

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^{1,4,*}

¹Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, United States; ²Department of Physics, California Institute of Technology, Pasadena, CA, United States;

³Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA, United States; ⁴Department of Applied Physics, California Institute of Technology, Pasadena, CA, United States

† contributed equally

* phillips@pboc.caltech.edu

Contents

A	Inferring Allosteric Parameters from Previous Data	S3
A.1	Degenerate Parameter Values	S3
A.2	Computing $\Delta\varepsilon_{AI}$	S3
B	Induction of Simple Repression with Multiple Promoters or Competitor Sites	S7
B.1	Chemical Potential Formulation to Calculate Fold-Change	S7
B.2	Variable Repressor Copy Number (R) with Multiple Specific Binding Sites ($N_S > 1$) .	S8
B.3	Variable Number of Specific Binding Sites N_S with Fixed Repressor Copy Number (R)	S8
B.4	Competitor Binding Sites	S8
B.5	Properties of the Induction Response	S9
C	Flow Cytometry	S12
C.1	Equipment	S12
C.2	Experimental Measurement	S12
C.3	Unsupervised Gating	S12
C.4	Comparison of Flow Cytometry with Other Methods	S13
D	Single-Cell Microscopy	S15
D.1	Strains and Growth Conditions	S15
D.2	Imaging Procedure	S15
D.3	Image Processing	S16
D.3.1	Correcting Uneven Illumination	S16
D.3.2	Cell Segmentation	S16
D.3.3	Calculation of Fold-Change	S17
D.4	Parameter Estimation and Comparison	S18
E	Fold-Change Sensitivity Analysis	S20
F	Global Fit of All Parameters	S22
G	Applicability of Theory to the Oid Operator Sequence	S27
H	Comparison of Parameter Estimation and Fold-Change Predictions across Strains	S29
I	Properties of Induction Titration Curves	S33
J	Applications to Other Regulatory Architectures	S36
J.1	Corepression	S36
J.2	Activation	S36
K	<i>E. coli</i> Primer and Strain List	S38

A Inferring Allosteric Parameters from Previous Data

The fold-change profile described by Eq. (5) features three unknown parameters K_A , K_I , and $\Delta\varepsilon_{AI}$. In this section, we explore different conceptual approaches to determining these parameters. We first discuss how the induction titration profile of the simple repression constructs used in this paper are not sufficient to determine all three MWC parameters simultaneously, since multiple degenerate sets of parameters can produce the same fold-change response. We then utilize an additional data set from Brewster *et al.* [1] to determine the parameter $\Delta\varepsilon_{AI} = 4.5 k_B T$, after which the remaining parameters K_A and K_I can be extracted from any induction profile with no further degeneracy.

A.1 Degenerate Parameter Values

In this section, we discuss how multiple sets of parameters may yield identical fold-change profiles. More precisely, we shall show that if we try to fit the data in Fig. 5C to the fold-change Eq. (5) and extract the three unknown parameters (K_A , K_I , and $\Delta\varepsilon_{AI}$), then multiple degenerate parameter sets would yield equally good fits. In other words, this data set alone is insufficient to uniquely determine the actual physical parameter values of the system. This problem persists even when fitting multiple data sets simultaneously as in Appendix F.

In Fig. S1A, we fit the $R = 260$ data by fixing $\Delta\varepsilon_{AI}$ to the value shown on the x -axis and determine the parameters K_A and K_I given this constraint. We use the fold-change function Eq. (5) but with $\beta\Delta\varepsilon_{RA}$ modified to the form $\beta\Delta\tilde{\varepsilon}_{RA}$ in Eq. (S4) to account for the underlying assumptions used when fitting previous data (see Appendix A.2 for a full explanation of why this modification is needed).

The best-fit curves for several different values of $\Delta\varepsilon_{AI}$ are shown in Fig. S1B. Note that these fold-change curves are nearly overlapping, demonstrating that different sets of parameters can yield nearly equivalent responses. Without more data, the relationships between the parameter values shown in Fig. S1A represent the maximum information about the parameter values that can be extracted from the data. Additional experiments which independently measure any of these unknown parameters could resolve this degeneracy. For example, NMR measurements could be used to directly measure the fraction $(1 + e^{-\beta\Delta\varepsilon_{AI}})^{-1}$ of active repressors in the absence of IPTG [2,3].

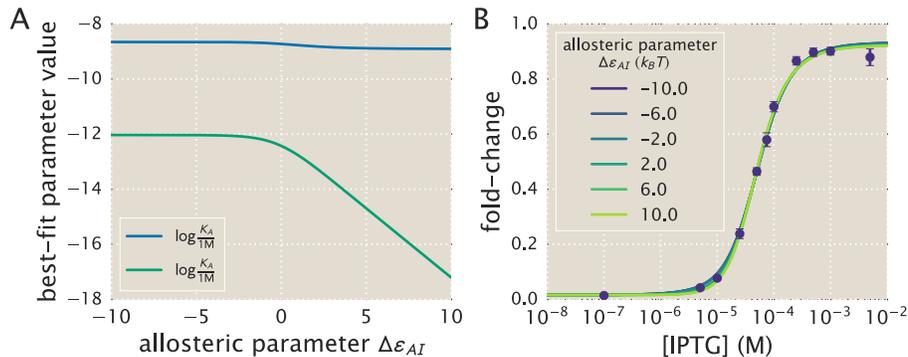


Figure S1. Multiple sets of parameters yield identical fold-change responses. (A) The data for the O2 strain ($\Delta\varepsilon_{RA} = -13.9 k_B T$) with $R = 260$ in Fig. 5C was fit using Eq. (5) with $n = 2$. $\Delta\varepsilon_{AI}$ is forced to take on the value shown on the x -axis, while the K_A and K_I parameters are fit freely. (B) The resulting best-fit functions for several value of $\Delta\varepsilon_{AI}$ all yield nearly identical fold-change responses.

A.2 Computing $\Delta\varepsilon_{AI}$

As shown in the previous section, the fold-change response of a single strain is not sufficient to determine the three MWC parameters (K_A , K_I , and $\Delta\varepsilon_{AI}$), since degenerate sets of parameters yield nearly identical fold-change responses. To circumvent this degeneracy, we now turn to some previous data from

the *lac* system in order to determine the value of $\Delta\varepsilon_{AI}$ in Eq. (5) for the induction of the Lac repressor. Specifically, we consider two previous sets of work from: (1) Garcia and Phillips [4] and (2) Brewster *et al.* [1], both of which measured fold-change with the same simple repression system in the absence of inducer ($c = 0$) but at various repressor copy numbers R . The original analysis for both data sets assumed that in the absence of inducer all of the Lac repressors were in the active state. As a result, the effective binding energies they extracted were a convolution of the DNA binding energy $\Delta\varepsilon_{RA}$ and the allosteric energy difference $\Delta\varepsilon_{AI}$ between the Lac repressor's active and inactive states. We refer to this convoluted energy value as $\Delta\tilde{\varepsilon}_{RA}$. We first disentangle the relationship between these parameters in Garcia and Phillips and then use this relationship to extract the value of $\Delta\varepsilon_{AI}$ from the Brewster *et al.* dataset.

Garcia and Phillips determined the total repressor copy numbers R of different strains using quantitative Western blots. Then they measured the fold-change at these repressor copy numbers for simple repression constructs carrying the O1, O2, O3, and Oid *lac* operators integrated into the chromosome. These data were then fit to the following thermodynamic model to determine the repressor-DNA binding energies $\Delta\tilde{\varepsilon}_{RA}$ for each operator,

$$\text{fold-change}(c = 0) = \left(1 + \frac{R}{N_{NS}} e^{-\beta\Delta\tilde{\varepsilon}_{RA}}\right)^{-1}. \quad (\text{S1})$$

Note that this functional form does not exactly match our fold-change Eq. (5) in the limit $c = 0$,

$$\text{fold-change}(c = 0) = \left(1 + \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}}} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}}\right)^{-1}, \quad (\text{S2})$$

since it is missing the factor $\frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}}}$ which specifies what fraction of repressors are in the active state in the absence of inducer,

$$\frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}}} = p_A(0). \quad (\text{S3})$$

In other words, Garcia *et al.* assumed that in the absence of inducer, all repressors were active. In terms of our notation, the convoluted energy values $\Delta\tilde{\varepsilon}_{RA}$ extracted by Garcia *et al.* (namely, $\Delta\tilde{\varepsilon}_{RA} = -15.3 k_B T$ for O1 and $\Delta\tilde{\varepsilon}_{RA} = -17.0 k_B T$ for Oid) represent

$$\beta\Delta\tilde{\varepsilon}_{RA} = \beta\Delta\varepsilon_{RA} - \log\left(\frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}}}\right). \quad (\text{S4})$$

Note that if $e^{-\beta\Delta\varepsilon_{AI}} \ll 1$, then nearly all of the repressors are active in the absence of inducer so that $\Delta\tilde{\varepsilon}_{RA} \approx \Delta\varepsilon_{RA}$. In simple repression systems where we definitively know the value of $\Delta\varepsilon_{RA}$ and R , we can use Eq. (S2) to determine the value of $\Delta\varepsilon_{AI}$ by comparing with experimentally determined fold-change values. However, the binding energy values that we use from Ref. [4] are effective parameters $\Delta\tilde{\varepsilon}_{RA}$. In this case, we are faced with an undetermined system in which we have more variables than equations, and we are thus unable to determine the value of $\Delta\varepsilon_{AI}$. In order to obtain this parameter, we must turn to a more complex regulatory scenario which provides additional constraints that allow us to fit for $\Delta\varepsilon_{AI}$.

A variation on simple repression in which multiple copies of the promoter are available for repressor binding (for instance, when the simple repression construct is on plasmid) can be used to circumvent the problems that arise when using $\Delta\tilde{\varepsilon}_{RA}$. This is because the behavior of the system is distinctly different when the number of active repressors $p_A(0)R$ is less than or greater than the number of available promoters N . Repression data for plasmids with known copy number N allows us to perform a fit for the value of $\Delta\varepsilon_{AI}$.

To obtain an expression for a system with multiple promoters N , we follow Ref. [5], writing the fold-change in terms of the the grand canonical ensemble as

$$\text{fold-change} = \frac{1}{1 + \lambda_r e^{-\beta\Delta\varepsilon_{RA}}}, \quad (\text{S5})$$

where $\lambda_r = e^{\beta\mu}$ is the fugacity and μ is the chemical potential of the repressor. The fugacity will enable us to easily enumerate the possible states available to the repressor.

To determine the value of λ_r , we first consider that the total number of repressors in the system, R_{tot} , is fixed and given by

$$R_{\text{tot}} = R_S + R_{NS}, \quad (\text{S6})$$

where R_S represents the number of repressors specifically bound to the promoter and R_{NS} represents the number of repressors nonspecifically bound throughout the genome. The value of R_S is given by

$$R_S = N \frac{\lambda_r e^{-\beta\Delta\varepsilon_{RA}}}{1 + \lambda_r e^{-\beta\Delta\varepsilon_{RA}}}, \quad (\text{S7})$$

where N is the number of available promoters in the cell. Note that in counting N , we do not distinguish between promoters that are on plasmid or chromosomally integrated provided that they both have the same repressor-operator binding energy [5]. The value of R_{NS} is similarly give by

$$R_{NS} = N_{NS} \frac{\lambda_r}{1 + \lambda_r}, \quad (\text{S8})$$

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*).

Substituting in Eqs. (S7) and (S8) into the modified Eq. (S6) yields the form

$$p_A(0)R_{\text{tot}} = \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}}} R_{\text{tot}} = N \frac{\lambda_r e^{-\beta\Delta\varepsilon_{RA}}}{1 + \lambda_r e^{-\beta\Delta\varepsilon_{RA}}} + N_{NS} \frac{\lambda_r}{1 + \lambda_r}, \quad (\text{S9})$$

where we recall from Eq. (S4) that $\beta\Delta\varepsilon_{RA} = \beta\Delta\tilde{\varepsilon}_{RA} + \log\left(\frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}}}\right)$. Numerically solving for λ_r and plugging the value back into Eq. (S5) yields a fold-change function in which the only unknown parameter is $\Delta\varepsilon_{AI}$.

Fig. S2A shows how tuning $\Delta\varepsilon_{AI}$ leads to significantly different fold-change response curves, although curves with $\Delta\varepsilon_{AI} \gtrsim 5$ will still be indistinguishable. It should be noted that these different responses occur in spite of the fact that the energy term used for these curves is the convolved energy $\Delta\tilde{\varepsilon}_{AI}$. Thus, analyzing the specific fold-change response of any strain with a known plasmid copy number N will fix $\Delta\varepsilon_{AI}$. Interestingly, the inflection point of Eq. (S9) occurs near $p_A(0)R_{\text{tot}} = N$, so that merely knowing where the fold-change response transitions from concave down to concave up is sufficient to obtain a rough value for $\Delta\varepsilon_{AI}$. In addition, once the energy gets sufficiently large ($\Delta\varepsilon_{AI} \gtrsim 5 k_B T$), nearly all of the repressors are in the active state and increasing $\Delta\varepsilon_{AI}$ further does not affect the fold-change.

Fig. S2B shows measurements of fold-change for two O1 promoters with $N = 64$ and $N = 52$ copy numbers and one Oid promoter with $N = 10$ from Brewster *et al.* [1]. By fitting these data to Eq. (S5), we extracted the parameter value $\Delta\varepsilon_{AI} = 4.5 k_B T$. Substituting this value into Eq. (S3) shows that 99% of the repressors are in the active state in the absence of inducer and $\Delta\tilde{\varepsilon}_{RA} \approx \Delta\varepsilon_{RA}$, so that all of the previous energies and calculations made by Garcia and Phillips and Brewster *et al.* [1, 4] were accurate.

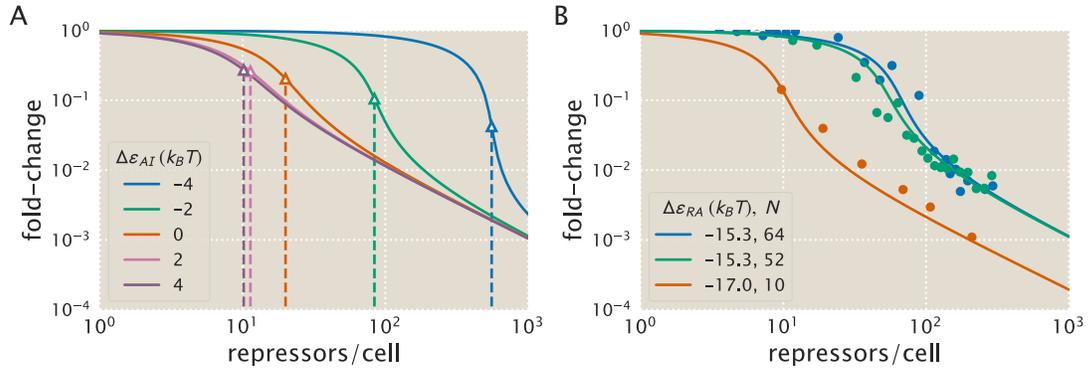


Figure S2. Fold-change of multiple identical genes. (A) In the presence of $N = 10$ identical promoters, the fold-change Eq. (S5) depends strongly on the allosteric energy difference $\Delta\epsilon_{AI}$ between the Lac repressor's active and inactive states. The vertical dotted lines represent the number of repressors at which $R_A = N$ for each value of $\Delta\epsilon_{AI}$. (B) Using fold-change measurements from [1] for the operators and gene copy numbers shown, we can determine the most likely value $\Delta\epsilon_{AI} = 4.5 k_B T$ for LacI.

B 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 [6–13], 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.

B.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 A, we can write the fold-change Eq. (2) in the grand canonical ensemble as

$$\text{fold-change} = \frac{1}{1 + \lambda_r e^{-\beta \Delta \varepsilon_{RA}}}, \quad (\text{S10})$$

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

$$R_{\text{tot}} = R_S + R_{NS} + R_C, \quad (\text{S11})$$

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

$$R_S = N_S \frac{\lambda_r e^{-\beta \Delta \varepsilon_{RA}}}{1 + \lambda_r e^{-\beta \Delta \varepsilon_{RA}}}, \quad (\text{S12})$$

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

$$R_{NS} = N_{NS} \frac{\lambda_r}{1 + \lambda_r}, \quad (\text{S13})$$

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

$$R_C = N_C \frac{\lambda_r e^{-\beta \Delta \varepsilon_C}}{1 + \lambda_r e^{-\beta \Delta \varepsilon_C}}, \quad (\text{S14})$$

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. (S11) 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)),

$$p_A(c) = \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{-\beta\Delta\epsilon_{AI}} \left(1 + \frac{c}{K_I}\right)^n}. \quad (\text{S15})$$

Substituting in Eqs. (S12)-(S14) into the modified Eq. (S11) yields the form

$$p_A(c)R_{\text{tot}} = N_S \frac{\lambda_r e^{-\beta\Delta\epsilon_{RA}}}{1 + \lambda_r e^{-\beta\Delta\epsilon_{RA}}} + N_{NS} \frac{\lambda_r}{1 + \lambda_r} + N_C \frac{\lambda_r e^{-\beta\Delta\epsilon_C}}{1 + \lambda_r e^{-\beta\Delta\epsilon_C}}. \quad (\text{S16})$$

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. (S10) 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. (S16) 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.

B.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. 6). Here we predict the induction profiles for similar strains in which R is varied, but $N_S > 1$, as shown in Fig. S3. 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 .

B.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. S4). 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.

B.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 [14, 15]. If the number of competitor sites and their average binding energy is

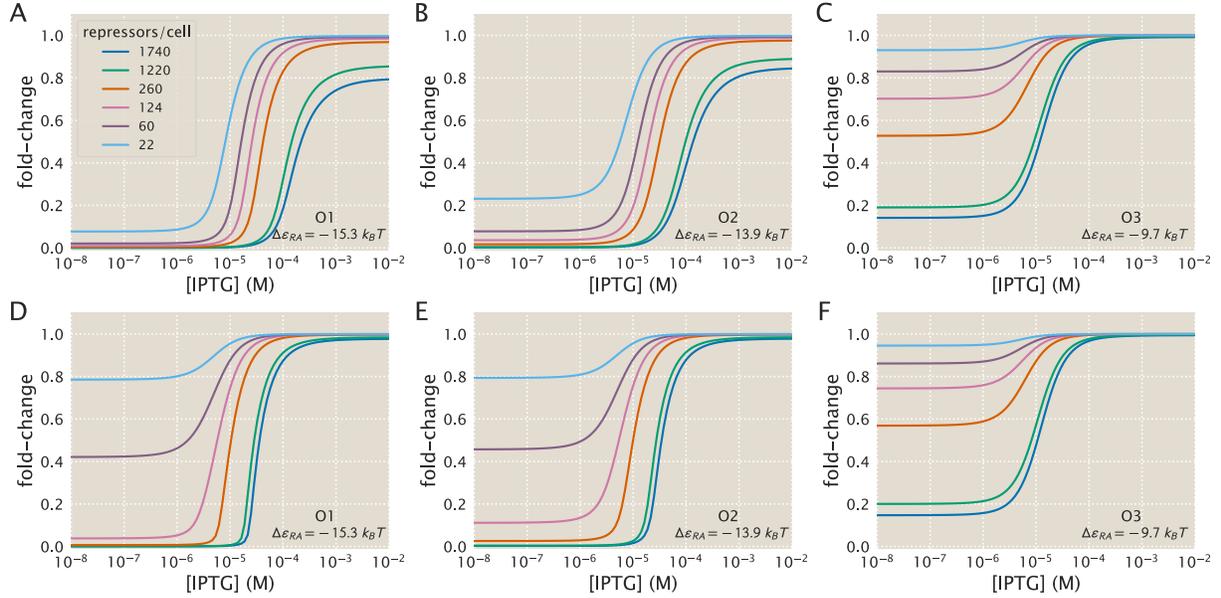


Figure S3. Induction with variable R and multiple specific binding sites. Induction profiles are shown for strains with variable R and $\Delta\epsilon_{RA} = -15.3, -13.9,$ or $-9.7 k_B T$. (A-C) The number of specific sites, N_S , is held constant at 10 as R and $\Delta\epsilon_{RA}$ are varied. (D-F) N_S is held constant at 100 as R and $\Delta\epsilon_{RA}$ are varied. These situations mimic the common scenario in which a promoter construct is placed on either a low or high copy number plasmid.

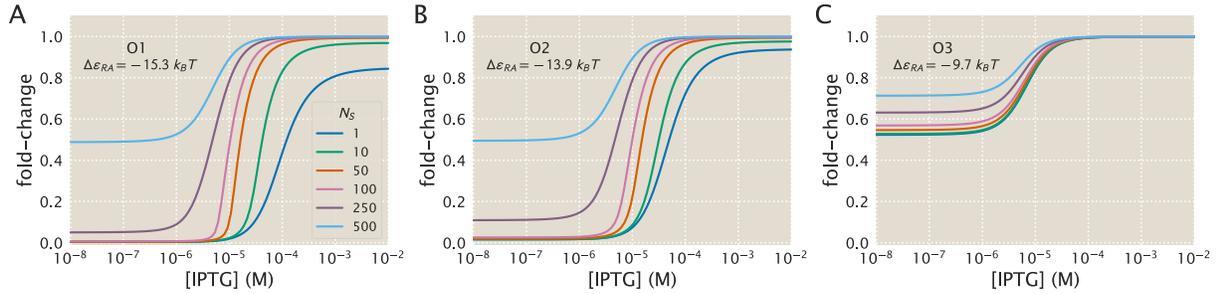


Figure S4. Induction with variable specific sites and fixed R. Induction profiles are shown for strains with $R = 260$ and (A) $\Delta\epsilon_{RA} = -15.3 k_B T$, (B) $\Delta\epsilon_{RA} = -13.9 k_B T$, or (C) $\Delta\epsilon_{RA} = -9.7 k_B T$. The number of specific sites N_S is varied from 1 to 500.

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 $\Delta\epsilon_C = -17.0 k_B T$. 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 $\Delta\epsilon_{RA} = -9.7 k_B T$, 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.

B.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. (S16) in the absence of inducer, $c = 0$, and in the presence of

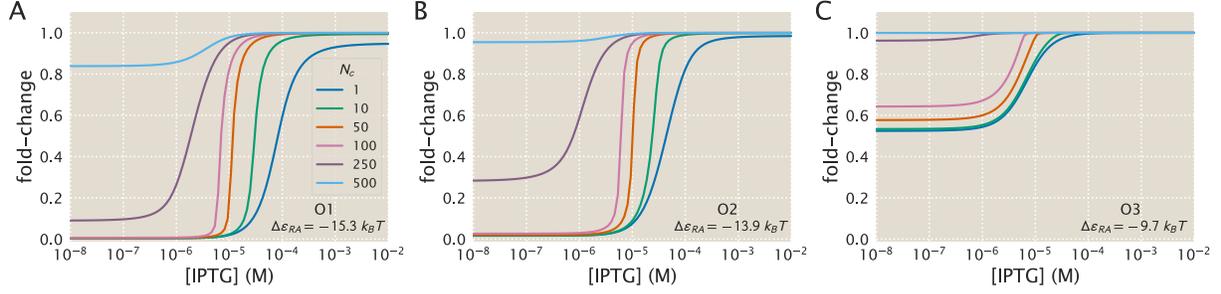


Figure S5. Induction with variable competitor sites, a single specific site, and fixed R. Induction profiles are shown for strains with $R = 260$, $N_s = 1$, and (A) $\Delta\varepsilon_{RA} = -15.3 k_B T$ for the O1 operator, (B) $\Delta\varepsilon_{RA} = -13.9 k_B T$ for the O2 operator, or (C) $\Delta\varepsilon_{RA} = -9.7 k_B T$ 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.

saturation inducer $c \rightarrow \infty$. Using Eq. (S15), the former case is given by

$$R_{\text{tot}} \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}}} = N_S \frac{\lambda_r e^{-\beta\Delta\varepsilon_{RA}}}{1 + \lambda_r e^{-\beta\Delta\varepsilon_{RA}}} + N_{NS} \frac{\lambda_r}{1 + \lambda_r} + N_C \frac{\lambda_r e^{-\beta\Delta\varepsilon_C}}{1 + \lambda_r e^{-\beta\Delta\varepsilon_C}}, \quad (\text{S17})$$

whereupon substituting in the value of λ_r into Eq. (S10) will yield the leakiness. Similarly, the limit of saturating inducer is found by determining λ_r from the form

$$R_{\text{tot}} \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^2} = N_S \frac{\lambda_r e^{-\beta\Delta\varepsilon_{RA}}}{1 + \lambda_r e^{-\beta\Delta\varepsilon_{RA}}} + N_{NS} \frac{\lambda_r}{1 + \lambda_r} + N_C \frac{\lambda_r e^{-\beta\Delta\varepsilon_C}}{1 + \lambda_r e^{-\beta\Delta\varepsilon_C}}. \quad (\text{S18})$$

In Fig. S6 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. S7 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 multiple specific sites shown in Fig. S6, 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.

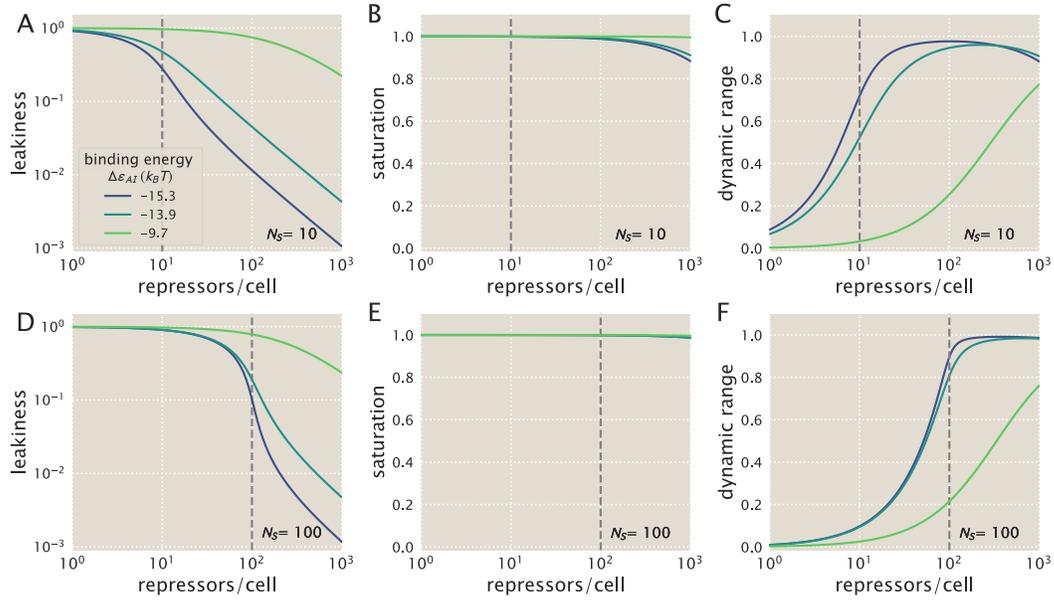


Figure S6. Phenotypic properties of induction with multiple specific binding sites. The leakiness (A, D), saturation (B, E), and dynamic range (C, F) are shown for systems with number of specific binding sites $N_S = 10$ (A-C) or $N_S = 100$ (D-F). The dashed vertical line indicates the point at which $N_S = R$.

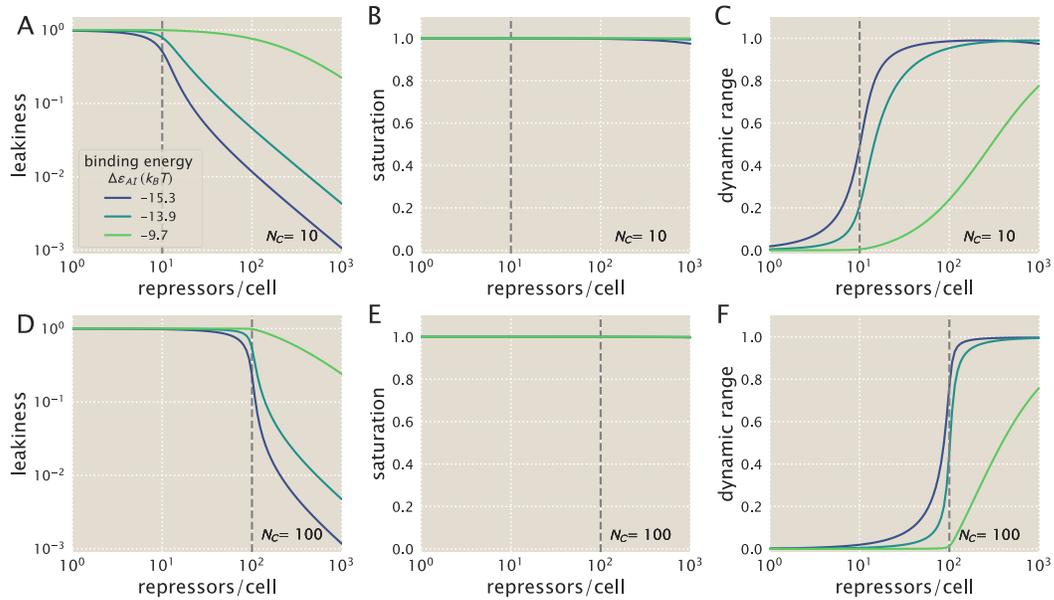


Figure S7. Phenotypic properties of induction with a single specific site and multiple competitor sites. The leakiness (A, D), saturation (B, E), and dynamic range (C, F) are shown for systems with a single specific binding site $N_S = 1$ and a number of competitor sites $N_C = 10$ (A-C) or $N_C = 100$ (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 $N_C = R$.

C Flow Cytometry

In this section, we provide information regarding the equipment used to make experimental measurements of the fold-change in gene expression in the interests of transparency and reproducibility. We also provide a summary of our unsupervised method of gating the flow cytometry measurements for consistency between experimental runs.

C.1 Equipment

Due to past experience using the Miltenyi Biotec MACSQuant flow cytometer during the Physiology summer course at the Marine Biological Laboratory, we used the same flow cytometer for the formal measurements in this work. All measurements were made using an excitation wavelength of 488 nm with an emission filter set of 525/50 nm. This excitation wavelength provides approximately 40% of the maximum YFP absorbance [16], and this was found to be sufficient for the purposes of these experiments. A useful feature of modern flow cytometry is the high-sensitivity signal detection through the use of photomultiplier tubes (PMT) whose response can be tuned by adjusting the voltage. Thus, the voltage for the forward-scatter (FSC), side-scatter (SSC), and gene expression measurements were tuned manually to maximize the dynamic range between autofluorescence signal and maximal expression without losing the details of the population distribution. Once these voltages were determined, they were used for all subsequent measurements. Extremely low signal producing particles were discarded before data storage by setting a basal voltage threshold, thus removing the majority of spurious events. The various instrument settings for data collection are given in Table S1.

Table S1. Instrument settings for data collection using the Miltenyi Biotec MACSQuant flow cytometer. All experimental measurements were collected using these values.

Laser	Channel	Sensor Voltage
488 nm	Forward-Scatter (FSC)	423 V
488 nm	Side-Scatter (SSC)	537 V
488 nm	Intensity (B1 Filter, 525/50nm)	790 V
488 nm	Trigger (debris threshold)	24.5 V

C.2 Experimental Measurement

A single data set consisted of seven bacterial strains, all sharing the same operator, with varying repressor copy numbers ($R = 0, 22, 60, 124, 260, 1220, \text{ and } 1740$), in addition to an autofluorescent strain, under twelve IPTG concentrations. Data collection took place over two to three hours. During this time, the cultures were held at approximately 4°C by placing the 96-well plate on a MACSQuant ice block. Because the ice block thawed over the course of the experiment, the samples measured last were approximately at room temperature. This means that samples may have grown slightly by the end of the experiment. To confirm that this continued growth did not alter the measured results, a subset of experiments were run in reverse meaning that the fully induced cultures were measured first and the uninduced samples last. The plate arrangements and corresponding fold-change measurements are shown in Fig. S8A and Fig. S8B, respectively. The measured fold-change values in the reverse ordered plate appear to be drawn from the same distribution as those measured in the forward order, meaning that any growth that might have taken place during the experiment did not significantly affect the results. Both the forward and reverse data sets were used in our analysis.

C.3 Unsupervised Gating

As explained in the Methods, we used an automatic unsupervised gating procedure to filter the flow cytometry data based on the front and side-scattering values returned by the MACSQuant flow cytometer.

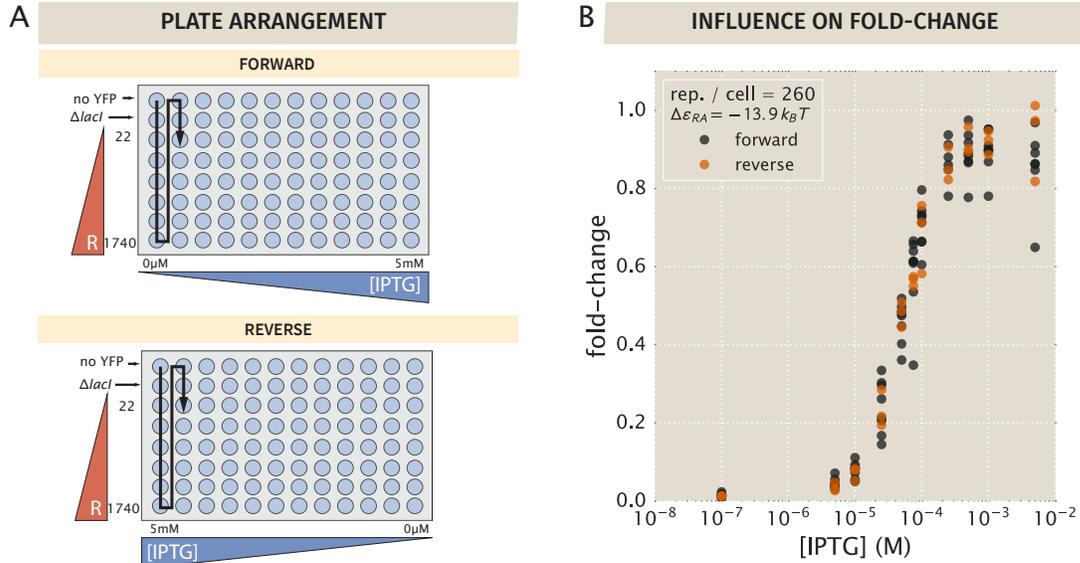


Figure S8. Plate arrangements for flow cytometry. (A) Samples were measured primarily in the forward arrangement with a subset of samples measured in reverse. The black arrow indicates the order in which samples were processed by the flow cytometer. (B) The experimentally measured fold-change values for the two sets of plate arrangements show that samples measured in the forward arrangement appear to be indistinguishable from those measured in reverse order.

We assume that the region with highest density of points in these two channels corresponds to single-cell measurements. Everything extending outside of this region was discarded in order to exclude sources of error such as cell clustering, particulates, or other spurious events.

In order to define the gated region we fit a two-dimensional Gaussian function to the \log_{10} forward-scattering (FSC) and the \log_{10} side-scattering (SSC) data. We then kept a fraction $\alpha \in [0, 1]$ of the data by defining an elliptical region given by

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \leq \chi_{\alpha}^2(p), \quad (\text{S19})$$

where \mathbf{x} is the 2×1 vector containing the $\log(\text{FSC})$ and $\log(\text{SSC})$, $\boldsymbol{\mu}$ is the 2×1 vector representing the mean values of $\log(\text{FSC})$ and $\log(\text{SSC})$ as obtained from fitting a two-dimensional Gaussian to the data, and $\boldsymbol{\Sigma}$ is the 2×2 covariance matrix also obtained from the Gaussian fit. $\chi_{\alpha}^2(p)$ is the quantile function for probability p of the chi-squared distribution with two degrees of freedom. Fig. S9 shows an example of different gating contours that would arise from different values of α in Eq. (S19). In this work, we chose $\alpha = 0.4$ which we deemed was a sufficient constraint to minimize the noise in the data. As explained in Appendix D we compared our high throughput flow cytometry data with single cell microscopy, confirming that the automatic gating did not introduce systematic biases to the analysis pipeline. The specific code where this gating is implemented can be found in [GitHub repository](#).

C.4 Comparison of Flow Cytometry with Other Methods

Previous work from our lab experimentally determined fold-change for similar simple repression constructs using a variety of different measurement methods [1, 17]. Garcia and Phillips used the same background strains as the ones used in this work, but gene expression was measured with Miller assays based on colorimetric enzymatic reactions with the LacZ protein [4]. Brewster *et al.* used a LacI dimer with the tetramerization region replaced with an mCherry tag, where the fold-change was measured as the ratio of the gene expression rate rather than a single snapshot of the gene output [1].

Fig. S10 shows the comparison of these methods along with the flow cytometry method used in this work. The consistency of these three readouts validates the quantitative use of flow cytometry

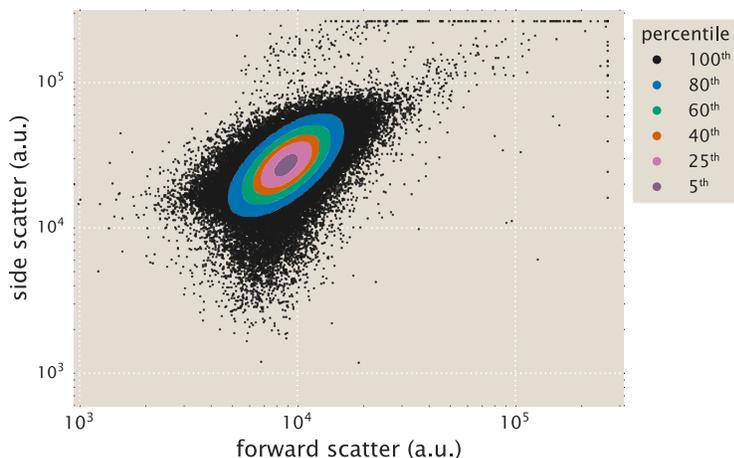


Figure S9. Representative unsupervised gating contours. Points indicate individual flow cytometry measurements of forward scatter and side scatter. Colored points indicate arbitrary gating contours ranging from 100% ($\alpha = 1.0$) to 5% ($\alpha = 0.05$). All measurements for this work were made computing the mean fluorescence from the 40th percentile ($\alpha = 0.4$), shown as orange points.

and unsupervised gating to determine the fold-change in gene expression. However, one important caveat revealed by this figure is that the sensitivity of flow cytometer measurements is not sufficient to accurately determine the fold-change for the high repressor copy number strains in O1 without induction. Instead, a method with a large dynamic range such as the Miller assay is needed to accurately resolve the fold-change at such low expression levels.

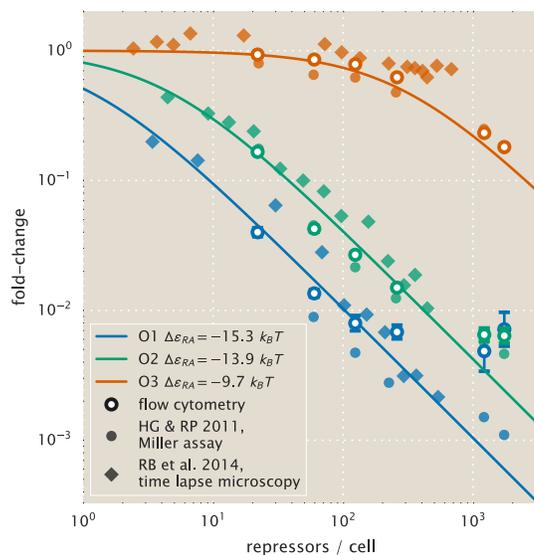


Figure S10. Comparison of experimental methods to determine the fold-change. The fold-change in gene expression for equivalent simple-repression constructs has been determined using three independent methods: flow cytometry (this work), colorimetric Miller assays [4], and video microscopy [1]. All three methods give consistent results, although flow cytometry measurements lose accuracy for fold-change less than 10^{-2} . Note that the repressor-DNA binding energies $\Delta\epsilon_{RA}$ used for the theoretical predictions were determined in [4].

D 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 [4] or through single-cell microscopy [1]. 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. S11.

D.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.

D.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 [18]. 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 upon request.

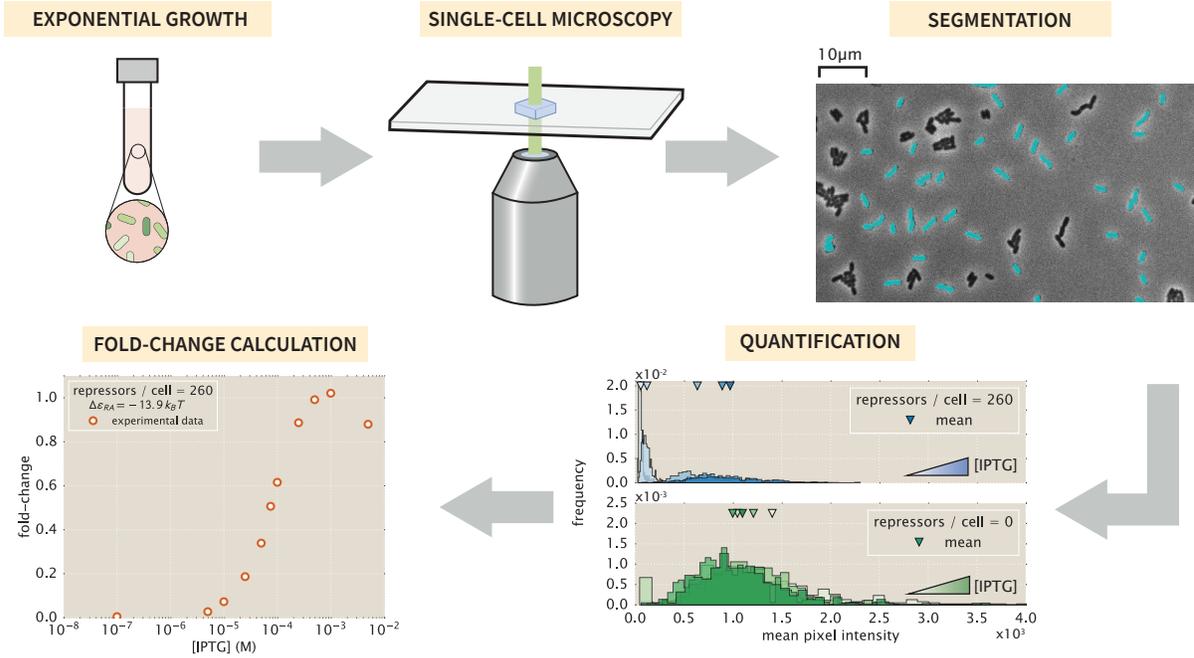


Figure S11. Experimental workflow for single-cell microscopy. For comparison with the flow cytometry results, the cells were grown in an identical manner to those described in the main text. Once cells had reached mid to late exponential growth, the cultures were diluted and placed on agarose substrates 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.

D.3 Image Processing

D.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

$$I_{\text{flat}} = \frac{I - I_{\text{dark}}}{I_{\text{YFP}} - I_{\text{dark}}} \langle I_{\text{YFP}} - I_{\text{dark}} \rangle, \quad (\text{S20})$$

where I_{flat} is the renormalized image and I is the original fluorescence image. An example of this correction can be seen in Fig. S12.

D.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 [19] which identifies edges by taking the second

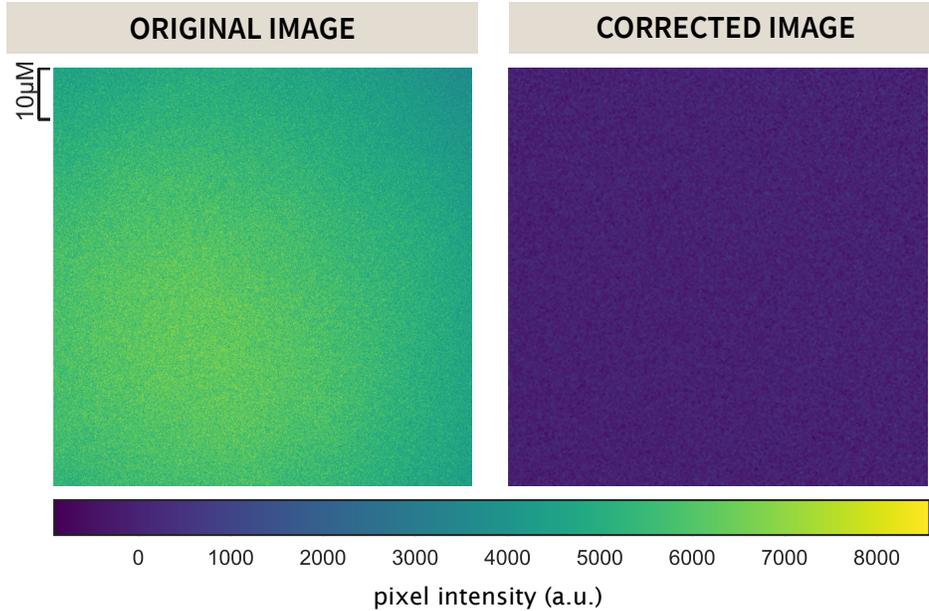


Figure S12. Correction for uneven illumination. A representative image of the illumination profile of the 512 nm excitation beam on a homogeneously fluorescent slide is shown in the left panel. This is corrected for using equation Eq. (S20) and is shown in the right panel.

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

$$\text{eccentricity} = \sqrt{1 - \left(\frac{b}{a}\right)^2}, \quad (\text{S21})$$

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. S13A) and eccentricity filter (Fig. S13B), the remaining objects were considered cells of interest (Fig. S13C) and the mean fluorescence intensity of each cell was extracted.

D.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

$$\text{fold-change} = \frac{\langle I_{R>0} \rangle - \langle I_{\text{auto}} \rangle}{\langle I_{R=0} \rangle - \langle I_{\text{auto}} \rangle}, \quad (\text{S22})$$

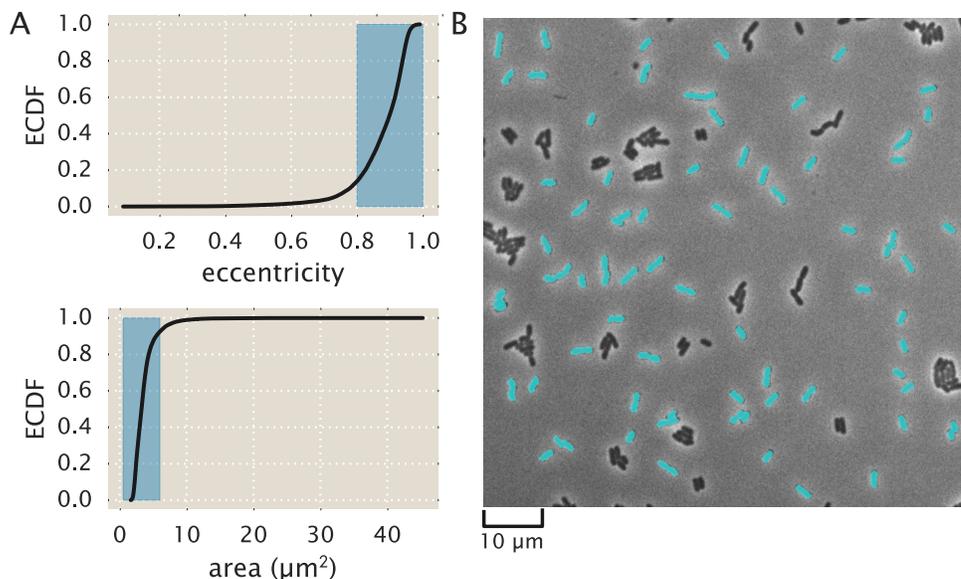


Figure S13. Segmentation of single bacterial cells. (A) Objects were selected if they had an eccentricity greater than 0.8 and an area between $0.5 \mu\text{m}^2$ and $6 \mu\text{m}^2$. Highlighted in blue are the regions considered to be representative of single cells. The black lines correspond to 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.

D.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. S14A 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. S14B. Specifically, these values were $K_A = 142_{-34}^{+40} \mu\text{M}$ and $K_I = 0.6_{-0.1}^{+0.1} \mu\text{M}$ from microscopy and $K_A = 149_{-12}^{+14} \mu\text{M}$ and $K_I = 0.57_{-0.02}^{+0.03} \mu\text{M}$ from the flow cytometry data. We note that the credible regions from the microscopy data shown in Fig. S14B are much broader than those from flow cytometry due to the fewer number of replicates performed.

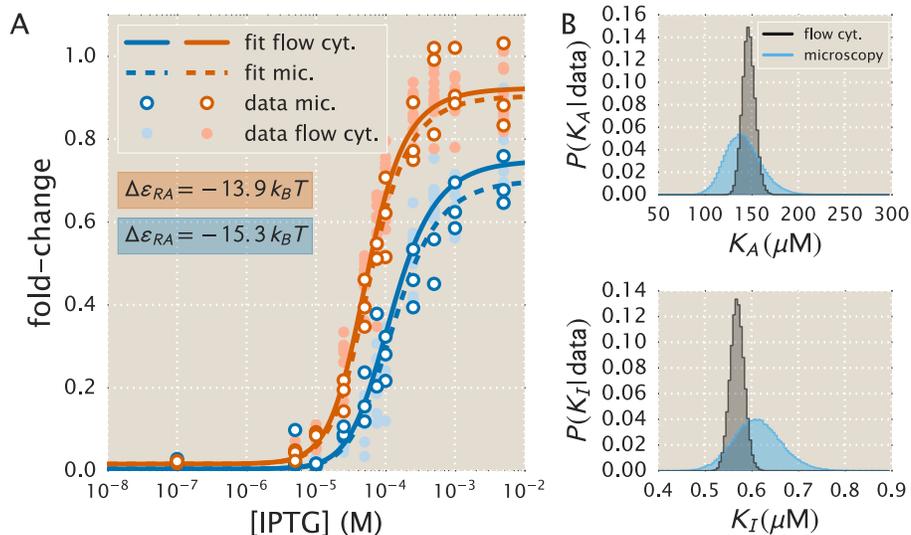


Figure S14. Comparison of measured fold-change between flow cytometry and single-cell microscopy. (A) Experimentally measured fold-change values obtained through single-cell microscopy and flow cytometry are shown as white filled and solid colored circles, respectively. Solid and dashed lines indicate 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 [4]. (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).

E Fold-Change Sensitivity Analysis

In Fig. 6 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. 6 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

$$\Delta \text{fold-change}_{K_A} \equiv \text{fold-change}(K_A, K_I^{\text{fit}}) - \text{fold-change}(K_A^{\text{fit}}, K_I^{\text{fit}}), \quad (\text{S23})$$

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

$$\Delta \text{fold-change}_{K_I} = \text{fold-change}(K_A^{\text{fit}}, K_I) - \text{fold-change}(K_A^{\text{fit}}, K_I^{\text{fit}}) \quad (\text{S24})$$

to measure the sensitivity of the fold-change to K_I at a fixed K_A^{fit} . Fig. S15 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

$$\frac{\partial \text{fold-change}}{\partial K_A} = \frac{e^{-\beta\Delta\varepsilon_{AI}} \frac{n}{K_I} \left(\frac{K_A}{K_I}\right)^{n-1}}{\left(1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^n\right)^2} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}} \left(1 + \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^n} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}}\right)^{-2}. \quad (\text{S25})$$

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

$$\frac{\partial \text{fold-change}}{\partial K_I} = -\frac{e^{-\beta\Delta\varepsilon_{AI}} \frac{n}{K_I} \left(\frac{K_A}{K_I}\right)^n}{\left(1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^n\right)^2} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}} \left(1 + \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^n} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}}\right)^{-2}. \quad (\text{S26})$$

From Eqs. (S25) and (S26), 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. S15.

Lastly, we note that Eqs. (S25) and (S26) 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.

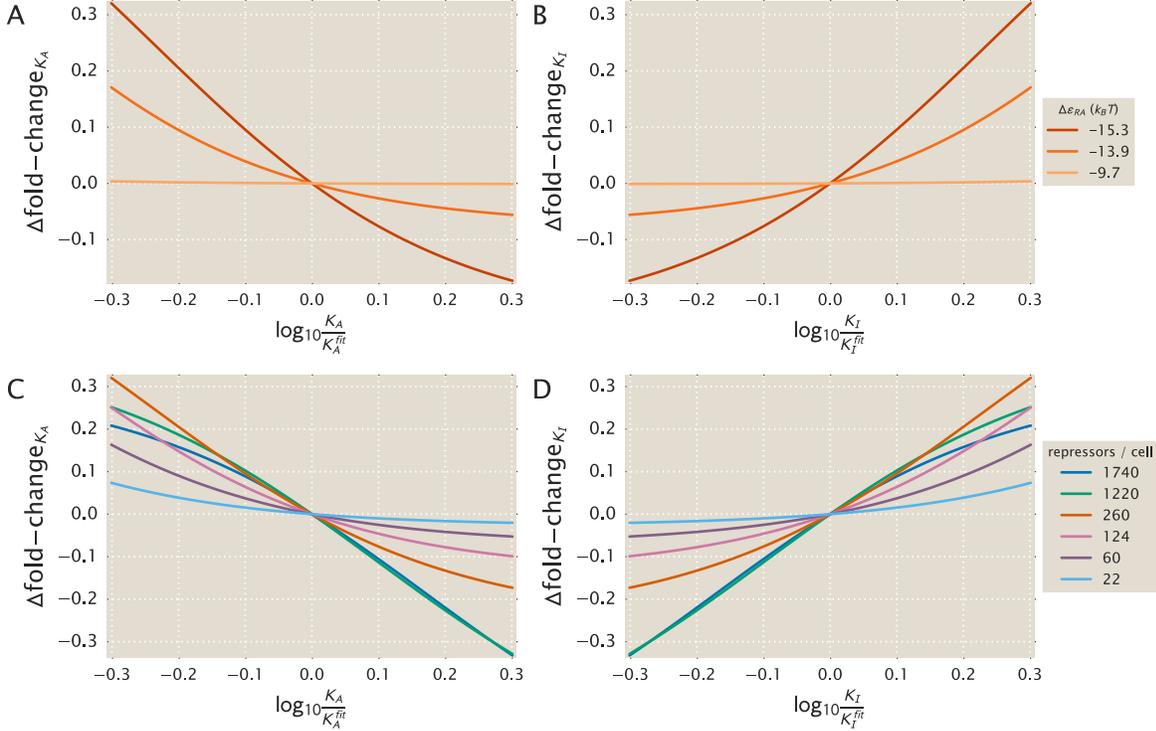


Figure S15. Determining how sensitive the fold-change values are to the fit values of the dissociation constants. (A) The difference $\Delta \text{fold-change}_{K_A}$ in fold change when the dissociation constant K_A is slightly offset from its best-fit value $K_A = 139^{+29}_{-22} \times 10^{-6}$ M, as given by Eq. (S23). Fold-change is computed in the limit of saturating inducer concentration ($c \rightarrow \infty$, see Eq. (7)) where the credible regions in Fig. 6 are widest. The O3 strain ($\Delta \epsilon_{RA} = -9.7 k_B T$) 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. 6 have tighter credible regions. All curves were made using $\Delta \epsilon_{RA} = -13.9 k_B T$. (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.

F Global Fit of All Parameters

In the main text, we used the repressor copy numbers R and repressor-DNA binding energies $\Delta\varepsilon_{RA}$ as reported by Garcia and Phillips [4]. However, any error in these previous measurements of R and $\Delta\varepsilon_{RA}$ will necessarily propagate into our own fold-change predictions. In this section we take an alternative approach to fitting the physical parameters of the system to that used in the main text. First, rather than fitting only a single strain, we fit the entire data set in Fig. 6 along with microscopy data for the synthetic operator Oid (see Appendix G). In addition, we also simultaneously fit the parameters R and $\Delta\varepsilon_{RA}$ using the prior information given by the previous measurements. By using the entire data set and fitting all of the parameters, we obtain the best possible characterization of the statistical mechanical parameters of the system given our current state of knowledge. As a point of reference, we state all of the parameters of the MWC model derived in the text in Table S2.

To fit all of the parameters simultaneously, we follow a similar approach to the one detailed in the Methods section. Briefly, we perform a Bayesian parameter estimation of the dissociation constants K_A and K_I , the six different repressor copy numbers R corresponding to the six *lacI* ribosomal binding sites used in our work, and the four different binding energies $\Delta\varepsilon_{RA}$ characterizing the four distinct operators used to make the experimental strains. As in the main text, we fit the logarithms $\tilde{k}_A = -\log \frac{K_A}{1M}$ and $\tilde{k}_I = -\log \frac{K_I}{1M}$ of the dissociation constants which grants better numerical stability.

As in Eqs. (15) and (16), we assume that deviations of the experimental fold-change from the theoretical predictions are normally distributed with mean zero and standard deviation σ . We begin by writing Bayes' theorem,

$$P(\tilde{k}_A, \tilde{k}_I, \mathbf{R}, \Delta\varepsilon_{RA}, \sigma | D) = \frac{P(D | \tilde{k}_A, \tilde{k}_I, \mathbf{R}, \Delta\varepsilon_{RA}, \sigma) P(\tilde{k}_A, \tilde{k}_I, \mathbf{R}, \Delta\varepsilon_{RA}, \sigma)}{P(D)}, \quad (\text{S27})$$

where \mathbf{R} is an array containing the six different repressor copy numbers to be fit, $\Delta\varepsilon_{RA}$ is an array containing the four binding energies to be fit, and D is the experimental fold-change data. The term $P(\tilde{k}_A, \tilde{k}_I, \mathbf{R}, \Delta\varepsilon_{RA}, \sigma | D)$ gives the probability distributions of all of the parameters given the data. The term $P(D | \tilde{k}_A, \tilde{k}_I, \mathbf{R}, \Delta\varepsilon_{RA}, \sigma)$ represents the likelihood of having observed our experimental data given some value for each parameter. $P(\tilde{k}_A, \tilde{k}_I, \mathbf{R}, \Delta\varepsilon_{RA}, \sigma)$ contains all the prior information on the values of these parameters. Lastly, $P(D)$ serves as a normalization constant and hence can be ignored.

Given n independent measurements of the fold-change, the first term in Eq. (S27) can be written as

$$P(D | \tilde{k}_A, \tilde{k}_I, \mathbf{R}, \Delta\varepsilon_{RA}, \sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \prod_{i=1}^n \exp \left[-\frac{(\text{fc}_{\text{exp}}^{(i)} - \text{fc}(\tilde{k}_A, \tilde{k}_I, R^{(i)}, \Delta\varepsilon_{RA}^{(i)}, c^{(i)}))^2}{2\sigma^2} \right], \quad (\text{S28})$$

where $\text{fc}_{\text{exp}}^{(i)}$ is the i^{th} experimental fold-change and $\text{fc}(\dots)$ is the theoretical prediction. Note that the standard deviation σ of this distribution is not known and hence needs to be included as a parameter to be fit.

The second term in Eq. (S27) represents the prior information of the parameter values. We assume that all parameters are independent of each other, so that

$$P(\tilde{k}_A, \tilde{k}_I, \mathbf{R}, \Delta\varepsilon_{RA}, \sigma) = P(\tilde{k}_A) \cdot P(\tilde{k}_I) \cdot \prod_i P(R^{(i)}) \cdot \prod_j P(\Delta\varepsilon_{RA}^{(j)}) \cdot P(\sigma), \quad (\text{S29})$$

where the superscript (i) indicates the repressor copy number of index i and the superscript (j) denotes the binding energy of index j . As above, we note that a prior must also be included for the unknown parameter σ .

Because we knew nothing about the values of \tilde{k}_A , \tilde{k}_I , and σ before performing the experiment, we assign maximally uninformative priors to each of these parameters. More specifically, we assign uniform priors to \tilde{k}_A and \tilde{k}_I and a Jeffreys prior to σ , indicating that K_A , K_I , and σ are scale parameters [20]. We do, however, have prior information for the repressor copy numbers and the repressor-DNA binding energies from Ref. [4]. This prior knowledge is included within our model using an informative prior for

these two parameters, which we assume to be Gaussian. Hence each of the $R^{(i)}$ repressor copy numbers to be fit satisfies

$$P(R^{(i)}) = \frac{1}{\sqrt{2\pi\sigma_{R_i}^2}} \exp\left(-\frac{(R^{(i)} - \bar{R}^{(i)})^2}{2\sigma_{R_i}^2}\right), \quad (\text{S30})$$

where $\bar{R}^{(i)}$ is the mean repressor copy number and σ_{R_i} is the variability associated with this parameter as reported in Ref. [4]. Note that we use the given value of σ_{R_i} from previous measurements rather than leaving this as a free parameter.

Similarly, the binding energies $\Delta\varepsilon_{RA}^{(j)}$ are also assumed to have a Gaussian informative prior of the same form. We write it as

$$P(\Delta\varepsilon_{RA}^{(j)}) = \frac{1}{\sqrt{2\pi\sigma_{\varepsilon_j}^2}} \exp\left(-\frac{(\Delta\varepsilon_{RA}^{(j)} - \bar{\Delta\varepsilon}_{RA}^{(j)})^2}{2\sigma_{\varepsilon_j}^2}\right), \quad (\text{S31})$$

where $\bar{\Delta\varepsilon}_{RA}^{(j)}$ is the binding energy and σ_{ε_j} is the variability associated with that parameter around the mean value as reported in Ref. [4].

The σ_{R_i} and σ_{ε_j} parameters will constrain the range of values for $R^{(i)}$ and $\Delta\varepsilon_{RA}^{(j)}$ found from the fitting. For example, if for some i the standard deviation σ_{R_i} is very small, it implies a strong confidence in the previously reported value. Mathematically, the exponential in Eq. (S30) will ensure that the best-fit $R^{(i)}$ lies within a few standard deviations of $\bar{R}^{(i)}$. Since we are interested in exploring which values could give the best fit, the errors are taken to be wide enough to allow the parameter estimation to freely explore parameter space in the vicinity of the best estimates. Putting all these terms together, we use Markov chain Monte Carlo to sample the posterior distribution $P(\tilde{k}_A, k_I, \mathbf{R}, \mathbf{\Delta\varepsilon}_{RA}, \sigma \mid D)$, enabling us to determine both the most likely value for each physical parameter as well as its associated credible region (see the [GitHub repository](#) for the implementation).

Fig. S16 shows the result of this global fit. When compared with Fig. 6 we can see that fitting for the binding energies and the repressor copy numbers improves the agreement between the theory and the data. Table S3 summarizes the values of the parameters as obtained with this MCMC parameter inference. We note that even though we allowed the repressor copy numbers and repressor-DNA binding energies to vary, the resulting fit values were very close to the previously reported values. The fit values of the repressor copy numbers were all within one standard deviation of the previous reported values provided in Ref. [4]. And although some of the repressor-DNA binding energies differed by a few standard deviations from the reported values, the differences were always less than $1 k_B T$, which represents a small change in the biological scales we are considering. The biggest discrepancy between our fit values and the previous measurements arose for the synthetic Oid operator, which we discuss in more detail in Appendix G.

Fig. S17 shows the same key properties as in Fig. 7, but uses the parameters obtained from this global fitting approach. We note that even by increasing the number of degrees of freedom in our fit, the result does not change substantially, due to in general, only minor improvements between the theoretical curves and data. For the O3 operator data, again, agreement between the predicted $[EC_{50}]$ and the effective Hill coefficient remain poor due the theory being unable to capture the steepness of the response curves.

Table S2. 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
$\Delta\varepsilon_{RA}, \Delta\varepsilon_{RI}$	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

Table S3. Global fit of all parameter values using the entire data set in Fig. 6. In addition to fitting the repressor inducer dissociation constants K_A and K_I as was done in the text, we also fit the repressor DNA binding energy $\Delta\varepsilon_{RA}$ as well as the repressor copy numbers R for each strain. The middle columns show the previously reported values for all $\Delta\varepsilon_{RA}$ and R values, with \pm representing the standard deviation of three replicates. The right column shows the global fits from this work, with the subscript and superscript notation denoting the 95% credible region. Note that there is overlap between all of the repressor copy numbers and that the net difference in the repressor-DNA binding energies is less than $1 k_B T$. The logarithms $\tilde{k}_A = -\log \frac{K_A}{1\text{M}}$ and $\tilde{k}_I = -\log \frac{K_I}{1\text{M}}$ of the dissociation constants were fit for numerical stability.

	Reported Values [4]	Global Fit
\tilde{k}_A	–	$-5.33^{+0.06}_{-0.05}$
\tilde{k}_I	–	$0.31^{+0.05}_{-0.06}$
K_A	–	$205^{+11}_{-12} \mu\text{M}$
K_I	–	$0.73^{+0.04}_{-0.04} \mu\text{M}$
R_{22}	22 ± 4	20^{+1}_{-1}
R_{60}	60 ± 20	74^{+4}_{-3}
R_{124}	124 ± 30	130^{+6}_{-6}
R_{260}	260 ± 40	257^{+9}_{-11}
R_{1220}	1220 ± 160	1191^{+32}_{-55}
R_{1740}	1740 ± 340	1599^{+75}_{-87}
O1 $\Delta\varepsilon_{RA}$	$-15.3 \pm 0.2 k_B T$	$-15.2^{+0.1}_{-0.1} k_B T$
O2 $\Delta\varepsilon_{RA}$	$-13.9 \pm 0.2 k_B T$	$-13.6^{+0.1}_{-0.1} k_B T$
O3 $\Delta\varepsilon_{RA}$	$-9.7 \pm 0.1 k_B T$	$-9.4^{+0.1}_{-0.1} k_B T$
Oid $\Delta\varepsilon_{RA}$	$-17.0 \pm 0.2 k_B T$	$-17.7^{+0.2}_{-0.1} k_B T$

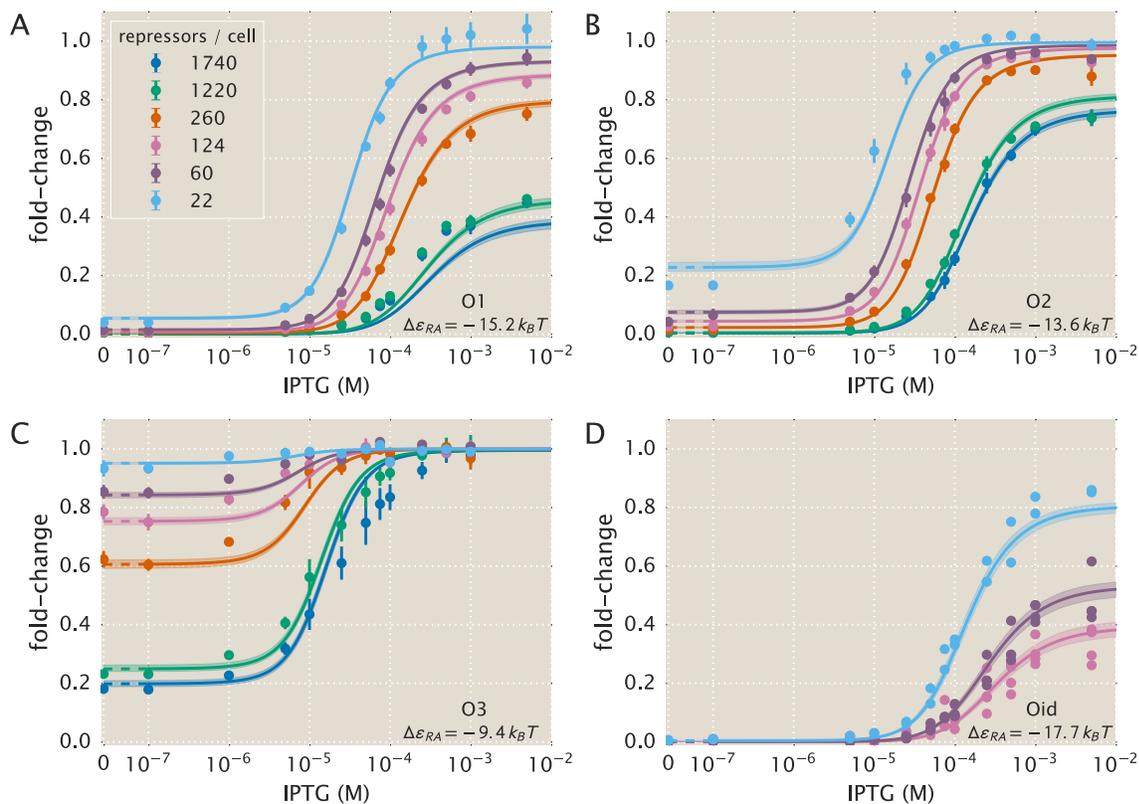


Figure S16. Global fit of dissociation constants, repressor copy numbers and binding energies. Theoretical predictions resulting from simultaneously fitting the dissociation constants K_A and K_I , the six repressor copy numbers R , and the four repressor-DNA binding energies $\Delta\epsilon_{RA}$ using the entire data set from Fig. 6 as well as the microscopy data for the Oid operator. 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 error bars are not visible, they are smaller than the point itself. For the Oid operator, all of the data points are shown since a smaller number of replicates were taken. The shaded regions are significantly smaller than in Fig. 6 because this fit was based on all data points, and hence the fit parameters are much more tightly constrained. The dashed lines at 0 IPTG indicates a linear scale, whereas solid lines represent a log scale.

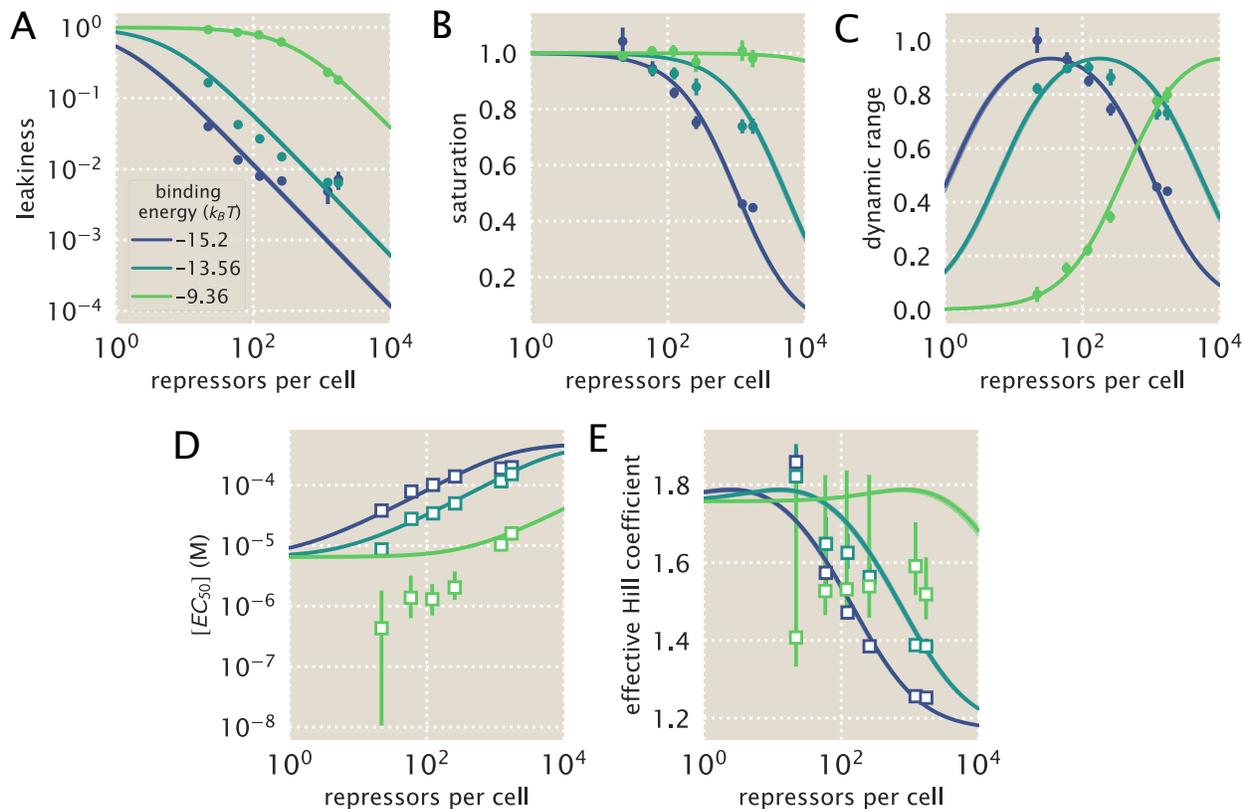


Figure S17. Key properties of induction profiles as predicted with a global fit using all available data. Data for the (A) leakiness, (B) saturation, and (C) dynamic range are obtained from fold-change measurements in Fig. 6 in the absence and presence of IPTG. All prediction curves were generated using the parameters listed in S3. Both the (D) $[EC_{50}]$ and (E) effective Hill coefficient are inferred by individually fitting all parameters $-K_A, K_I, R, \Delta\epsilon_{RA}$ to each operator-repressor pairing in Fig. 6A-C separately to Eq. (5) in order to smoothly interpolate between the data points. Note that where error bars are not visible, this indicates that the error bars are smaller than the point itself.

G 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 ($\Delta\varepsilon_{RA} = -17.0 k_B T$) [17]. Accordingly, fluorescence data for Oid were obtained using microscopy, which is more sensitive than flow cytometry. Appendix D 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. S18A, 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 F) 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 [4]. In Fig. S18B 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.

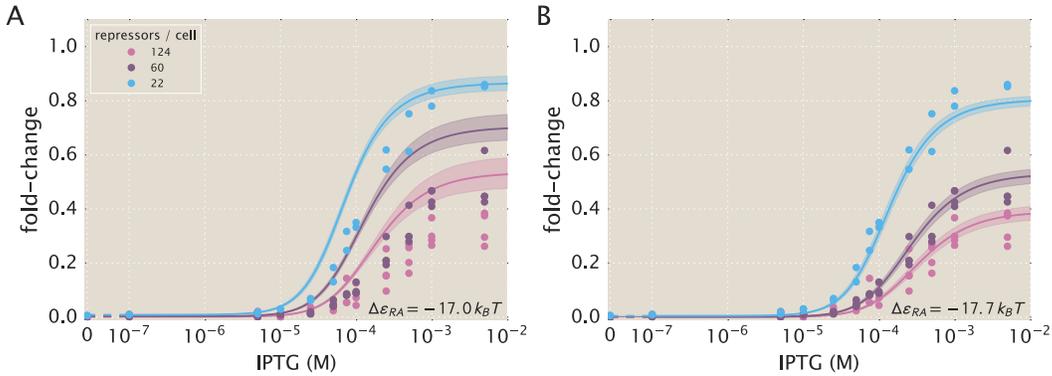


Figure S18. Predictions of fold-change for strains with an Oid binding sequence versus experimental measurements with different repressor copy numbers. (A) Experimental data is plotted against the parameter-free predictions that are based on our fit to the O2 strain with $R = 260$. Here we use the previously measured binding energy $\Delta\varepsilon_{RA} = -17.0 k_B T$ [4]. (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 F). 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. S19 shows the cumulative data from Garcia and Phillips [4] and Brewster *et al.* [1], 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 $\Delta\varepsilon_{RA} = -17.7 k_B T$, compare reasonably well with all previous measurements.

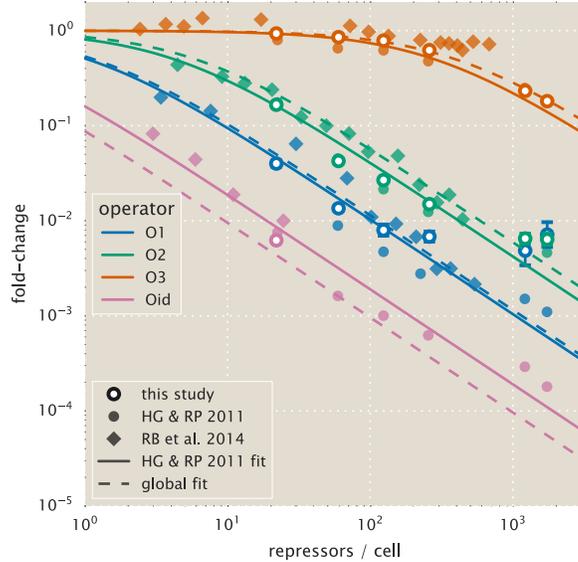


Figure S19. Comparison of fold-change predictions based on binding energies from Garcia and Phillips and those inferred from this work. Fold-change curves for the different repressor-DNA binding energies $\Delta\epsilon_{RA}$ are plotted as a function of repressor copy number when IPTG concentration $c = 0$. Solid curves use the binding energies determined from Garcia and Phillips [4], 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) [4] (solid circles), and [1] (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. S18 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.

H Comparison of Parameter Estimation and Fold-Change Predictions across Strains

The inferred parameter values for K_A and K_I in the main text were determined by fitting to induction fold-change measurements from a single strain ($R = 260$, $\Delta\varepsilon_{RA} = -13.9 k_B T$, $n = 2$, and $\Delta\varepsilon_{AI} = 4.5 k_B T$). After determining these parameters, we were able to predict the fold-change of the remaining strains without any additional fitting. However, the theory should be independent of the specific strain used to estimate K_A and K_I ; using any alternative strain to fit K_A and K_I should yield similar predictions. For the sake of completeness, here we discuss the values for K_A and K_I that are obtained by fitting to each of the induction data sets individually. These fit parameters are shown in Fig. 6D of the main text, where we find close agreement between strains, but with some deviation and poorer inferences observed with the O3 operator strains. Overall, we find that regardless of which strain is chosen to determine the unknown parameters, the predictions laid out by the theory closely match the experimental measurements. Here we present a comparison of the strain specific predictions and measured fold-change data for each of the three operators considered.

We follow the approach taken in the main text and use Eq. (5) to infer values for K_A and K_I by fitting to each combination of binding energy $\Delta\varepsilon_{RA}$ and repressor copy number R . We then use these fitted parameters to predict the induction curves of all other strains. In Fig. S20 we plot these fold-change predictions along with experimental data for each of our strains that contains an O1 operator. To make sense of this plot consider the first row as an example. In the first row, K_A and K_I were estimated using data from the strain containing $R = 22$ and an O1 operator (top leftmost plot, shaded in gray). The remaining plots in this row show the predicted fold-change using these values for K_A and K_I . In each row, we then infer K_A and K_I using data from a strain containing a different repressor copy number ($R = 60$ in the second row, $R = 124$ in the third row, and so on). In Fig. S21 and Fig. S22, we similarly apply this inference to our strains with O2 and O3 operators, respectively. We note that the overwhelming majority of predictions closely match the experimental data. The notable exception is that using the $R = 22$ strain provides poor predictions for the strains with large copy numbers (especially $R = 1220$ and $R = 1740$), though it should be noted that predictions made from the $R = 22$ strain have considerably broader credible regions. This loss in predictive power is due to the poorer estimates of K_A and K_I for the $R = 22$ strain as shown in Fig. 6D.

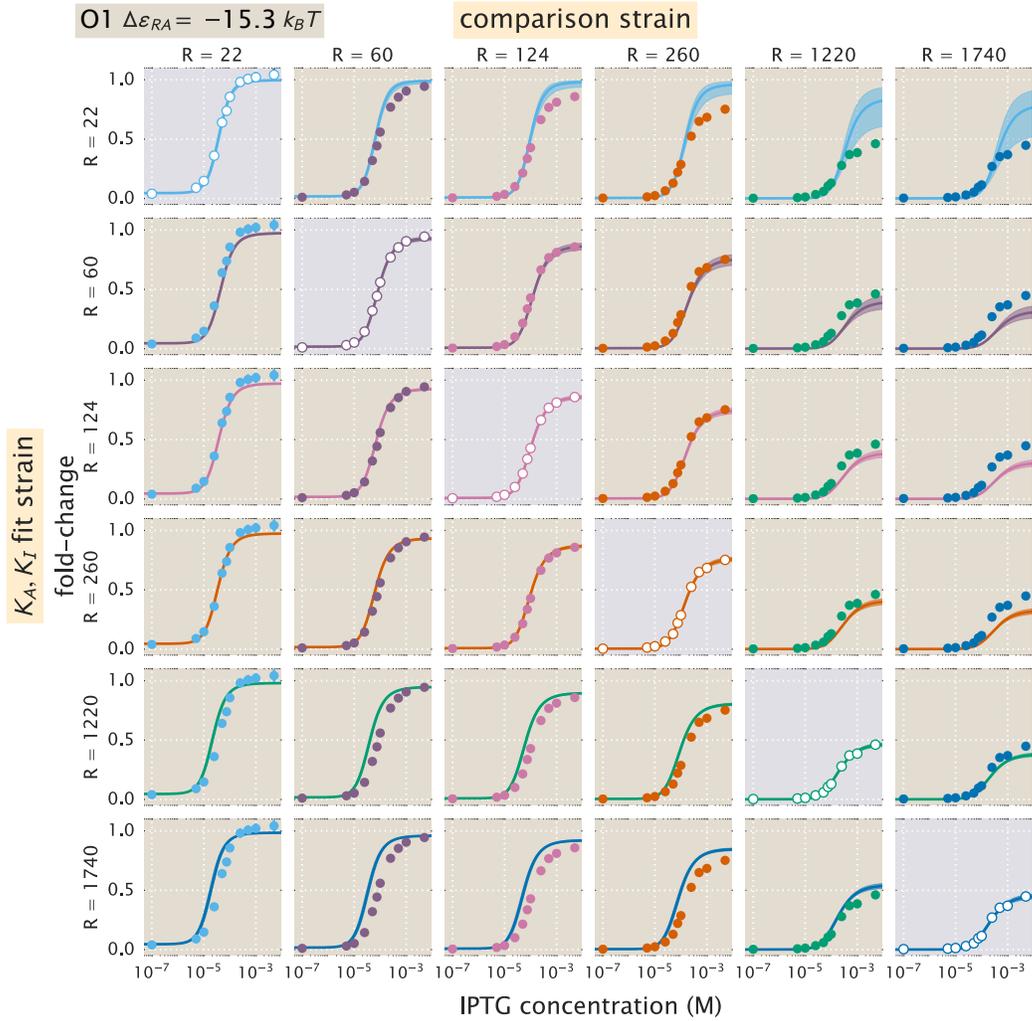


Figure S20. O1 strain fold-change predictions based on strain-specific parameter estimation of K_A and K_I . Fold-change in expression is plotted as a function of IPTG concentration for all strains containing an O1 operator. The solid points correspond to the mean experimental value. The solid lines correspond to Eq. (5) using the parameter estimates of K_A and K_I . Each row uses a single set of parameter values based on the strain noted on the left axis. The shaded plots along the diagonal are those where the parameter estimates are plotted along with the data used to infer them. Values for repressor copy number and operator binding energy are from [4]. The shaded region on the curve represents the uncertainty from our parameter estimates and reflect the 95% highest probability density region of the parameter predictions.

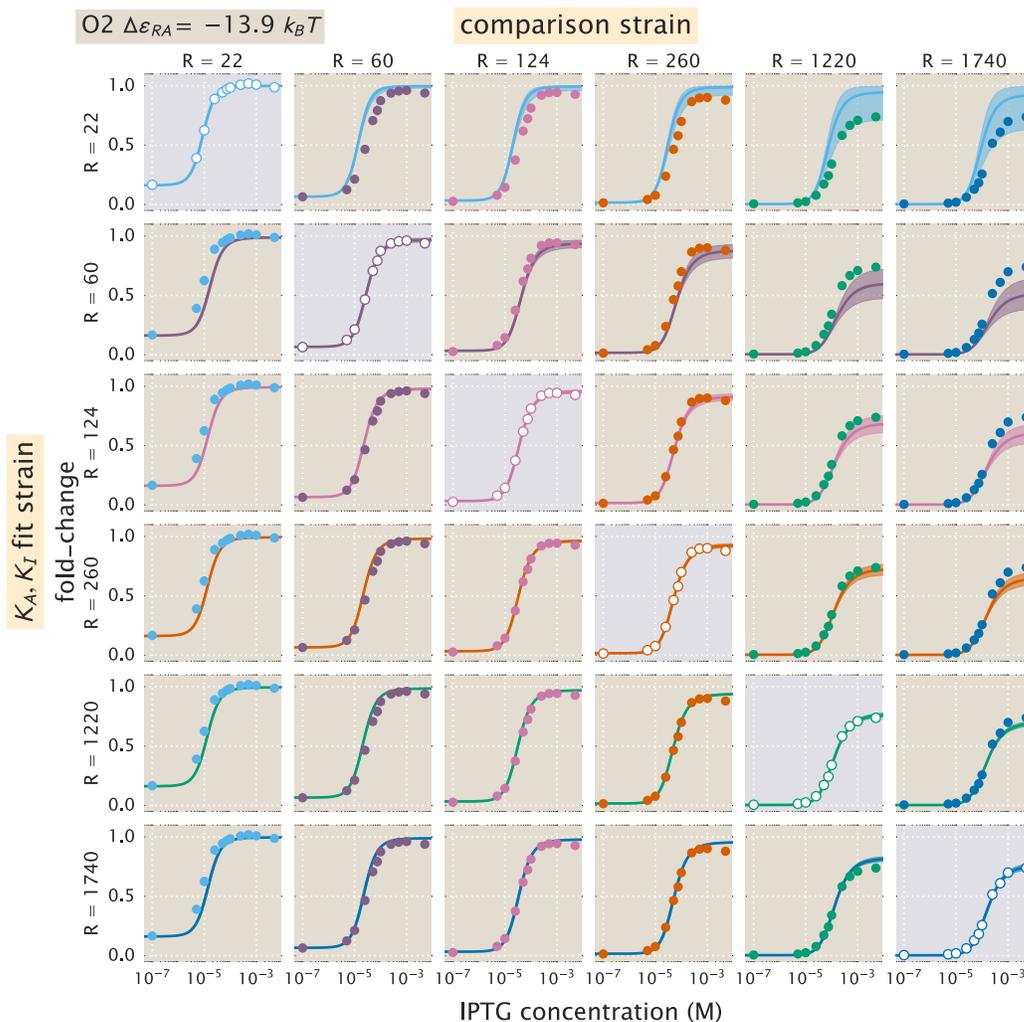


Figure S21. O2 strain fold-change predictions based on strain-specific parameter estimation of K_A and K_I . Fold-change in expression is plotted as a function of IPTG concentration for all strains containing an O2 operator. The plots and data shown are analogous to Fig. S20, but for the O2 operator.

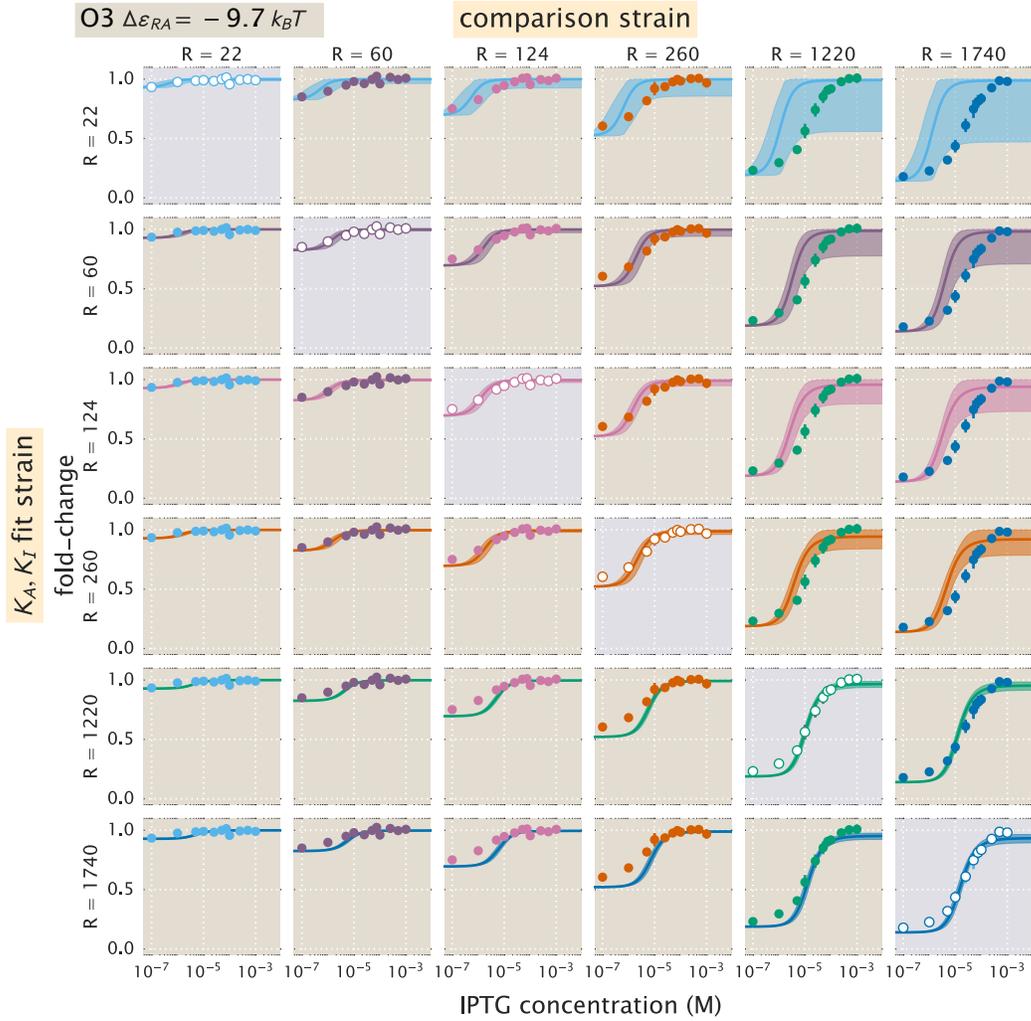


Figure S22. O3 strain fold-change predictions based on strain-specific parameter estimation of K_A and K_I . Fold-change in expression is plotted as a function of IPTG concentration for all strains containing an O3 operator. The plots and data shown are analogous to Fig. S20, but for the O3 operator. We note that when using the $R = 22$ O3 strain to predict K_A and K_I , the large uncertainty in the estimates of these parameters (see Fig. 6D) leads to correspondingly wider credible regions.

I Properties of Induction Titration Curves

In this section, we expand on the phenotypic properties of the induction response that were explored in the main text (see Fig. 1). We begin by expanding on our discussion of dynamic range and then show the analytic form of the $[EC_{50}]$ for simple repression.

As stated in the main text, the dynamic range is defined as the difference between the maximum and minimum system response, or equivalently, as the difference between the saturation and leakiness of the system. Using Eqs. (6)-(8), the dynamic range is given by

$$\text{dynamic range} = \left(1 + \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}} \left(\frac{K_A}{K_I}\right)^n} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}} \right)^{-1} - \left(1 + \frac{1}{1 + e^{-\beta\Delta\varepsilon_{AI}} \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}}} \right)^{-1}. \quad (\text{S32})$$

The dynamic range, along with saturation and leakiness were plotted with our experimental data in Fig. 7A-C as a function of repressor copy number. Fig. S23 shows how these properties are expected to vary as a function of the repressor-operator binding energy. Note that the resulting curves for all three properties have the same shape as in Fig. 7A-C, since the dependence of the fold-change upon the repressor copy number and repressor-operator binding energy are both contained in a single multiplicative term, $Re^{-\beta\Delta\varepsilon_{RA}}$. Hence, increasing R on a logarithmic scale (as in Fig. 7A-C) is equivalent to decreasing $\Delta\varepsilon_{RA}$ on a linear scale (as in Fig. S23).

An interesting aspect of the dynamic range is that it exhibits a peak as a function of either the repressor copy number (or equivalently of the repressor-operator binding energy). Differentiating the dynamic range Eq. (S32) and setting it equal to zero, we find that this peak occurs at

$$\frac{R^*}{N_{NS}} = e^{-\beta(\Delta\varepsilon_{AI} - \Delta\varepsilon_{RA})} \sqrt{e^{\Delta\varepsilon_{AI}} + 1} \sqrt{e^{\Delta\varepsilon_{AI}} + \left(\frac{K_A}{K_I}\right)^n}. \quad (\text{S33})$$

The magnitude of the peak is given by

$$\text{max dynamic range} = \frac{\left(\sqrt{e^{\Delta\varepsilon_{AI}} + 1} - \sqrt{e^{\Delta\varepsilon_{AI}} + \left(\frac{K_A}{K_I}\right)^n} \right)^2}{\left(\frac{K_A}{K_I}\right)^n - 1}, \quad (\text{S34})$$

which is independent of the repressor-operator binding energy $\Delta\varepsilon_{RA}$ or R , and will only cause a shift in the location of the peak but not its magnitude.

We now consider the two remaining properties, the $[EC_{50}]$ and effective Hill coefficient, which determine the horizontal properties of a system - that is, they determine the range of inducer concentration in which the system's response goes from its minimum to maximum values. The $[EC_{50}]$ denotes the inducer concentration required to generate fold-change halfway between its minimum and maximum value and was defined implicitly in Eq. (9). For the simple repression system, the $[EC_{50}]$ is given by

$$\frac{[EC_{50}]}{K_A} = \frac{\frac{K_A}{K_I} - 1}{\frac{K_A}{K_I} - \left(\frac{\left(1 + \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}}\right) + \left(\frac{K_A}{K_I}\right)^n \left(2e^{-\beta\Delta\varepsilon_{AI}} + \left(1 + \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}}\right)\right)}{2\left(1 + \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{RA}}\right) + e^{-\beta\Delta\varepsilon_{AI}} + \left(\frac{K_A}{K_I}\right)^n e^{-\beta\Delta\varepsilon_{AI}}} \right)^{\frac{1}{n}}} - 1. \quad (\text{S35})$$

Using this expression, we can then find the effective Hill coefficient h , which equals twice the log-log slope of the normalized fold-change evaluated at $c = [EC_{50}]$ (see Eq. (10)). In Fig. 7D-E we show how these two properties vary with repressor copy number, and in Fig. S24 we demonstrate how they depend on the repressor-operator binding energy. Both the $[EC_{50}]$ and h vary significantly with repressor copy number for sufficiently strong operator binding energies. Interestingly, for weak operator binding energies on the order of the O3 operator, it is predicted that the effective Hill coefficient should not vary with repressor copy number. In addition, the maximum possible Hill coefficient is roughly 1.75, which stresses the point that the effective Hill coefficient should not be interpreted as the number of inducer binding sites, which is exactly 2.

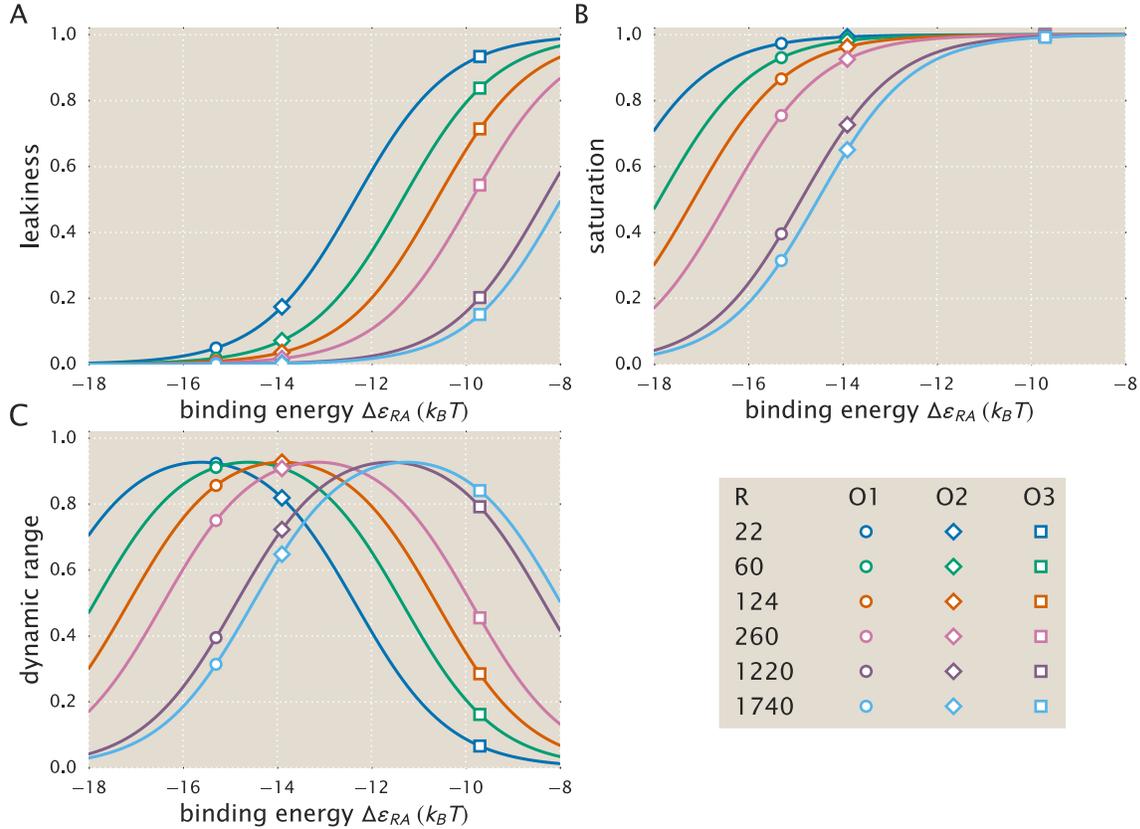


Figure S23. Dependence of leakiness, saturation, and dynamic range on the operator binding energy and repressor copy number. Increasing repressor copy number or decreasing the repressor-operator binding energy suppresses gene expression and decreases both the (A) leakiness and (B) saturation. (C) The dynamic range retains its shape but shifts right as the repressor copy number increases. The peak in the dynamic range can be understood by considering the two extremes for $\Delta\epsilon_{RA}$: for small repressor-operator binding energies, the leakiness is small but the saturation increases with $\Delta\epsilon_{RA}$; for large repressor-operator binding energies the saturation is near unity and the leakiness increases with $\Delta\epsilon_{RA}$, thereby decreasing the dynamic range. Repressor copy number does not affect the maximum dynamic range (see Eq. (S34)). Circles, diamonds, and squares represent $\Delta\epsilon_{RA}$ values for the O1, O2, and O3 operators, respectively, demonstrating the expected values of the properties using those strains.

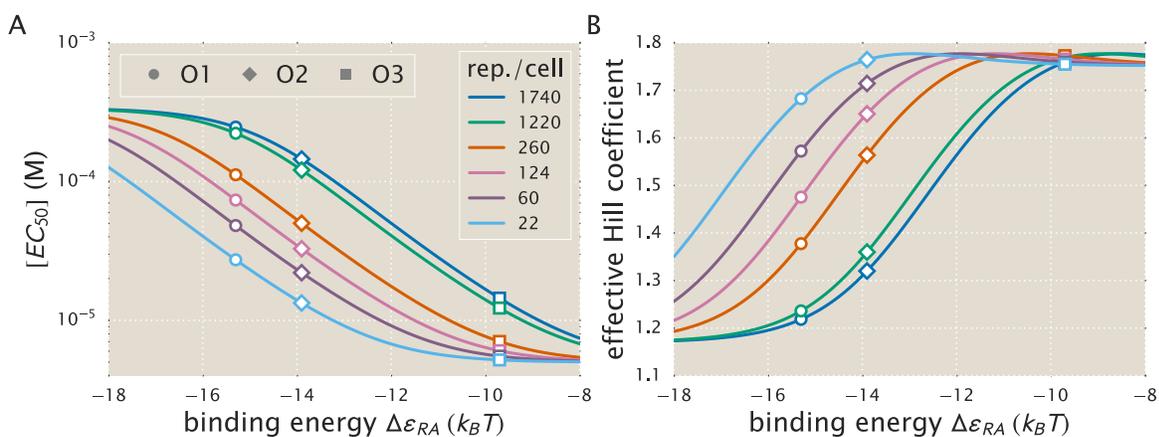


Figure S24. $[EC_{50}]$ and effective Hill coefficient depend strongly on repressor copy number and operator binding energy. (A) $[EC_{50}]$ values range from very small and tightly clustered at weak operator binding energies (e.g. O3) to relatively large and spread out for stronger operator binding energies (O1 and O2). (B) The effective Hill coefficient generally decreases with increasing repressor copy number, indicating a flatter normalized response. The maximum possible Hill coefficient is roughly 1.75 for all repressor-operator binding energies. Circles, diamonds, and squares represent $\Delta\epsilon_{RA}$ values for the O1, O2, and O3 operators, respectively.

J 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 Refs. [21, 22].

J.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),

$$\text{fold-change} \approx \left(1 + \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{\beta\Delta\epsilon_{AI}} \left(1 + \frac{c}{K_I}\right)^n} \frac{R}{N_{NS}} e^{-\beta\Delta\epsilon_{RA}} \right)^{-1}, \quad (\text{S36})$$

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. 5). Much as was performed in the main text, we can make some predictions about the how the response of a corepressor. In Fig. S25A, 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.

J.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

$$\text{fold-change} = \frac{1 + \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{\beta\Delta\epsilon_{AI}} \left(1 + \frac{c}{K_I}\right)^n} \frac{A}{N_{NS}} e^{-\beta\Delta\epsilon_{AA}} e^{-\beta\epsilon_{AP}}}{1 + \frac{\left(1 + \frac{c}{K_A}\right)^n}{\left(1 + \frac{c}{K_A}\right)^n + e^{\beta\Delta\epsilon_{AI}} \left(1 + \frac{c}{K_I}\right)^n} \frac{A}{N_{NS}} e^{-\beta\Delta\epsilon_{AA}}}, \quad (\text{S37})$$

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. S25B 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. (S37) 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.

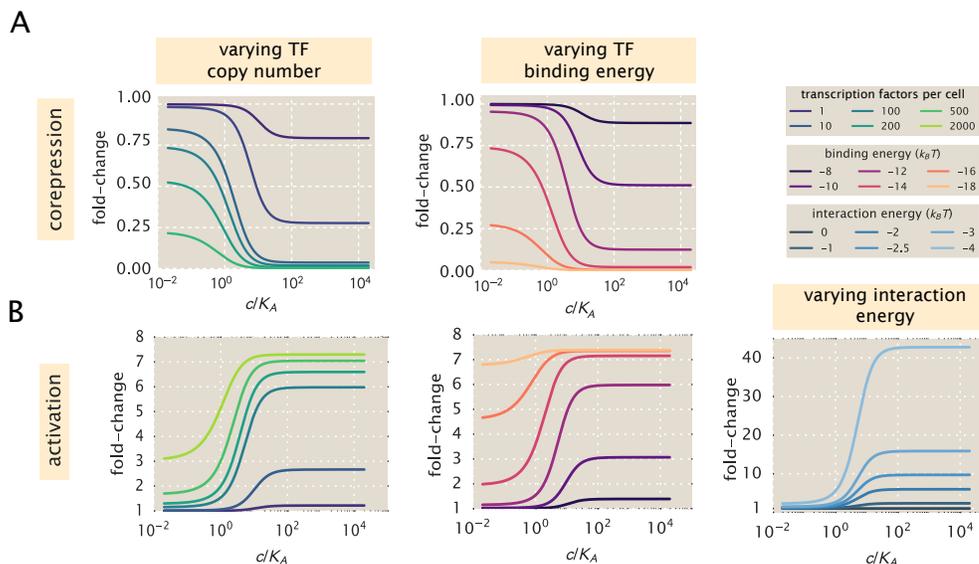


Figure S25. Representative fold-change predictions for allosteric corepression and activation. (A) Contrary to the case of induction described in the main text, addition of a corepressor decreases fold-change in gene expression. The left and right panels demonstrate how varying the values of the repressor 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.

K *E. coli* Primer and Strain List

Here we provide additional details about the genotypes of the strains used, as well as the primer sequences used to generate them. *E. coli* strains were derived from K12 MG1655. For those containing $R = 22$, we used strain HG104 which additionally has the *lacYZA* operon deleted (positions 360,483 to 365,579) but still contains the native *lacI* locus. All other strains used strain HG105, where both the *lacYZA* and *lacI* operons have both been deleted (positions 360,483 to 366,637).

All 25x+11-yfp expression constructs were integrated at the *galK* locus (between positions 1,504,078 and 1,505,112) while the 3*1x-lacI constructs were integrated at the *ybcN* locus (between positions 1,287,628 and 1,288,047). Integration was performed with λ Red recombineering [23] as described in [4] using the primers listed in Table S4. We follow the notation of Lutz and Bujard [24] for the nomenclature of the different constructs used. Specifically, the first number refers to the antibiotic resistance cassette that is present for selection (2 = kanamycin, 3 = chloramphenicol, and 4 = spectinomycin) and the second number refers to the promoter used to drive expression of either YFP or LacI (1 = $P_{LtetO-1}$, and 5 = *lacUV5*). Note that in 25x+11-yfp, x refers to the LacI operator used, which is centered at +11 (or alternatively, begins at the transcription start site). For the different LacI constructs, 3*1x-lacI, x refers to the different ribosomal binding site modifications that provide different repressor copy numbers and follows from Garcia *et al.* [4]. The asterisk refers to the presence of FLP recombinase sites flanking the chloramphenicol resistance gene that can be used to lose this resistance. However, we maintained the resistance gene in our constructs. A summary of the final genotypes of each strain is listed in Table S5. In addition each strain also contained the plasmid pZS4*1-mCherry and provided constitutive expression of the mCherry fluorescent protein. This pZS plasmid is a low copy (SC101 origin of replication) where like with 3*1x-lacI, mCherry is driven by a $P_{LtetO-1}$ promoter.

Table S4. Primers used in this work. Lower case sequences denote homology to a chromosomal locus used for integration of the construct into the *E. coli* chromosome. Uppercase sequences refer to the sequences used for PCR amplification.

Primer	Sequence	Comment
General sequencing primers:		
pZSForwSeq2	TTCCCAACCTTACCAGAGGGC	Forward primer for 3*1x-lacI
251F	CCTTTCGTCTTCACCTCGA	Forward primer for 25x+11-yfp
YFP1	ACTAGCAACACCAGAACAGCCC	Reverse primer for 3*1x-lacI and 25x+11-yfp
Integration primers:		
HG6.1 (<i>galK</i>)	gtttgcgcgagtcagcgatatccatthttcggaatccgg agtgaagaaACTAGCAACACCAGAACAGCC	Reverse primer for 25x+11-yfp with homology to <i>galK</i> locus.
HG6.3 (<i>galK</i>)	ttcatattgttcagcgacagcttgctgtacggcaggcacc agctcttccggGCTAATGCACCCAGTAAGG	Forward primer for 25x+11-yfp with homology to <i>galK</i> locus.
galK-control-upstream1	TTCATATTGTTACGCGACAGCTTG	To check integration.
galK-control-downstream1	CTCCGCCACCGTACGTAATT	To check integration.
HG11.1 (<i>ybcN</i>)	acctctgaggagggaagcgtgaacctctcacaagacggc atcaaattacACTAGCAACACCAGAACAGCC	Reverse primer for 3*1x-lacI with homology to <i>ybcN</i> locus.
HG11.3 (<i>ybcN</i>)	ctgtagatgtgtccgttcatgacacgaataagcgggtgtag ccattacgccGGCTAATGCACCCAGTAAGG	Forward primer for 3*1x-lacI with homology to <i>ybcN</i> locus.
ybcN-control-upstream1	AGCGTTTGACCTCTGCGGA	To check integration.
ybcN-control-downstream1	GCTCAGGTTTACGCTTACGACG	To check integration.

Table S5. *E. coli* strains used in this work. Each strain contains a unique operator-yfp construct for measurement of fluorescence and R refers to the dimer copy number as measured by Garcia *et al.* [4].

Strain	Genotype
O1, $R = 0$	HG105::galK \langle 25O1+11-yfp
O1, $R = 22$	HG104::galK \langle 25O1+11-yfp
O1, $R = 60$	HG105::galK \langle 25O1+11-yfp, ybcN \langle 3*1RBS1147-lacI
O1, $R = 124$	HG105::galK \langle 25O1+11-yfp, ybcN \langle 3*1RBS1027-lacI
O1, $R = 260$	HG105::galK \langle 25O1+11-yfp, ybcN \langle 3*1RBS446-lacI
O1, $R = 1220$	HG105::galK \langle 25O1+11-yfp, ybcN \langle 3*1RBS1-lacI
O1, $R = 1740$	HG105::galK \langle 25O1+11-yfp, ybcN \langle 3*1-lacI (RBS1L)
O2, $R = 0$	HG105::galK \langle 25O2+11-yfp
O2, $R = 22$	HG104::galK \langle 25O2+11-yfp
O2, $R = 60$	HG105::galK \langle 25O2+11-yfp, ybcN \langle 3*1RBS1147-lacI
O2, $R = 124$	HG105::galK \langle 25O2+11-yfp, ybcN \langle 3*1RBS1027-lacI
O2, $R = 260$	HG105::galK \langle 25O2+11-yfp, ybcN \langle 3*1RBS446-lacI
O2, $R = 1220$	HG105::galK \langle 25O2+11-yfp, ybcN \langle 3*1RBS1-lacI
O2, $R = 1740$	HG105::galK \langle 25O2+11-yfp, ybcN \langle 3*1-lacI (RBS1L)
O3, $R = 0$	HG105::galK \langle 25O3+11-yfp
O3, $R = 22$	HG104::galK \langle 25O3+11-yfp
O3, $R = 60$	HG105::galK \langle 25O3+11-yfp, ybcN \langle 3*1RBS1147-lacI
O3, $R = 124$	HG105::galK \langle 25O3+11-yfp, ybcN \langle 3*1RBS1027-lacI
O3, $R = 260$	HG105::galK \langle 25O3+11-yfp, ybcN \langle 3*1RBS446-lacI
O3, $R = 1220$	HG105::galK \langle 25O3+11-yfp, ybcN \langle 3*1RBS1-lacI
O3, $R = 1740$	HG105::galK \langle 25O3+11-yfp, ybcN \langle 3*1-lacI (RBS1L)
Oid, $R = 0$	HG105::galK \langle 25Oid+11-yfp
Oid, $R = 22$	HG104::galK \langle 25Oid+11-yfp
Oid, $R = 60$	HG105::galK \langle 25Oid+11-yfp, ybcN \langle 3*1RBS1147-lacI
Oid, $R = 124$	HG105::galK \langle 25Oid+11-yfp, ybcN \langle 3*1RBS1027-lacI
Oid, $R = 260$	HG105::galK \langle 25Oid+11-yfp, ybcN \langle 3*1RBS446-lacI
Oid, $R = 1220$	HG105::galK \langle 25Oid+11-yfp, ybcN \langle 3*1RBS1-lacI
Oid, $R = 1740$	HG105::galK \langle 25Oid+11-yfp, ybcN \langle 3*1-lacI (RBS1L)

References

1. 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.
2. Gardino AK, Volkman BF, Cho HS, Lee SY, Wemmer DE, Kern D. 2003. The NMR solution structure of BeF₃-activated Spo0F reveals the conformational switch in a phosphorelay system. *Journal of Molecular Biology* **331**: 245–254.
3. Boulton S, Melacini G. 2016. Advances in NMR methods to map allosteric sites: from models to translation. *Chemical Reviews* **116**: 6267–6304.
4. 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.
5. Weinert FM, Brewster RC, Rydenfelt M, Phillips R, Kegel WK. 2014. Scaling of gene expression with transcription-factor fugacity. *Physical Review Letters* **113**: 1–5.
6. Daber R, Sharp K, Lewis M. 2009. One is not enough. *Journal of Molecular Biology* **392**: 1133–1144.
7. Daber R, Sochor MA, Lewis M. 2011. Thermodynamic analysis of mutant Lac repressors. *Journal of Molecular Biology* **409**: 76–87.
8. 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.
9. 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.
10. Oehler S, Alberti S, Müller-Hill B. 2006. Induction of the lac promoter in the absence of DNA loops and the stoichiometry of induction. *Nucleic Acids Research* **34**: 606–612.
11. 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.
12. Sochor MA. 2014. *In vitro* transcription accurately predicts Lac repressor phenotype *in vivo* in *Escherichia coli*. *PeerJ* **2**: e498.
13. Shis DL, Hussain F, Meinhardt S, Swint-Kruse L, Bennett MR. 2014. Modular, multi-input transcriptional logic gating with orthogonal LacI/GalR family chimeras. *ACS Synthetic Biology* **3**: 645–651.
14. 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.
15. Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M. 2015. The quantitative and condition-dependent *Escherichia coli* proteome. *Nature Biotechnology* **34**: 104–110.
16. Chroma Technology Corporation. 2016. Chroma spectra viewer. URL <https://www.chroma.com/spectra-viewer?fluorochromes=10455>.
17. 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.
18. Edelstein AD, Tsuchida MA, Amodaj N, Pinkard H, Vale RD, Stuurman N. 2014. Advanced methods of microscope control using μ Manager software. *Journal of Biological Methods* **1**: 10–10.

19. Marr D, Hildreth E. 1980. Theory of edge detection. *Proceedings of the Royal Society B: Biological Sciences* **207**: 187–217.
20. Sivia D, Skilling J. 2006. Data analysis: a Bayesian tutorial. OUP Oxford.
21. Marzen S, Garcia HG, Phillips R. 2013. Statistical mechanics of Monod-Wyman-Changeux (MWC) models. *Journal of Molecular Biology* **425**: 1433–1460.
22. 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.
23. Sharan SK, Thomason LC, Kuznetsov SG, Court DL. 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols* **4**: 206–223.
24. 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.