, R. S., and Phillips, R. (2014b). The Influence of Promoter Architectures and Regulatory Motifs on Gene Expression in *Escherichia coli*. *PLoS ONE* 9.12, pp. 1–31.
- Schmidt, A. et al. (2015). The quantitative and condition-dependent *Escherichia coli* proteome. *Nature Biotechnology* 34.1, pp. 104–110.

Note that a complete list of references can be found in the main text associated with this work.