University of Massachusetts Medical School

eScholarship@UMMS

2020-08-18

Inherent regulatory asymmetry emanating from network architecture in a prevalent autoregulatory motif

Zulfikar Ali
University of Massachusetts Medical School

Et al.

Let us know how access to this document benefits you.

Repository Citation

Inherent regulatory asymmetry emanating from network architecture in a prevalent autoregulatory motif

Md Zulfikar Ali[1,2†], **Vinuselvi Parisutham**[1,2†], **Sandeep Choubey**[3], **Robert C. Brewster**[1,2*]

*For correspondence:
Robert.Brewster@umassmed.edu (FMS)

†These authors contributed equally to this work

[1]Program in Systems Biology, University of Massachusetts Medical School, 368 Plantation St., Worcester, MA 01605; [2]Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, 368 Plantation St., Worcester, MA 01605.; [3]Max Planck Institute for the Physics of Complex Systems, Nothnitzer strasse 38, 01187 Dresden, Germany

Abstract Predicting gene expression from DNA sequence remains a major goal in the field of gene regulation. A challenge to this goal is the connectivity of the network, whose role in altering gene expression remains unclear. Here, we study a common autoregulatory network motif, the negative single-input module, to explore the regulatory properties inherited from the motif. Using stochastic simulations and a synthetic biology approach in *E. coli*, we find that the TF gene and its target genes have inherent asymmetry in regulation, even when their promoters are identical; the TF gene being more repressed than its targets. The magnitude of asymmetry depends on network features such as network size and TF binding affinities. Intriguingly, asymmetry disappears when the growth rate is too fast or too slow and is most significant for typical growth conditions. These results highlight the importance of accounting for network architecture in quantitative models of gene expression.

Introduction

The genomics revolution has enabled biology with the ability to read, write and assemble DNA at the genome scale with single base pair resolution. These advancements have provided an important tool for the field of gene regulation that aims to predict gene expression from the regulatory code, inscribed in DNA (*Carey et al.* (*2013*); *Kosuri et al.* (*2013*); *Sharon et al.* (*2012*)) This approach relies on quantitative measurements of gene expression as the regulatory DNA is systematically designed to induce regulation by various transcription factors (TFs) at specific positions or with differing affinities. However, success in predicting expression levels of natural genes from sequence alone has been relatively modest. One obvious complication is that genes are not isolated but rather exist in dense, interconnected networks. The concept of network motifs, defined as overrepresented patterns of connections between genes and TFs in the network, helps to digest these large networks into smaller subgraphs with specific properties; each of these motifs can be interpreted as performing a particular "information processing" function that is determined by the connectivity and regulatory role of the genes in the motif (*Alon* (*2006*, 2007); *Davidson* (*2006*); *Mangan and Alon* (*2003*); *Tkačik et al.* (*2008*). In this study, we dissect a prevalent gene regulation

motif, the single-input module (SIM), to demonstrate the influence of network size and connectivity on the regulation of a network motif.

The SIM is a network motif where a single TF regulates the expression of a set of genes, including itself (Fig. 1A). In *E. coli* this motif is prevalent; the majority of TFs are autoregulated and have multiple targets (*Santos-Zavaleta et al.* (*2018*)). Typically, this group of genes have related functions and the purpose of this motif is to coordinate, in both time and magnitude, expression of these related genes (*Alon* (*2006*)). There are mounting examples, from diverse topics that range from metabolism (Fig. 1B, (*Zaslaver et al.* (*2004*))), stress response (Fig. 1C, (*Friedman et al.* (*2005*); *Ronen et al.* (*2002*))), development (*Arnone* (*2002*); *Gaudet and Mango* (*2002*); *Kalir et al.* (*2001*)), and cancer (*Lorenzin et al.* (*2016*)), where temporal ordering of gene expression in the motif naturally follows the functional order of the genes in the physiological pathway. Mechanistically, it is thought that this ordering is set through differential affinity for the TF amongst the various target genes in the motif (*Alon* (*2006*)), although in some experiments temporal ordering was not observed implying a dependence on physiology or another experimental detail that is yet unrecognized (*Gerosa et al.* (*2013*); *Schmidt et al.* (*2016*)). Due to the broad importance of these motifs, a quantitative understanding of how SIM motifs can be encoded, designed and optimized, will be instrumental in gaining a deep and fundamental understanding of the spatial and temporal features of a diverse set of cellular phenomena.

To quantitatively explore the input-output relationship of the SIM motif, we use a synthetic biology approach that boils the motif down to its most basic components: an autoregulated TF gene, a sample target gene, and competing binding sites. Using *E. coli* as a model organism we build this motif *in vivo*. We use non-functional "decoy" binding sites to exert competition for the TF and mimic the demand of the other genes in the motif (which will depend on the size of the network, Fig. 1D (*Gillespie* (*1977*); *Shen-Orr et al.* (*2002*)). However, the demand for the TF could also stem from a litany of sources such as random non-functional sites in the genome (*Bakk and Metzler* (*2004*); *Kemme et al.* (*2016*); *Lee and Maheshri* (*2012*); *Mirny* (*2009*)) or non-DNA based obstruction or localization effects that transiently interfere with a TFs ability to bind DNA. Because of the design, our results do not depend on the nature of the TF competition. SIM TFs typically exert the same regulatory role on all targets of the motif (*Shen-Orr et al.* (*2002*)). As such, in this work we will focus on a TF that is a negative regulator of its target genes and itself; this is the most common regulation strategy in *Escherichia coli* where roughly 60% of TF genes are autoregulated and almost 70% of those TFs negatively regulate their own expression (inset Fig. 1D, (*Shen-Orr et al.* (*2002*)).

We use stochastic simulations of kinetic models (*Gillespie* (*1977*, 2007); *Kaern et al.* (*2005*); *Shahrezaei and Swain* (*2008*)), to predict how the overall level of gene expression depends on parameters characterizing cellular environment such as TF binding affinities and the number of competing binding sites. To test these predictions *in vivo*, we built a synthetic system with LacI as a model TF, and individually tune each of these parameters. Past work with LacI has demonstrated the ability to control with precision the regulatory function, binding affinity and TF copy number through basic sequence level manipulations (*Brewster et al.* (*2014*); *Choi et al.* (*2008*); *Garcia and Phillips* (*2011*); *Jones et al.* (*2014*); *Kuhlman et al.* (*2007*); *Oehler et al.* (*1990*); *Razo-Mejia et al.* (*2018*)); Here we use that detailed knowledge to inform our simulations which then guide our experiments (and vice versa).

Our approach reveals that the presence of competing TF binding sites can have counterintuitive effects on the mean expression levels of the TF and its target genes due to the opposing relationship between free TFs and total TFs (total TF is the sum of free TF and TF bound to promoters and decoy binding sites). Furthermore, we find that the TF and target gene experience quantitatively different levels of regulation in the same cell, and with the same regulatory sequence. We show that this regulatory asymmetry is sensitive to features such as the degradation rate, TF binding affinity and the number of competing binding sites for the TF. The stochastic simulation makes accurate predictions of the asymmetry and its dependence on the parameters of the model that we confirm through *in vivo* measurements. Interestingly, regulatory asymmetry is not captured by a

91 simple deterministic model which is based on translating the stochastic reactions to kinetic rates
92 through mass action equilibrium kinetics (which have been shown to accurately predict target gene
93 expression in other studies (*Brewster et al.* (*2014*); *Garcia and Phillips* (*2011*); *Garcia et al.* (*2012*);
94 *Jones et al.* (*2014*); *Razo-Mejia et al.* (*2018*)). In fact, this deterministic model fails to accurately
95 predict expression of either gene. A revised deterministic model, which explicitly allows for different
96 microenvironments in each "regulatory state", predicts asymmetry although it still does not recover
97 quantitative agreement with stochastic simulations.

Results

Matching molecular biology with simulation methodology

100 We use a combination of theory and experimental *in vivo* measurements on engineered *E. coli*
101 strains to study the interplay between TF gene, target gene, and additional binding sites of a
102 negative autoregulatory SIM network motif. The basic regulatory system is outlined in Fig. 1E. We
103 use a stochastic model of the SIM motif to explore how the expression of the TF gene and one
104 target gene depends on parameters such as TF binding affinity and number of other binding sites in
105 the network (here modeled and controlled through competing, non-regulatory decoy sites (*Burger*
106 *et al.* (*2010*)). In this model, the TF gene and target gene can be independently bound by a free TF
107 to shut off gene expression until the TF unbinds. The two genes (TF-encoding and target) compete
108 with decoy binding sites which can also bind free TFs. Each free TF can bind any open operator site
109 with equal probability (set by the binding rate). The unbinding rate can be set individually for the
110 TF gene, target gene and decoy sites and is related to the specific base pair identity of the bound
111 operator site (*Kinney et al.* (*2010*); *Maerkl and Quake* (*2007*); *Stormo* (*2000*); *Weirauch et al.* (*2013*)).
112 We employ stochastic simulations to make specific predictions for how the expression level of the TF
113 and target genes depend on the various parameters of the model. Furthermore, we translate these
114 stochastic processes into a deterministic ODE model using equilibrium mass action kinetics (see
115 Appendix 6: Deterministic solution). A thorough discussion on how we chose the kinetic parameters
116 of our model is presented in the methods section.

117 In experiments, the corresponding system is constructed with an integrated copy of both the
118 TF (LacI-mCherry) and target gene (YFP) with expression of both genes controlled by identical
119 promoters with a single LacI binding site centered at +11 relative to their transcription start sites
120 (*Brewster et al.* (*2014*); *Garcia and Phillips* (*2011*)). As demonstrated in Fig. 1F, decoy binding sites
121 are added by introducing a plasmid with an array of TF binding sites (between 0 to 5 sites per
122 plasmid) enabling control of up to roughly 300 binding sites per cell (for average plasmid copy
123 number measured by qPCR, see methods and Appendix 3 Figure 1). TF unbinding rate is controlled
124 by changing the sequence identity of the operator sites; the binding sequence assessed in this
125 study include (in order of increasing affinity) O2, O1 and Oid. The decoy binding site arrays are
126 constructed using the Oid operator site. We quantify regulation through measurements of fold-
127 change (FC) in expression which is defined as the expression level of a gene in a given condition
128 (typically a specific number of decoy binding sites) divided by the expression of that gene when
129 it is unregulated. For the target gene we can always measure unregulated expression simply by
130 measuring expression in a LacI knockout strain. However, it is challenging to measure unregulated
131 expression for the autoregulated gene. For autoregulation this unregulated expression can be
132 measured by exchanging the TF binding site with a mutated non-binding version of the site. For O1
133 there is a mutated sequence (NoO1v1 *Oehler et al.* (*1994*)) that we have shown relieves repression
134 of the target gene comparable to a strain expressing no TF (see Appendix 4 Figure 1A) which
135 allows us to calculate fold-change even for the autorepressed gene. Despite testing many different
136 mutated sites and strategies, we could not find a corresponding sequence for O2 and Oid so we
137 focus primarily on studying a TF gene regulated by O1 (see Appendix 4: Constitutive values for
138 autoregulatory gene, for more discussion).

Decoy sites increase expression of the auto-repressed gene and its targets

We first investigate the negatively regulated SIM motif where the TF and target gene have identical promoters and TF binding sites (O1) and the number of (identical) competing binding sites are varied systematically (schematically shown in Fig. 1E, F). Simulation and experimental data for Fold-change of the TF gene as a function of number of decoys is shown in Fig. 2A as red lines (simulation) and red points (experiments). We find that increasing the number of decoy sites increases the expression of the auto-repressed TF gene monotonically. To interpret why the TF level increases, in Fig. 2B we plot the number of "free" TFs in our simulation (defined as TFs not bound to an operator site) as a function of decoy site number. The solid line demonstrates that on average, despite the increased average number of TFs in the cell, the number of unbound TFs decreases as the number of competing binding sites increases (*Nevozhay et al.* (*2009*)). Therefore, because the number of available repressors decreases, the overall level of repression also decreases and thus the mean expression of the TF gene rises.

Now we consider the effect of competition on the expression of SIM target genes. We measure our system with O1 as the regulatory binding site for both TF and target genes. In Fig. 2A, the expression of the target gene is shown as blue points (experiments) and blue lines (simulation) for the SIM motif with different numbers of decoy TF binding sites (from 0 sites up to 5 per plasmid). Just as in the case of the TF gene, we once again see that the expression of the target gene increases as more decoy binding sites are added even though the total number of TFs is also increasing (red points and line). Qualitatively, we expected this result since the free TF number is expected to decrease (Fig. 2B) and, in turn, the expression of any gene targeted by the autoregulated repressing TF will increase. While the mechanism is more obvious in this controlled system, it is important to note that this is a case where more repressors correlate with more expression of the repressed gene. It is easy to see how this relationship could be misinterpreted as activation in more complex *in vivo* system if the competition level of the TF is (advertently or otherwise) altered in experiments.

Asymmetry in gene regulation between TF and target genes

Quantitative inspection of Fig. 2A reveals an interesting detail: Even when the regulatory region of the auto-repressed gene and the target gene are identical, we find that the expression (fold-change or FC) is higher for the target gene, raising the question of how two genes with identical promoters and regulatory binding sites in the same cell can have different regulation levels. In this data, both the TF gene and target gene are regulated by a single repressor binding site (O1) immediately downstream of the promoter. This regulatory scheme is often referred to as "simple repression" (*Bintu et al.* (*2005*); *Garcia and Phillips* (*2011*); *Phillips et al.* (*2013*)). Drawing our intuition from a simple deterministic model of regulation based on translating the stochastic reactions to kinetic rate equations (Fig. 2C and Appendix 6: Deterministic solution), we find that regardless of the network architecture (autoregulation, constitutive TF production, number of competing sites, *etc.*), the fold-change of any gene is expected to follow a simple scaling relation,

$$\text{Fold} - \text{change} = \frac{1}{1 + R^*},$$

$$R^* = R_{\text{free}} \frac{k_{\text{on}}}{k_{\text{off}} + \gamma}.$$

where, R_{free} is the number of free (unbound) TFs and $k_{\text{on}}/(k_{\text{off}} + \gamma)$ represents the affinity of the specific TF binding site in the thermodynamic framework (*Rydenfelt et al.* (*2014*)). This calculation is applicable for both the TF and the target gene and would predict a "symmetric" response for identical regulatory regions. This model performs well for this same promoter in a related system where the TF is induced or constitutively expressed and predicts the fold-change for a wide range of perturbations such as promoter strength, TF binding site, induction condition and TF competition levels are tuned (data accumulated in Fig. 3A, adapted from (*Phillips et al.* (*2019*)). However, it has been shown that the regulation of an autorepressed gene can diverge from this prediction (*Hahl*

184 *and Kremling* (*2016*); *Hornos et al.* (*2005*); *Milias-Argeitis et al.* (*2015*)). In Fig. 3, we show simulation
185 data for the fold-change versus number of scaled-free TFs (R^*) for the autoregulatory gene (red line)
186 and its target gene (blue line) with O1 (Fig. 3C) or O2 (Fig. 3B) binding sites, where we are changing
187 the number of free TFs by tuning the number of competing binding sites. In each plot, we also show
188 simulations for the fold-change of a single target gene with a TF undergoing constitutive (constant
189 in time) expression where the TF is controlled by either changing the expression level of the TF
190 (purple stars) or adding competing binding sites while maintaining a set constitutive expression
191 level (purple circles). In both cases, where TFs are made constitutively, the simulation data agrees
192 well with the deterministic model predictions. However, for the autoregulatory circuits, we find that
193 for strong binding sites (O1) neither the target nor the TF gene follow the deterministic solution
194 (black dashed line). In this case, the asymmetry occurs with the TF gene being more repressed and
195 the target gene less repressed than expected.

196   Since "free TF concentration" is not readily available in experiments, we demonstrate asymmetry
197 in experimental results explicitly in Fig. 3D, where we plot the fold-change of the target gene against
198 fold-change of the TF gene. In this figure, the data points are derived from measurements made
199 in six different competition levels (from 0 to 5 decoy binding sites per plasmid). Each data point
200 represents the average expression level of each gene for a given number of competing binding
201 sites. The lines represent results from the stochastic simulations where we systematically vary
202 competition levels by introducing decoy binding sites and the fold-change of both the TF and target
203 gene are calculated. The simple deterministic model prediction that identical promoters (yellow
204 data, Fig. 3D) should experience identical levels of regulation (see Appendix 6 Figure 1C, (*Sanchez*
205 *et al.* (*2011*)) would cause the data to fall on the black dashed one-to-one line. However, for both
206 simulations and experiments of this system the TF gene is clearly more strongly regulated than the
207 target gene subject to identical regulatory sequences.

208   To examine the extent of asymmetry in this system, we adjust the target binding site to be of
209 higher affinity (Oid, blue lines and data points in Fig. 3D) or weaker (O2, purple lines and data points
210 in Fig. 3D). Clearly, this should change the symmetry of the regulation, after all the TF binding sites
211 on the promoters are now different and symmetry is no longer to be expected. The experiments
212 and simulations once again agree well. However, when Oid regulates the target gene and O1
213 regulates the TF gene, the regulation is now roughly symmetric despite the target gene having a
214 much stronger binding site; in this case, the size of the inherent regulatory asymmetry effect is on
215 par with altering the binding site to a stronger operator resulting in symmetric overall regulation of
216 the genes.

Mechanism of asymmetric gene regulation

218 The difference in expression between the TF and its target can be understood by studying the
219 TF-operator occupancy for each gene, drawn schematically in Fig. 4A. This cartoon shows the four
220 possible promoter occupancy states of the system: (1) both genes unbound by TF, (2) target gene
221 bound by TF, TF gene unbound, (3) TF gene bound by TF, target gene unbound, and (4) both genes
222 bound by TF. It should be clear that state 1 and state 4 cannot be the cause of asymmetry; both
223 genes are either fully on (state 1) or fully off (state 4). As such the asymmetry must originate from
224 differences in states 2 and 3. In state 2, the TF gene is "on" while the target gene is fully repressed
225 and in state 3 the opposite is true. Since we know that the asymmetry appears as more regulation
226 of the TF gene than the target gene, then it must be the case that the system spends less time in
227 state 2 than in state 3. There are two paths to exit either of these states: unbinding of the TF from
228 the bound operator or binding of the TF to the free operator. Since unbinding rate of a TF is identical
229 for both promoters in our model, the asymmetry must originate from differences in binding of free
230 TF in state 2 and in state 3; specifically state 2 must have an (on average) higher concentration of
231 TF than state 3. This makes sense since the system is still making TF in state 2, while production
232 of TF is shut off in state 3. Fig. 4B validates this interpretation as we can see that state 2 has on
233 average more free TFs than state 3, and as a result, the system spends less time in state 2 than

234 in state 3 in our simulations. As such, the asymmetry comes from the fact that the two genes,
235 despite being in the same cell and experiencing the same average intracellular TF concentrations,
236 are exposed to systematically different concentrations of TF when the TF and target gene are in their
237 respective "active" states. To quantify regulatory asymmetry, we define asymmetry as the difference
238 in fold-change of the target and the fold-change of the TF gene (asymmetry $= \text{FC}_{\text{target}} - \text{FC}_{\text{TF}}$). Using
239 the chemical master equation (CME) approach, we find that the asymmetry is exactly equal to the
240 difference in time spent in state 3 and state 2, for any condition or parameter choice (Fig. 4C and
241 Appendix 9 Eqn. A9-9: CME for minimal model). Furthermore, the asymmetry can be written as the
242 difference of TF concentration in state 2 and state 3 and is given by

$$\text{Asymmetry} = \frac{k_{\text{on}}}{k_{\text{off}} + \gamma}(n_2 - n_3),$$

243 where, n_2 and n_3 are the TF concentrations in state 2 and state 3, respectively. In Fig. 4D we show
244 that the asymmetry obtained using the difference in TF concentration precisely match with the
245 asymmetry calculated from the fold-change expression. However, it is important to note, that this
246 is not a complete analytic solution for asymmetry because n_2 and n_3 are unsolved functions of the
247 model parameters.
248   The asymmetry in the expression of TF and target genes stems from systematically differential
249 TF concentration in the states when the TF gene is occupied (and target gene is expressing) and
250 when the target gene is occupied (and the TF gene is expressing). The general approach of ODEs
251 outlined above (Fig. 2C) does not account for this differential TF concentration and hence shows no
252 asymmetry. Armed with the knowledge that individual states have this systematic TF difference,
253 we can rewrite the basic deterministic model where we instead keep track individually of each
254 state and the specific TF concentration of that state using the same equilibrium mass action kinetic
255 approach (details in Appendix 10: Modified ODEs for the minimal model). Like the stochastic CMEs
256 the modified ODEs predict that the asymmetry arises from the difference in the TF concentrations
257 in different states and solely depends on the difference in time spent in state 3 (only target gene
258 occupied) and state 2 (only TF gene occupied). Although we find the modified deterministic model
259 can predict asymmetry, it still does not quantitatively agree with the results of stochastic modeling
260 due to the deterministic model not accounting for variability in TF number in each state (see
261 Appendix 10: Modified ODEs for the minimal model). As a result, in the following sections, we will
262 compare our experiments to stochastic simulations based on the full CME formalism.

Dependence of regulatory asymmetry on TF degradation and binding affinity

264 According to the above proposed mechanism, the regulatory asymmetry stems from differences
265 in the cellular TF concentration when the TF is bound to the target versus when it is bound to
266 the autoregulatory gene, as such we expect that binding affinity will play a central role in setting
267 asymmetry levels. This is also evident from Fig. 3B, C where we find that the deviation of the
268 expression of both TF and target gene is more prominent for a strong binding site (Oid or O1)
269 compared to a weaker binding site (O2). Furthermore, there are many parameters associated
270 with the production and decay of TF and target mRNA and protein which could also influence the
271 asymmetry. To reveal which (if any) of these parameters is important to asymmetry, we calculate
272 the maximum asymmetry (the maximum value of asymmetry found as competing site number is
273 controlled, Appendix 7 Figure 1A) using simulation as these production and degradation parameters
274 are tuned. First, we find that tuning the rates of target gene production and decay has no effect on
275 asymmetry (Appendix 7 Figure 1B and Appendix 11 Figure 1B). On the other hand, for TF production
276 and decay each parameter has some effect on asymmetry. However, we find that the biggest driver
277 of asymmetry in this set of parameters is the protein degradation rate (Appendix 7 Figure 1B).
278 As such, we focus on two crucial parameters that control the asymmetry: TF binding affinity and
279 TF degradation rate. In Fig. 4E we show a heat map of the maximum asymmetry as a function
280 of the rate of protein degradation and binding affinity of the TF. We see from this figure that

strong binding produces enhanced asymmetry, but the degradation rate displays an interesting intermediate maximum in asymmetry – degradation that is too fast, or too slow will not show asymmetry, but a maximum asymmetry is expected for TF lifetimes between 10 and 100 minutes. Crucially, this maximum coincides with typical doubling time of *E. coli* (which sets the TF half-life (*Marr* (*1991*); *Neidhardt and Curtiss* (*1996*)) and thus regulatory asymmetry in this motif is most relevant in common physiological conditions.

The non-monotonic behavior of asymmetry with degradation rate of TF can be explained by the TF-promoter occupancy (alternatively, residence time) of the TF and the target gene. Analytically, the asymmetry is given by the difference of occupancy of state 2 and state 3 (Appendix 9 Eqn. A9-7: CME for minimal model). For slow degradation, the number of TFs in a cell is high, favoring the transition to state 4 very quickly, thereby reducing the residence times of both state 2 and 3. On the other extreme, when degradation is fast, the TF number is too low for the cell to be in the state 2 or 3; the cell spends most of the time in state 1. In both the cases, the difference of residence times between state 2 and state 3 is low and hence the asymmetry is small. In the intermediate regime of degradation, the number of TFs is optimum to maximize the difference between residence times in state 2 and 3, which leads to maximum asymmetry.

To experimentally test the theory predictions for the role of TF degradation in setting regulatory asymmetry, we introduced several ssrA degradation tags to the LacI in our experiments (*McGinness et al.* (*2006*)). The data, shown in Fig. 4F includes degradation by a "weak" or "slow" tag (DAS with a rate of 0.00063 per minute per enzyme (*McGinness et al.* (*2007*)), blue points), a slightly faster tag (DAS+4 with a rate of 0.0011 per minute per enzyme (*McGinness et al.* (*2007*)), green points) and a very fast tag (LAA tag with a rate of 0.21 per minute per enzyme (*McGinness et al.* (*2007*)), red points) . In addition, the data without a tag is shown as yellow points. Here we see that the slowest tag (blue points) introduces strong asymmetry. However, for the next fastest tag (green points) we see a significant decrease in asymmetry and the level of regulatory asymmetry is similar to what is seen in the absence of tags (yellow points). Finally, the fastest tag (red points) shows no asymmetry at all. It is worth pointing out that the qualitative order of degradation rates in these experiments can be inferred from how far the data "reaches", faster degradation will lead to higher overall fold-changes for a given competition level. Importantly, controlling the protein degradation rate through this synthetic tool agrees with our model predictions, although the actual *in vivo* protein degradation rates are difficult to estimate from tag sequence alone, the asymmetry follows the expected trends based on the known (and observed) effectiveness of each tag (see schematic inset Fig. 4F).

In the absence of targeted degradation, the degradation rate of most protein in *E. coli*, is naturally set by the growth rate. According to the model predictions in Fig. 4E, the asymmetry should be highest for fast growing cells (roughly 20-minute division rate for our growth conditions which is well below the degradation rate for peak asymmetry ∼ 10 minutes, Fig. 4F) and decrease (or vanish) for very slow growing cells. To test this, we take the system with O1 regulatory binding sites on both the target and the TF promoter (yellow data in Fig. 3D grown in M9 + glucose, 55-minute doubling time) and grow in a range of doubling times between 22 minutes (rich defined media) up to 215 minutes (M9 + acetate) (see Appendix 2 Figure 1A). Importantly, when we change the growth rate, other rates such as the transcription and translation rates will also be impacted (*Bremer and Dennis* (*2008*); *Klumpp et al.* (*2009*)), while these parameters will change the quantitative values of the asymmetry curve, the qualitative ordering and features of the asymmetry are not expected to be impacted (see Appendix 11 Figure 1C). The data for these growth conditions is shown in Fig. 5A. As predicted, faster growing cells show more regulatory asymmetry and slower growing cells show little-to-no regulatory asymmetry. We also test the role of growth rate in asymmetric regulation when O2 (a lower affinity site) and Oid (a higher affinity site) are used as the regulatory binding sites instead of O1. This data is shown in Fig. 5B (O2) and 5C (Oid). As discussed above, we could not find a suitable mutant for O2 and Oid that both relieved regulation from LacI and completely restored the expression of target gene (see Appendix 4: Constitutive values for autoregulatory gene.). This

means we cannot explicitly measure the 1-1 correlation between the two axes in our data when using O2 or Oid for the TF gene. To this end, we find this correspondence by fitting the glucose data to our simulation of the same system and use that value to normalize all other growth rates for that operator. Despite this complication, it is clear that O2 regulation is symmetric at all studied growth rates while Oid regulation is asymmetric for all growth rates with faster growth rates appearing more asymmetric.

Importantly, the regulatory asymmetry is not due to a small population of outliers, bimodality or any other "rare" phenotype. In Fig. 5D, we show a histogram of single cell asymmetry values (defined as asymmetry = $FC_{Target} - FC_{TF}$) for each condition. As can be seen, expression in each media condition are roughly symmetric for most cells at the lowest competition levels (top panel). However, as competition levels are increased, the fast-growing conditions shift to higher asymmetry levels; strikingly at the highest growth rate almost every single cell is expressing target at a higher level than TF (bottom panel).

Discussion

The single-input module (SIM) is a prevalent regulation strategy in both bacteria (*Ma et al.* (*2004*); *Shen-Orr et al.* (*2002*)) and higher organisms (*Lee et al.* (*2002*); *Segal et al.* (*2003*); *Yu et al.* (*2003*)). While the role of TF autoregulation (positive and negative) has been extensively studied (*Acar et al.* (*2008*); *Assaf et al.* (*2011*); *Becskei and Serrano* (*2000*); *Ochab-Marcinek et al.* (*2017*); *Rodrigo et al.* (*2016*); *Rosenfeld et al.* (*2002*); *Savageau* (*1975*); *Semsey et al.* (*2009*)), the focus here is on the combined influence of an autoregulated TF and its target genes and how the shared need for that TF influences the quantitative features of its regulatory behaviors. We find that there is a fundamental asymmetry in gene regulation that can occur in the SIM regulatory motif. This asymmetry is not related to distinctions in the biological processes or an unexpected difference in our *in vivo* experiment, but rather an inherent asymmetry originating from the way the motif itself is wired. Although two identical promoters are in the same cell with the same average protein concentrations, they experience distinct regulatory environments. This is particularly relevant for the SIM motif because the primary function of the motif, organizing and coordinating gene expression patterns, operates on the premise of differential affinities amongst target genes; here we have shown that the TF gene has an inherent "affinity advantage" due to being exposed to systematically higher TF concentrations than its target genes. This implies that the TF gene will respond "earlier" than expected based on the raw affinity of its binding site and may necessitate weaker sites on autoregulating TF genes in order to achieve similar timing in expression compared to its targets. This may also shed light on the discrepancies in Arg pathway timing between different experiments which have used plasmid reporters (essentially changing network size) or different physiological growth conditions; the asymmetry is critically sensitive to both of these features. Although, here we are using *E. coli* as a model organism where it is easy to build and manipulate these regulatory motifs, we expect this phenomenon to apply broadly to other regulatory systems.

Regulatory asymmetry is intrinsic to the negative SIM motif even in the absence of decoys, but it can be greatly exacerbated by competing TF binding sites. Due to the promiscuous nature of TF binding, this highlights the importance of considering not just the "closed" system of a TF and a given target but also the impact of other binding sites (or inactivating interactions) for the TF in predicting regulation as well as the regulatory motif at play in the system. In our system, the magnitude of the asymmetry is enough to compensate for swapping the wild-type proximal O1 LacI binding site on the target gene with the "ideal" operator Oid.

The cause of this asymmetry is a systematic difference in the TF concentration when the TF gene is active compared to when the target gene is active. As such, asymmetry is magnified by anything that enhances this concentration difference. Here we have identified TF binding affinity and TF degradation rate (controlled both directly and through modulating growth rate) as primary drivers of asymmetry in this motif. Although the relationship between growth rate and expression levels is well established (*Bremer and Dennis* (*2008*); *Klumpp and Hwa* (*2008*); *Klumpp et al.* (*2009*); *Scott*

et al. (*2010*); *Volkmer and Heinemann* (*2011*)), effects such as this add a layer of complexity to this relationship.

In studies of quantitative gene regulation, the typical goal is to predict the output of a gene based on the regulatory composition of that gene's promoter and the number and identity of regulatory proteins. This work clearly presents a challenge for the drive to "read" and predict regulation levels from the promoter DNA alone, in this case the regulatory motif is responsible for altering the observed regulation and must be considered as well. It has previously been demonstrated that features of a transcript can impact its regulation by effects such as targeted degradation, stabilization or posttranslational modification and regulation (*Schikora-Tamarit et al.* (*2018*)), it is important to point out that regulatory asymmetry in this motif is a distinct phenomenon that does not operate through an enzymatic processes but rather is a fundamental feature of the network.

Finally, here we demonstrate regulatory asymmetry using a specific (but common) regulatory motif. The more general problem of quantifying the role of asymmetry in other network motifs may be an important step in expanding the predictive power of models based on single genes. The broader point that specific genes can be exposed to systematically different levels of regulatory TFs even in the absence of specific cellular mechanisms such as cytoplasmic compartmentalization, protein localization or DNA accessibility is likely more generally relevant. Understanding and quantifying these mechanisms can be an important piece towards improving our ability to predict and design gene regulatory circuits.

Materials and Methods

Key Resources Table				
Reagent type (species) resources	Designation	Source or reference	Identifiers	Additional information
gene (*E. coli*)	*ybcN<>25XX+ 11-lacI-mcherry*	GeneBank	MT726947	TF gene; XX can be O1, O2 or Oid operator
gene (*E. coli*)	*galK<>3*5XX+ 11-yfp*	GeneBank	MT726948	Target gene; XX can be O1, O2 or Oid operator
strain, strain background (*E. coli*)	*E. coli* MG1655	Lab stock	CGSC#6300	Wild type
strain, strain background (*E. coli*)	HG105	*Garcia and Phillips* (*2011*)		*E. coli* MG1655 with *lac* operon deleted
strain, strain background (*E. coli*)	HG105 $\Delta sspB$	This study		*E. coli* HG105 with $sspB$ gene deleted
Other	M9 minimal media	BD Diagnostics	DF0485-17	commercial media
Other	Rich defined media	Teknova	#M2105	commercial media
software, algorithm	Matlab code	Schnitzcells *Rosenfeld et al.* (*2005*)		
Other	C code for simulations	GitHub link This study		

Bacterial Strains

All strains used in this study are constructed from the parent strain *E. coli* HG105 which is MG1655 with the *lac* operon deleted (MG1655$\Delta lacIZYA$). Auto-regulated TF (*lacI-mCherry*) is expressed from the *ybcN* locus and the TF-repressed target (*yfp*) is expressed from the *galK* locus with identical promoter sequence for both the TF and the target. Decoys are introduced on the pZE plasmid. In order to tune the degradation rate of the TF, three different ssrA tags were added to the C-terminus of the LacI-mCherry fusion protein. The tags used in this study are wildtype LAA tag (AANDENYALAA), DAS tag (AANDENYADAS) and DAS+4 tag (AANDENYSENYADAS)(*McGinness et al.* (*2006*)). For protein degradation tag experiments with LacI-mCherry fusion protein, HG105 with ΔsspB knockout is used as a parent strain to substantially moderate the protein degradation rate. It is also noteworthy that deletion of *sspB* gene did not affect the growth rate in any of the strains tested. Primers used in this study are listed in Table 1.

Microscopy

Bacterial cultures are grown overnight in 1 mL of LB in a 37°C incubator shaking at 250 rpm. Unless otherwise stated cultures grown overnight are diluted 2.5×10^3 fold to an initial OD of 0.002 into 1 mL of fresh M9 minimal media supplemented with 0.5% of one of the three different carbon

428 sources (Glucose, Glycerol or Acetate) or in Rich Defined Media (RDM, Teknova #M2105), allowed to
429 grow at 37°C until they reach an OD600 of 0.2 to 0.4 (0.1 for acetate) and harvested for microscopy.
430 Cells are diluted 1:3 in 1X PBS (in order to obtain isolated cells in microscope images) and 1μL is
431 spotted on a 2% low melting agarose pad (Invitrogen #16520050) made with 1X PBS. Cells grown in
432 RDM are cross-linked with paraformaldehyde before imaging to prevent shrinkage and osmotic
433 shock to the cells. An automated fluorescent microscope (Nikon TI-E) with a heating chamber set at
434 37°C is used to record multiple fields per sample (between 8-12 unique fields of view) resulting in
435 roughly 500 to 1000 individual cells per sample.

qPCR measurements for average plasmid copy number

437 We performed qPCR measurements in order to quantify the average copy number of the pZE
438 plasmid. Cells are grown as described for microscopic analysis and diluted 1:200 in Qiagen P1 lysis
439 buffer and allowed to sit on ice. Meanwhile, cells are plated at 10-5 dilutions on fresh LB plates in
440 order to determine the colony forming units per mL (CFU/mL). 25 μL of the lysate is diluted with 25
441 μL of 1X PBS and allowed to sit for 5 minutes. The cells are then diluted 1:100 into 1X cut smart
442 buffer from NEB. 20μL of the mixture is incubated with 0.5 μL of HindIII restriction enzyme for 30
443 minutes at 37°C followed by heat inactivation at 80°C for 20 minutes. The mixture is further diluted
444 1:10 and 4.2 μL is used as a template in a 20 μL qPCR reaction mixture. The pZE-1XOid plasmid is
445 purified using the Qiagen Plasmid Medi Prep kit and quantified using the Qubit dsDNA assay kit. A
446 standard curve is then prepared by diluting pZE-1XOid plasmid from 10^8 copies down to 10 copies.
447 The average copy number of the decoy plasmid per cell is computed by comparing the cT of the
448 sample to the standard curve and dividing by the number of cells in the sample.

Simulation methodology

450 To model the experiments and study the effect of decoy sites on the expression of a target gene
451 regulated by a negatively autoregulated TF gene, we develop a simple model of the experimental
452 system. In our model the auto-regulatory gene produces a protein (X) which forms a TF dimer
453 (R). We explicitly modeled TF as a dimer to incorporate the fact that LacI acts as a dimer in our
454 experimental system (the LacI-mCherry construct lacks the tetramerization domain (***Kipper et al.***
455 (***2018***)). Dimerization and de-dimerization steps occur at the rate k_p and k_m, respectively. The
456 TF binds to its own promoter (P_{TF}), to the promoter of the target gene (P_{target}), and to the decoy
457 sites (N) with a constant rate k_{on} per free TF per unit time. The off rate of the bound TF (k_{off}, the
458 unbinding rate) depends on the sequence identity and can be different for different promoters.
459 A bound TF unbinds from the promoters of the TF and target, and from the decoy sites at a rate
460 $k_{off,TF}$, $k_{off,target}$, and $k_{off,decoy}$ per unit time, respectively. A TF-free promoter produces an mRNA at
461 the rate β which is then translated into a protein at a rate α. The mRNA and the proteins are
462 degraded at the rate γ_m and γ, respectively. We assume that all proteins (free protein, TF bound to
463 promoter and TF bound to decoy sites) degrades with the same rate. Typically, the proteins in *E.*
464 *coli* are very stable with protein half-life greater than the cell cycle and the dominant contribution
465 to degradation comes from the dilution due to cell division. The degradation rate is thus given by
466 $\gamma = ln(2)/\tau_1 + ln(2)/\tau_2$, where τ_1 and τ_2 are protein half-life and cell division time, respectively. The
467 set of reactions describing the model above are listed in Appendix Figure 2A.

468 We implement the simulations for stochastic reaction systems using Gillespie's algorithm (***Gille-***
469 ***spie*** (***1977***)) in C programming. Each simulation is run for sufficiently long time ($\sim 10^6$ s) to reach
470 a steady state. Typically, for the rates used in this paper the steady state is achieved in 10^5 s or
471 less (see Appendix 6 Figure 1 for a sample time trace). Data for steady state distributions (TF and
472 target protein) are then recorded by sampling over time with a time interval (T_S) long enough for
473 the slowest reaction to occur 20 times on average (T_S = 20 over rate for slowest reaction). Mean
474 protein numbers in steady state for fold-change are calculated using at least 10^5 data points for
475 each single run.

Kinetic parameter estimation

To compare the results from experiments with our simulations we are required to find values for the kinetic on and off rate of LacI for different operator sites (Oid, O1 and O2), the transcription and translation rates, mRNA degradation rate, and the growth rates in different media. We directly measure growth rate for different media in our experiment (see Appendix 2). The on and off rates are related to the binding energy ($\Delta\epsilon$) through,

$$\frac{k_{\text{on}}}{k_{\text{off}} + \gamma} = \frac{exp(-\Delta\epsilon)}{N_{\text{ns}}}, \tag{1}$$

where $N_{\text{ns}} \sim 5 \times 10^6$ bps is the number of non-specific binding sites in the genome (which we take as the total number of bases) (*Phillips et al.* (*2013*)), k_{on} is the binding rate per free TF per unit time, k_{off} is the unbinding rate per unit time and γ is the decay rate of the TF. Experimental measurements of $\Delta\epsilon$ have been reported in many repeated experiments (*Brewster et al.* (*2014*); *Garcia and Phillips* (*2011*); *Razo-Mejia et al.* (*2018*)) and thus we constrain our choice of k_{on} and k_{off} such that we obtain affinities consistent with these measurements. Taking one data set (O1 regulated TF and O1 regulated target grown in glucose), we use maximum likelihood analysis to obtain the rates by varying k_{on} in a range 0.0015-0.003 s^{-1} (which sets the corresponding value of k_{off} to give $\Delta\epsilon_{\text{O1}} = 15.3k_{\text{B}}T$) (*Elf et al.* (*2007*); *Bremer and Dennis* (*2008*)), γ_{m}^{-1} in a range of 30-90 s (*Yu et al.* (*2006*); *Bremer and Dennis* (*2008*)), β in a range of 0.1-0.3 s^{-1} (*Kennell and Riezman* (*1977*)), and choosing α (*Cai et al.* (*2006*)) such that the constitutive number for the TF protein is in the range of 1000-2600; this parameter largely sets the "range" of our fold-change vs fold-change curves and this range of α reproduces the experimental range we see in those curves for this data set.

We then use this same on rate to derive the relevant off rates for O2 and Oid using their binding energies $\Delta\epsilon_{\text{O2}} = 13.9k_{\text{B}}T$, $\Delta\epsilon_{\text{Oid}} = 16.3k_{\text{B}}T$) and Equation 1. Interestingly, the binding affinity we measure for Oid is 0.7 $k_{\text{B}}T$ weaker than has been previously reported but is consistent with measurements of Oid binding affinity in our lab. Using this method, we find the k_{on} to be 0.0015 per TF per second, which yields k_{off} to be, O1=0.0015 s^{-1}, O2 = 0.0167 s^{-1} and Oid=0.0004 s^{-1}, consistent with previous findings (*Elf et al.* (*2007*); *Hammar et al.* (*2014*); *Jones et al.* (*2014*); *Razo-Mejia et al.* (*2018*)). All other rates are listed in Table 2. Importantly, this process is not meant to precisely determine the exact quantitative parameters of LacI binding, and it is not a formal fit, but rather an estimate that provides us with realistic prediction of regulation from our simulations using molecular parameters that are consistent with available direct kinetic measurements (*Chen et al.* (*2015*); *Elf et al.* (*2007*); *Sanchez et al.* (*2011*); *Yu et al.* (*2006*)).

Data Analysis

Data analysis is performed using a modified version of the Matlab code Schnitzcells (*Rosenfeld et al.* (*2005*)). We use this code to segment the phase images of each sample to identify single cells. Mean pixel intensities of YFP and mCherry signals are extracted from the segmented phase mask for each individual cell using regionprops, an inbuilt function in matlab. The background fluorescence is calculated by averaging the mean intensity of the inverse phase mask upon eroding the regions around the segmented cell masks. The background fluorescence value of a particular frame was subtracted from the mean pixel intensity of cells in the same frame (see Appendix 1). Finally, the autofluorescence value were calculated using the same procedure for cells that do not express either YFP or mCherry and the average autofluorescence value of these cells is subtracted from each measured YFP or mCherry value. Resulting mean pixel intensity of mCherry signal was corrected for the crosstalk from YFP signal. Crosstalk between different channels can be measured by determining the difference between the autofluorescence of a strain without a given fluorophore in the presence of the other fluorophore (highly expressed). We find that under our microscope 0.25% ($\gamma_{\text{cross}} = 0.0025$) of YFP signals can be seen in the mCherry channel whereas mCherry channel has no crosstalk in the YFP channel. Hence, we correct for this crosstalk by subtracting the mean pixel intensity of YFP signal times the γ_{cross} from the mean pixel intensity of mCherry signal. The

per-pixel fluorescence values of mCherry and YFP of each cell is then multiplied by the area of the cell to account for the total fluorescence. Fold-change in expression of the mCherry and YFP is calculated by dividing the corresponding values of the constitutive strains (discussed in Appendix 4). At least 500 individual cells were analyzed per sample and binned according to the mCherry values. Any bin with less than 50 data points is excluded. Unless otherwise stated, each data point represents the bootstrapped mean of all data points in a given bin and the error bar represents the standard deviation of the bootstrapped mean.

References

Acar M, Mettetal JT, van Oudenaarden A. Stochastic Switching as a Survival Strategy in Fluctuating Environments. Nature Genetics. 2008; 40(4):471–475. https://doi.org/10.1038/ng.110, doi: 10.1038/ng.110.

Alon U. An Introduction to Systems Biology: Design Principles of Biological Circuits. Taylor & Francis; 2006. https://books.google.com/books?id=tcxCkIxzCO4C.

Alon U. Network Motifs: Theory and Experimental Approaches. Nature Reviews Genetics. 2007; 8(6):450–461. https://doi.org/10.1038/nrg2102, doi: 10.1038/nrg2102.

Arnone MI. Bringing Order to Organogenesis. Nature Genetics. 2002; 30(4):348–350. https://doi.org/10.1038/ng0402-348, doi: 10.1038/ng0402-348.

Assaf M, Roberts E, Luthey-Schulten Z. Determining the Stability of Genetic Switches: Explicitly Accounting for mRNA Noise. Physical Review Letters. 2011; 106(24):248102. https://link.aps.org/doi/10.1103/PhysRevLett.106.248102, doi: 10.1103/PhysRevLett.106.248102.

Bakk A, Metzler R. Nonspecific Binding of the OR Repressors CI and Cro of Bacteriophage . Journal of Theoretical Biology. 2004; 231(4):525–533. http://www.sciencedirect.com/science/article/pii/S0022519304003285, doi: https://doi.org/10.1016/j.jtbi.2004.07.007.

Becskei A, Serrano L. Engineering Stability in Gene Networks by Autoregulation. Nature. 2000; 405(6786):590–593. https://doi.org/10.1038/35014651, doi: 10.1038/35014651.

Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Phillips R. Transcriptional Regulation by the Numbers: Models. Current Opinion in Genetics & Development. 2005; 15(2):116–124. http://www.sciencedirect.com/science/article/pii/S0959437X05000304, doi: https://doi.org/10.1016/j.gde.2005.02.007.

Bremer H, Dennis P. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. EcoSal Plus. 2008; https://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosal.5.2.3, doi: doi:10.1128/ecosal.5.2.3.

Brewster R, Weinert F, Garcia H, Song D, Rydenfelt M, Phillips R. The Transcription Factor Titration Effect Dictates Level of Gene Expression. Cell. 2014; 156(6):1312–1323. http://www.sciencedirect.com/science/article/pii/S0092867414002219, doi: https://doi.org/10.1016/j.cell.2014.02.022.

Burger A, Walczak AM, Wolynes PG. Abduction and Asylum in the Lives of Transcription Factors. Proceedings of the National Academy of Sciences. 2010; 107(9):4016–4021. https://www.pnas.org/content/pnas/107/9/4016.full.pdf, doi: 10.1073/pnas.0915138107.

Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. Nature. 2006; 440(7082):358–62. https://doi.org/10.1038/nature04599, doi: 10.1038/nature04599.

Carey LB, van Dijk D, Sloot PMA, Kaandorp JA, Segal E. Promoter Sequence Determines the Relationship between Expression Level and Noise. PLOS Biology. 2013; 11(4):e1001528. https://doi.org/10.1371/journal.pbio.1001528, doi: 10.1371/journal.pbio.1001528.

Chen H, Shiroguchi K, Ge H, Xie XS. Genome-wide Study of mRNA Degradation and Transcript Elongation in *Escherichia coli*. Molecular Systems Biology. 2015; 11(1):781. https://www.embopress.org/doi/abs/10.15252/msb.20145794, doi: 10.15252/msb.20145794.

Choi PJ, Cai L, Frieda K, Xie XS. A Stochastic Single-Molecule Event Triggers Phenotype Switching of a Bacterial Cell. Science. 2008; 322(5900):442–446. https://science.sciencemag.org/content/sci/322/5900/442.full.pdf, doi: 10.1126/science.1161427.

Cooper S, Helmstetter CE. Chromosome replication and the division cycle of Escherichia coli Br. Journal of Molecular Biology. 1968; 31(3):519–540. http://www.sciencedirect.com/science/article/pii/0022283668904257, doi: https://doi.org/10.1016/0022-2836(68)90425-7.

Davidson EH. The Regulatory Genome: Gene Regulatory Networks in Development and Evolution. Elsevier Science; 2006. https://books.google.com/books?id=xMCfngEACAAJ.

Elf J, Li GW, Xie XS. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. Science. 2007; 316(5828):1191–1194. https://science.sciencemag.org/content/sci/316/5828/1191.full.pdf, doi: 10.1126/science.1141967.

Friedman N, Vardi S, Ronen M, Alon U, Stavans J. Precise Temporal Modulation in the Response of the SOS DNA Repair Network in Individual Bacteria. PLOS Biology. 2005; 3(7):e238. https://doi.org/10.1371/journal.pbio.0030238, doi: 10.1371/journal.pbio.0030238.

Garcia HG, Phillips R. Quantitative Dissection of the Simple Repression Input–Output Function. Proceedings of the National Academy of Sciences. 2011; 108(29):12173–12178. https://www.pnas.org/content/pnas/108/29/12173.full.pdf, doi: 10.1073/pnas.1015616108.

Garcia H, Sanchez A, Boedicker J, Osborne M, Gelles J, Kondev J, Phillips R. Operator Sequence Alters Gene Expression Independently of Transcription Factor Occupancy in Bacteria. Cell Reports. 2012; 2(1):150–161. http://www.sciencedirect.com/science/article/pii/S2211124712001647, doi: https://doi.org/10.1016/j.celrep.2012.06.004.

Gaudet J, Mango SE. Regulation of Organogenesis by the Caenorhabditis elegans FoxA Protein PHA-4. Science. 2002; 295(5556):821–825. https://science.sciencemag.org/content/sci/295/5556/821.full.pdf, doi: 10.1126/science.1065175.

Gerosa L, Kochanowski K, Heinemann M, Sauer U. Dissecting specific and global transcriptional regulation of bacterial gene expression. Molecular Systems Biology. 2013; 9(1):658. https://www.embopress.org/doi/abs/10.1038/msb.2013.14, doi: 10.1038/msb.2013.14.

Gillespie DT. Exact Stochastic Simulation of Coupled Chemical Reactions. The Journal of Physical Chemistry. 1977; 81(25):2340–2361. https://doi.org/10.1021/j100540a008, doi: 10.1021/j100540a008.

Gillespie DT. Stochastic Simulation of Chemical Kinetics. Annual Review of Physical Chemistry. 2007; 58(1):35–55. https://www.annualreviews.org/doi/abs/10.1146/annurev.physchem.58.032806.104637, doi: 10.1146/annurev.physchem.58.032806.104637.

Hahl SK, Kremling A. A Comparison of Deterministic and Stochastic Modeling Approaches for Biochemical Reaction Systems: On Fixed Points, Means, and Modes. Frontiers in Genetics. 2016; 7(157). https://www.frontiersin.org/article/10.3389/fgene.2016.00157, doi: 10.3389/fgene.2016.00157.

Hammar P, Walldén M, Fange D, Persson F, Baltekin O, Ullman G, Leroy P, Elf J. Direct Measurement of Transcription Factor Dissociation Excludes a Simple Operator Occupancy Model for Gene Regulation. Nature Genetics. 2014; 46:405. https://doi.org/10.1038/ng.2905, doi: 10.1038/ng.2905 https://www.nature.com/articles/ng.2905##supplementary-information.

Hornos JEM, Schultz D, Innocentini GCP, Wang J, Walczak AM, Onuchic JN, Wolynes PG. Self-regulating gene: An exact solution. Physical Review E. 2005; 72(5):051907. https://link.aps.org/doi/10.1103/PhysRevE.72.051907, doi: 10.1103/PhysRevE.72.051907.

Jones DL, Brewster RC, Phillips R. Promoter Architecture Dictates Cell-to-Cell Variability in Gene Expression. Science. 2014; 346(6216):1533–1536. https://science.sciencemag.org/content/sci/346/6216/1533.full.pdf, doi: 10.1126/science.1255301.

Jun S, Si F, Pugatch R, Scott M. Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review. Reports on Progress in Physics. 2018; 81(5):056601. http://dx.doi.org/10.1088/1361-6633/aaa628, doi: 10.1088/1361-6633/aaa628.

Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in Gene Expression: From Theories to Phenotypes. Nature Reviews Genetics. 2005; 6(6):451–464. https://doi.org/10.1038/nrg1615, doi: 10.1038/nrg1615.

Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, Leibler S, Surette MG, Alon U. Ordering Genes in a Flagella Pathway by Analysis of Expression Kinetics from Living Bacteria. Science. 2001; 292(5524):2080–2083. https://science.sciencemag.org/content/sci/292/5524/2080.full.pdf, doi: 10.1126/science.1058758.

Kemme CA, Nguyen D, Chattopadhyay A, Iwahara J. Regulation of Transcription Factors via Natural Decoys in Genomic DNA. Transcription. 2016; 7(4):115–120. https://www.ncbi.nlm.nih.gov/pubmed/27384377https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4984682/, doi: 10.1080/21541264.2016.1188873.

Kennell D, Riezman H. Transcription and translation initiation frequencies of the Escherichia coli lac operon. Journal of Molecular Biology. 1977; 114(1):1–21. http://www.sciencedirect.com/science/article/pii/0022283677902790, doi: https://doi.org/10.1016/0022-2836(77)90279-0.

Kinney JB, Murugan A, Callan CG, Cox EC. Using Deep Sequencing to Characterize the Biophysical Mechanism of a Transcriptional Regulatory Sequence. Proceedings of the National Academy of Sciences. 2010; 107(20):9158–9163. https://www.pnas.org/content/pnas/107/20/9158.full.pdf, doi: 10.1073/pnas.1004290107.

Kipper K, Eremina N, Marklund E, Tubasum S, Mao G, Lehmann LC, Elf J, Deindl S. Structure-guided approach to site-specific fluorophore labeling of the lac repressor LacI. PLOS ONE. 2018; 13(6):e0198416. https://doi.org/10.1371/journal.pone.0198416, doi: 10.1371/journal.pone.0198416.

Klumpp S, Hwa T. Growth-Rate-Dependent Partitioning of RNA Polymerases in Bacteria. Proceedings of the National Academy of Sciences. 2008; 105(51):20245–20250. https://www.pnas.org/content/pnas/105/51/20245.full.pdf, doi: 10.1073/pnas.0804953105.

Klumpp S, Zhang Z, Hwa T. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. Cell. 2009; 139(7):1366–1375. https://www.ncbi.nlm.nih.gov/pubmed/20064380https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2818994/, doi: 10.1016/j.cell.2009.12.001.

Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, Arkin AP, Endy D, Church GM. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. Proceedings of the National Academy of Sciences. 2013; 110(34):14024–14029. https://www.pnas.org/content/pnas/110/34/14024.full.pdf, doi: 10.1073/pnas.1301301110.

Kuhlman T, Zhang Z, Saier MH, Hwa T. Combinatorial Transcriptional Control of the Lactose Operon of Escherichia coli. Proceedings of the National Academy of Sciences. 2007; 104(14):6043–6048. https://www.pnas.org/content/pnas/104/14/6043.full.pdf, doi: 10.1073/pnas.0606717104.

Kuhlman TE, Cox EC. Gene location and DNA density determine transcription factor distributions in Escherichia coli. Molecular Systems Biology. 2012; 8(1):610. https://www.embopress.org/doi/abs/10.1038/msb.2012.42, doi: 10.1038/msb.2012.42.

Lee TH, Maheshri N. A Regulatory Role for Repeated Decoy Transcription Factor Binding Sites in Target Gene Expression. Molecular Systems Biology. 2012; 8(1):576. https://www.embopress.org/doi/abs/10.1038/msb.2012.7, doi: 10.1038/msb.2012.7.

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, et al. Transcriptional Regulatory Networks in Saccharomyces cerevisiae. Science. 2002; 298(5594):799–804. https://science.sciencemag.org/content/sci/298/5594/799.full.pdf, doi: 10.1126/science.1075090.

Lorenzin F, Benary U, Baluapuri A, Walz S, Jung LA, von Eyss B, Kisker C, Wolf J, Eilers M, Wolf E. Different Promoter Affinities Account for Specificity in MYC-Dependent Gene Regulation. eLife. 2016; 5:e15161. https://doi.org/10.7554/eLife.15161, doi: 10.7554/eLife.15161.

Ma HW, Kumar B, Ditges U, Gunzer F, Buer J, Zeng AP. An Extended Transcriptional Regulatory Network of Escherichia coli and Analysis of its Hierarchical Structure and Network Motifs. Nucleic Acids Research. 2004; 32(22):6643–6649. https://doi.org/10.1093/nar/gkh1009, doi: 10.1093/nar/gkh1009.

Maerkl SJ, Quake SR. A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. Science. 2007; 315(5809):233–237. https://science.sciencemag.org/content/sci/315/5809/233.full.pdf, doi: 10.1126/science.1131007.

Mangan S, Alon U. Structure and Function of the Feed-Forward Loop Network Motif. Proceedings of the National Academy of Sciences. 2003; 100(21):11980–11985. https://www.pnas.org/content/pnas/100/21/11980.full.pdf, doi: 10.1073/pnas.2133841100.

Marr AG. Growth Rate of Escherichia coli. Microbiological reviews. 1991; 55(2):316–333. https://www.ncbi.nlm.nih.gov/pubmed/1886524https://www.ncbi.nlm.nih.gov/pmc/articles/PMC372817/.

**McGinness KE**, Baker TA, Sauer RT. Engineering Controllable Protein Degradation. Molecular Cell. 2006; 22(5):701–707. http://www.sciencedirect.com/science/article/pii/S1097276506003261, doi: https://doi.org/10.1016/j.molcel.2006.04.027.

McGinness KE, Bolon DN, Kaganovich M, Baker TA, Sauer RT. Altered Tethering of the SspB Adaptor to the ClpXP Protease Causes Changes in Substrate Delivery. Journal of Biological Chemistry. 2007; 282(15):11465–11473. http://www.jbc.org/content/282/15/11465.abstract, doi: 10.1074/jbc.M610671200.

Milias-Argeitis A, Engblom S, Bauer P, Khammash M. Stochastic focusing coupled with negative feedback enables robust regulation in biochemical reaction networks. Journal of The Royal Society Interface. 2015; 12(113):20150831. https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2015.0831, doi: doi:10.1098/rsif.2015.0831.

Mirny L. Nucleosome-Mediated Cooperativity Between Transcription Factors. Nature Precedings. 2009; https://doi.org/10.1038/npre.2009.2796.1, doi: 10.1038/npre.2009.2796.1.

Neidhardt FC, Curtiss R. Escherichia coli and Salmonella: Cellular and Molecular Biology. ASM Press; 1996. https://books.google.com/books?id=haqnAQAACAAJ.

Nevozhay D, Adams RM, Murphy KF, Josić K, Balázsi G. Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression. Proceedings of the National Academy of Sciences. 2009; 106(13):5123–5128. https://www.pnas.org/content/pnas/106/13/5123.full.pdf, doi: 10.1073/pnas.0809901106.

Ochab-Marcinek A, Jędrak J, Tabaka M. Hill Kinetics as a Noise Filter: the Role of Transcription Factor Autoregulation in Gene Cascades. Physical Chemistry Chemical Physics. 2017; 19(33):22580–22591. http://dx.doi.org/10.1039/C7CP00743D, doi: 10.1039/C7CP00743D.

Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, Müller-Hill B. Quality and Position of the Three lac Operators of E. coli Define Efficiency of Repression. The EMBO journal. 1994; 13(14):3348–3355. https://www.ncbi.nlm.nih.gov/pubmed/8045263https://www.ncbi.nlm.nih.gov/pmc/articles/PMC395232/.

Oehler S, Eismann ER, Krämer H, Müller-Hill B. The Three Operators of the lac Operon Cooperate in Repression. The EMBO Journal. 1990; 9(4):973–979. https://www.embopress.org/doi/abs/10.1002/j.1460-2075.1990.tb08199.x, doi: 10.1002/j.1460-2075.1990.tb08199.x.

Phillips R, Kondev J, Theriot J, Garcia H. Physical Biology of the Cell. Garland Science; 2013. https://books.google.com/books?id=iTBvQgAACAAJ.

Phillips R, Belliveau NM, Chure G, Garcia HG, Razo-Mejia M, Scholes C. Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression. Annual Review of Biophysics. 2019; 48(1):121–163. https://www.annualreviews.org/doi/abs/10.1146/annurev-biophys-052118-115525, doi: 10.1146/annurev-biophys-052118-115525.

Razo-Mejia M, Barnes SL, Belliveau NM, Chure G, Einav T, Lewis M, Phillips R. Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. Cell Systems. 2018; 6(4):456–469.e10. https://doi.org/10.1016/j.cels.2018.02.004, doi: 10.1016/j.cels.2018.02.004.

Rodrigo G, Bajic D, Elola I, Poyatos JF. Antagonistic Autoregulation Speeds up a Homogeneous Response in Escherichia coli. Scientific Reports. 2016; 6:36196. https://doi.org/10.1038/srep36196, doi: 10.1038/srep36196 https://www.nature.com/articles/srep36196##supplementary-information.

Ronen M, Rosenberg R, Shraiman BI, Alon U. Assigning Numbers to the Arrows: Parameterizing a Gene Regulation Network by Using Accurate Expression Kinetics. Proceedings of the National Academy of Sciences. 2002; 99(16):10555–10560. https://www.pnas.org/content/pnas/99/16/10555.full.pdf, doi: 10.1073/pnas.152046799.

Rosenfeld N, Elowitz MB, Alon U. Negative Autoregulation Speeds the Response Times of Transcription Networks. Journal of Molecular Biology. 2002; 323(5):785–793. http://www.sciencedirect.com/science/article/pii/S0022283602009944, doi: https://doi.org/10.1016/S0022-2836(02)00994-4.

Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. Gene Regulation at the Single-Cell Level. Science. 2005; 307(5717):1962–1965. https://science.sciencemag.org/content/sci/307/5717/1962.full.pdf, doi: 10.1126/science.1106914.

Rydenfelt M, Cox RS, Garcia H, Phillips R. Statistical mechanical model of coupled transcription from multiple promoters due to transcription factor titration. Physical Review E. 2014; 89(1):012702. https://link.aps.org/doi/10.1103/PhysRevE.89.012702, doi: 10.1103/PhysRevE.89.012702.

Sanchez A, Garcia HG, Jones D, Phillips R, Kondev J. Effect of Promoter Architecture on the Cell-to-Cell Variability in Gene Expression. PLOS Computational Biology. 2011; 7(3):e1001100. https://doi.org/10.1371/journal.pcbi.1001100, doi: 10.1371/journal.pcbi.1001100.

Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeida D, García-Sotelo JS, Alquicira-Hernández K, Muñiz-Rascado LJ, Peña-Loredo P, Ishida-Gutiérrez C, Velázquez-Ramírez DA, Del Moral-Chávez V, Bonavides-Martínez C, Méndez-Cruz CF, Galagan J, Collado-Vides J. RegulonDB v 10.5: Tackling Challenges to Unify Classic and High Throughput Knowledge of Gene Regulation in E. coli K-12. Nucleic Acids Research. 2018; 47(D1):D212–D220. https://doi.org/10.1093/nar/gky1077, doi: 10.1093/nar/gky1077.

Savageau MA. Significance of Autogenously Regulated and Constitutive Synthesis of Regulatory Proteins in Repressible Biosynthetic Systems. Nature. 1975; 258(5532):208–214. https://doi.org/10.1038/258208a0, doi: 10.1038/258208a0.

Schikora-Tamarit MA, Lopez-Grado I Salinas G, Gonzalez-Navasa C, Calderón I, Marcos-Fa X, Sas M, Carey LB. Promoter Activity Buffering Reduces the Fitness Cost of Misregulation. Cell Reports. 2018; 24(3):755–765. http://www.sciencedirect.com/science/article/pii/S2211124718309860, doi: https://doi.org/10.1016/j.celrep.2018.06.059.

Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M. The quantitative and condition-dependent Escherichia coli proteome. Nature Biotechnology. 2016; 34(1):104–110. https://doi.org/10.1038/nbt.3418, doi: 10.1038/nbt.3418.

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. Science. 2010; 330(6007):1099–1102. https://science.sciencemag.org/content/sci/330/6007/1099.full.pdf, doi: 10.1126/science.1192588.

Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N. Module Networks: Identifying Regulatory Modules and their Condition-Specific Regulators from Gene Expression Data. Nature Genetics. 2003; 34(2):166–176. https://doi.org/10.1038/ng1165, doi: 10.1038/ng1165.

Semsey S, Krishna S, Erdossy J, Horváth P, Orosz L, Sneppen K, Adhya S. Dominant Negative Autoregulation Limits Steady-State Repression Levels in Gene Networks. Journal of bacteriology. 2009; 191(14):4487–4491. https://www.ncbi.nlm.nih.gov/pubmed/19429616https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2704710/, doi: 10.1128/JB.00056-09.

Shahrezaei V, Swain PS. Analytical Distributions for Stochastic Gene Expression. Proceedings of the National Academy of Sciences. 2008; 105(45):17256–17261. https://www.pnas.org/content/pnas/105/45/17256.full.pdf, doi: 10.1073/pnas.0803850105.

Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, Keren L, Yakhini Z, Weinberger A, Segal E. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nature Biotechnology. 2012; 30(6):521–530. https://doi.org/10.1038/nbt.2205, doi: 10.1038/nbt.2205.

Shen-Orr SS, Milo R, Mangan S, Alon U. Network Motifs in the Transcriptional Regulation Network of Escherichia coli. Nature Genetics. 2002; 31(1):64–68. https://doi.org/10.1038/ng881, doi: 10.1038/ng881.

Stamatakis M, Zygourakis K. Deterministic and stochastic population-level simulations of an artificial lac operon genetic network. BMC Bioinformatics. 2011; 12(1):301. https://doi.org/10.1186/1471-2105-12-301, doi: 10.1186/1471-2105-12-301.

Stormo GD. DNA Binding Sites: Representation and Discovery. Bioinformatics. 2000; 16(1):16–23. https://doi.org/10.1093/bioinformatics/16.1.16, doi: 10.1093/bioinformatics/16.1.16.

Tkačik G, Callan CG, Bialek W. Information Flow and Optimization in Transcriptional Regulation. Proceedings of the National Academy of Sciences. 2008; 105(34):12265–12270. https://www.pnas.org/content/pnas/105/34/12265.full.pdf, doi: 10.1073/pnas.0806077105.

Volkmer B, Heinemann M. Condition-Dependent Cell Volume and Concentration of Escherichia coli to Facilitate Data Conversion for Systems Biology Modeling. PLOS ONE. 2011; 6(7):e23126. https://doi.org/10.1371/journal.pone.0023126, doi: 10.1371/journal.pone.0023126.

Weinert FM, Brewster RC, Rydenfelt M, Phillips R, Kegel WK. Scaling of Gene Expression with Transcription-Factor Fugacity. Physical Review Letters. 2014; 113(25):258101. https://link.aps.org/doi/10.1103/PhysRevLett.113.258101, doi: 10.1103/PhysRevLett.113.258101.

Weirauch MT, Cote A, Norel R, Annala M, Zhao Y, Riley TR, Saez-Rodriguez J, Cokelaer T, Vedenko A, Talukder S, Consortium D, Agius P, Arvey A, Bucher P, Callan Jr CG, Chang CW, Chen CY, Chen YS, Chu YW, Grau J, et al. Evaluation of Methods for Modeling Transcription Factor Sequence Specificity. Nature Biotechnology. 2013; 31:126. https://doi.org/10.1038/nbt.2486, doi: 10.1038/nbt.2486 https://www.nature.com/articles/nbt.2486##supplementary-information.

Yu H, Luscombe NM, Qian J, Gerstein M. Genomic Analysis of Gene Expression Relationships in Transcriptional Regulatory Networks. Trends in Genetics. 2003; 19(8):422–427. http://www.sciencedirect.com/science/article/pii/S0168952503001756, doi: https://doi.org/10.1016/S0168-9525(03)00175-6.

Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. Science. 2006; 311(5767):1600–1603. https://science.sciencemag.org/content/sci/311/5767/1600.full.pdf, doi: 10.1126/science.1119623.

Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U. Just-in-Time Transcription Program in Metabolic Pathways. Nature Genetics. 2004; 36(5):486–491. https://doi.org/10.1038/ng1348, doi: 10.1038/ng1348.

785 **Figures and Tables**
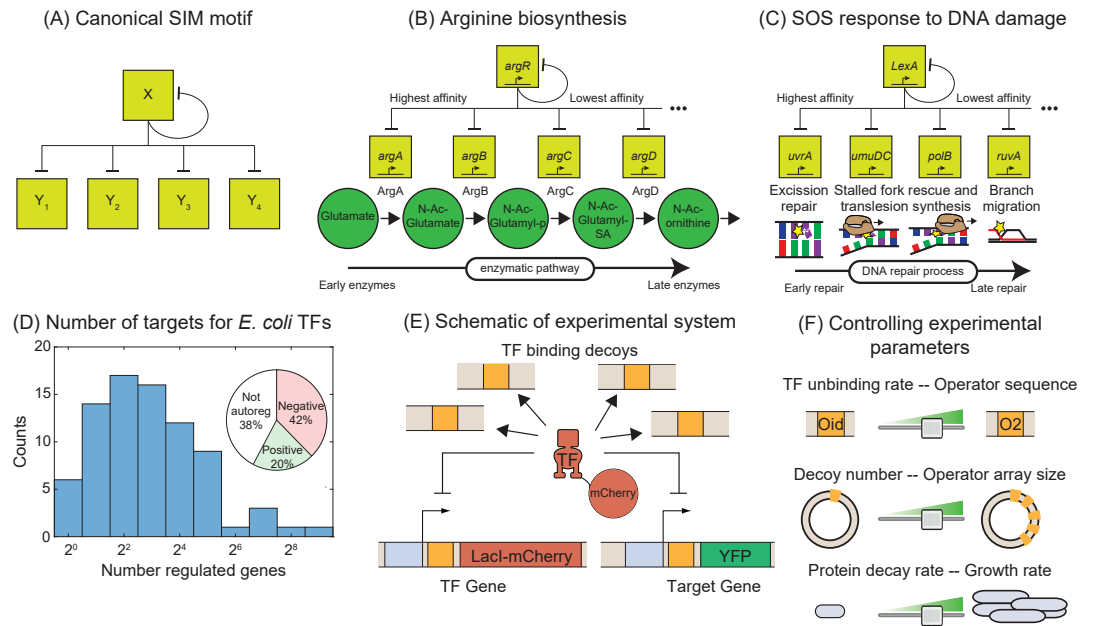
Figure 1. Synthetic approach to exploring the negative SIM motif. (A) Schematic of a canonical SIM motif: A single TF regulates itself and several other genes. (B and C) Examples of SIM motifs in *E. coli*. (B) ArgR is a transcriptional regulator of arginine biosynthesis. It auto-regulates itself and genes involved in different steps of arginine biosynthesis with precision in expression starting from the first enzyme of the pathway down to the last. This precise ordering is thought to originate from a corresponding ordering in TF binding affinities of the target genes. (C) LexA is the master regulator of SOS pathway and is actively degraded in response to DNA damage. LexA auto-represses itself and represses a set of other genes involved in DNA repair. In this case the early response genes have low affinity for the repressor while the late acting genes have high affinity, enabling temporal ordering of the response. (D) Histogram showing the number of known regulated genes for every TF in *E. coli*. Inset shows different modes of regulation of the TF genes. 62% of the TF genes are autoregulated with 42% negatively autoregulated and 20% positively auotregulated. (E) Schematic of the experimental model of a SIM motif used in this study. Here, LacI-mCherry is the model TF and YFP is the protein product of the target gene. Decoys sites are used to control the network size by simulating the demand of other target genes in the SIM motif. (F) Representation of the tunable parameter space detailed in this study. We can systematically tune the TF unbinding rate, number of decoys and protein degradation rate in the experimental system and adjust these parameters accordingly in simulations.
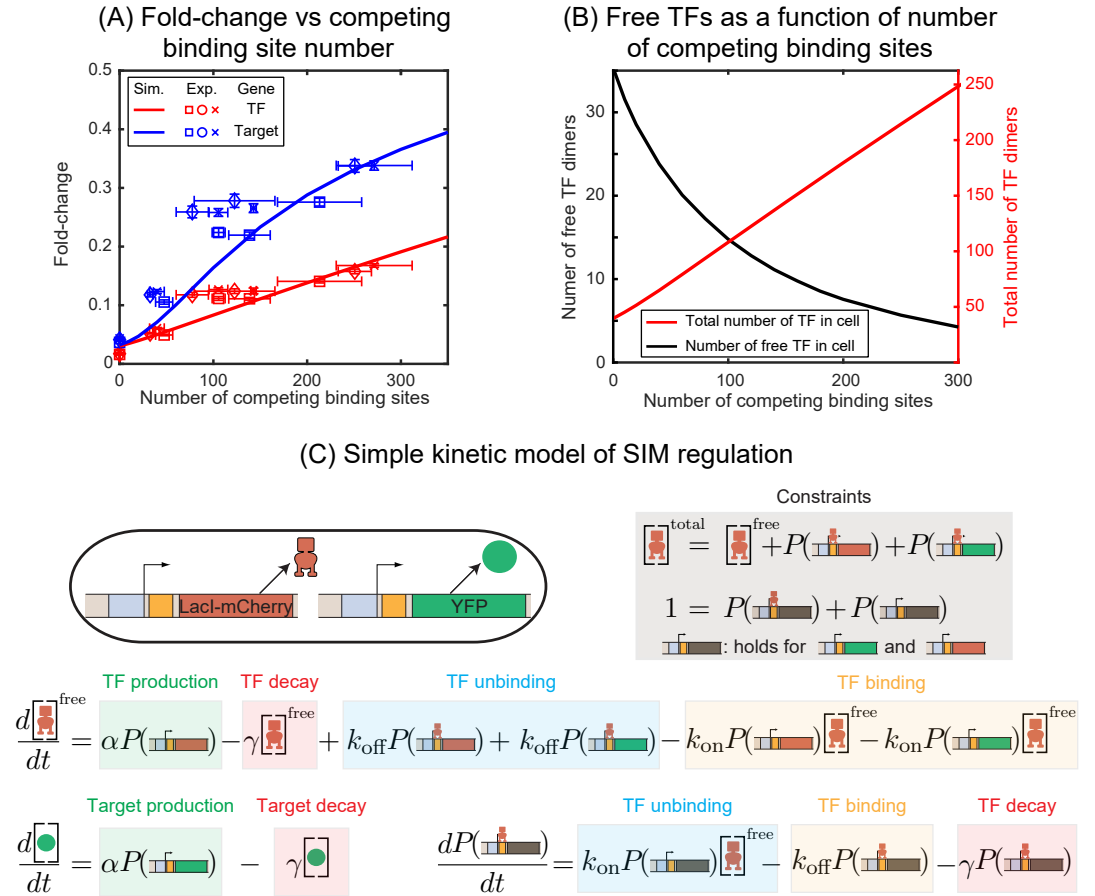
Figure 2. Fold-change in target and TF genes with network size. (A) Fold-change in the expression level of both the autoregulated gene (red) and the TF's target gene (blue) as a function of the number of competing binding sites present. Simulation data is shown as solid curves. Different symbols represent independent biological replicates. Each data point in y-axis is the bootstrapped mean of individual decoy strains and the error bars represent the standard deviation of bootstrapped mean. Each data point in x-axis is the mean of three technical replicates and the error bar is the corresponding standard deviation. (B) Increasing the number of competing binding sites increases the expression of both the TF (red line) and target genes by lowering the overall number of free TFs (black line). (C) Simple kinetic model describing the SIM motif using mass action equilibrium kinetics. For compactness of the figure the reactions involving the decoy binding sites, dimerization/dedimerization of TF monomers, and transcription steps are not shown. Full reactions of the model are described in Appendix 6.
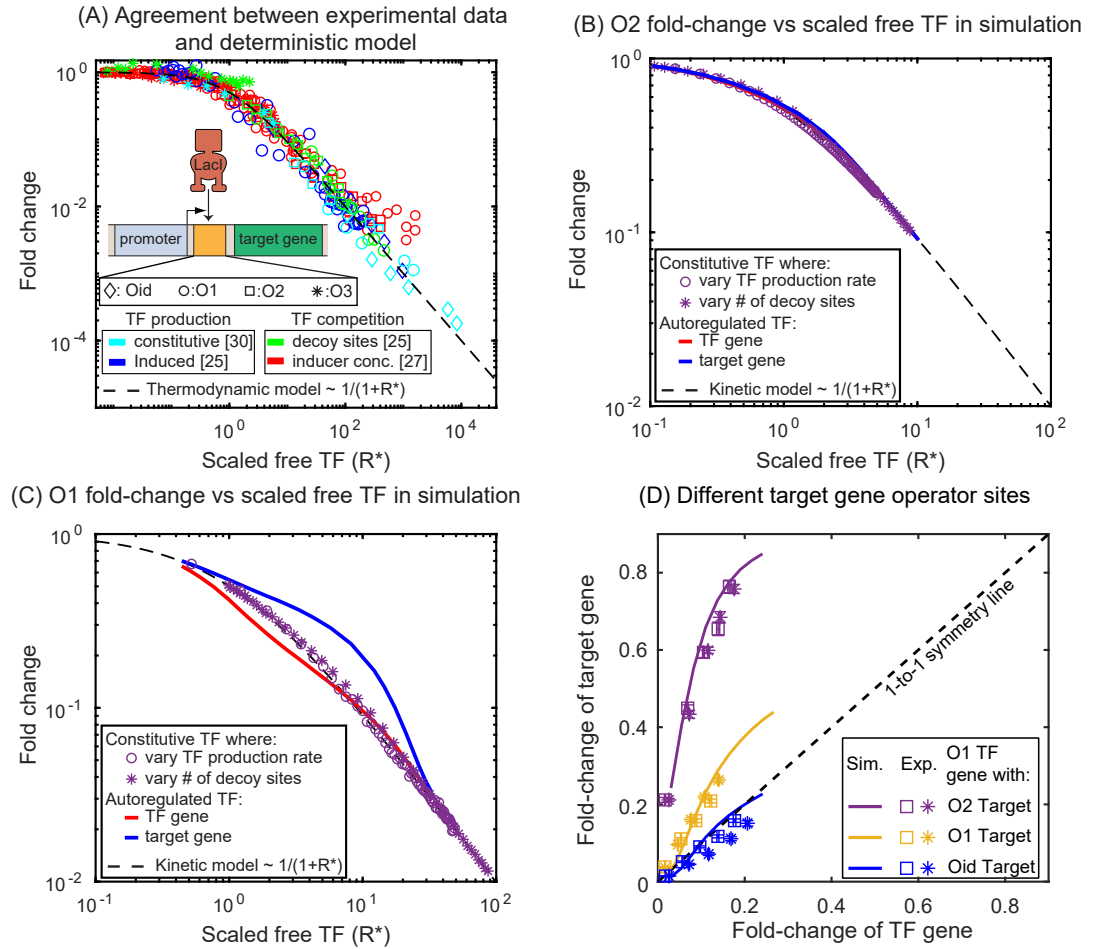
Figure 3. Comparison of SIM motif fold-change data to deterministic model predictions. (A) Fold-change vs scaled free TF in the thermodynamic model for a collection of simple repression data (open circles) where free TF is controlled through a diverse range of mechanisms. The data collapse to the deterministic model predictions (dashed curve). (B-C) Fold-change vs scaled free TF in simulations using the actual free TF obtained from simulation. The data for a constitutive expressed TF where free TF is varied by changing TF production rate (purple circles) or number of decoy sites (purple stars) collapses to the deterministic solution, however, the regulation of genes in the SIM motif (target: red line, TF gene: blue line) both diverge from the deterministic solution in opposing ways, giving rise not only to asymmetry but a disagreement with deterministic modeling for both genes. (D) Fold-change in the target gene versus fold-change in the TF gene. Each data point is the bootstrapped mean of fold-change in TF and target expression across hundreds of cells with a given number of competing binding sites and error bars represent the standard deviation of the bootstrapped mean. Different symbols represent independent biological replicates. In all cases the TF gene is regulated by an O1 binding sites whereas the target is regulated by (in order of weakest binding to strongest binding): O2 (purple), O1 (yellow) or Oid (blue). Simulation data is shown as solid curves.
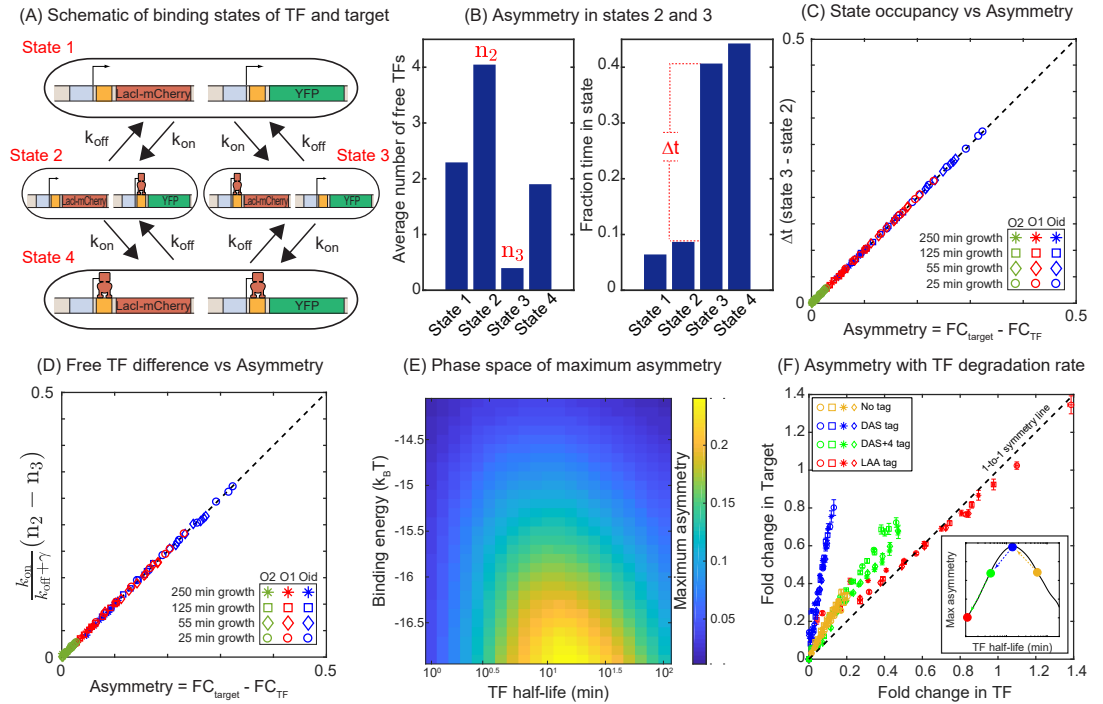
Figure 4. Mechanism of regulatory asymmetry. (A) Schematic of the TF-operator occupancy with their corresponding transition rates. The k_{on} for transition from state 1 to state 2 or state 3 will be identical and hence cannot account for the asymmetry. State 2 and state 3 on the other hand, will encounter a difference in the free TF concentration and hence the k_{on} for transition from one of these states to state 4 will be different; thus, accounting for the asymmetry in expression between the TF and the target. (B) Plot showing the average number of free TFs in different states and fraction of time cells spends in each of the given state in the simulation. (C) Plot showing asymmetry as a function of fractional time difference between state 2 and state 3. (D) Plot showing asymmetry as a function of difference in free TF concentration between state 3 and state 2. (E) Heat map showing the phase space of maximum asymmetry as a function of binding affinity for the TF and its half-life. (F) Tuning the TF degradation rate influences the extent of asymmetry observed in the SIM module. Yellow points corresponds to the system with no degradation tags; Blue points corresponds to degradation by a "weak" or "slow" tag (DAS tag with a rate of 0.00063 per min per enzyme); Green points corresponds to a slightly faster tag (DAS+4 with a rate of 0.0011 per min per enzyme); Red points corresponds to a very fast tag (LAA tag with a rate of 0.21 per min per enzyme). Different symbols represent independent biological replicates.
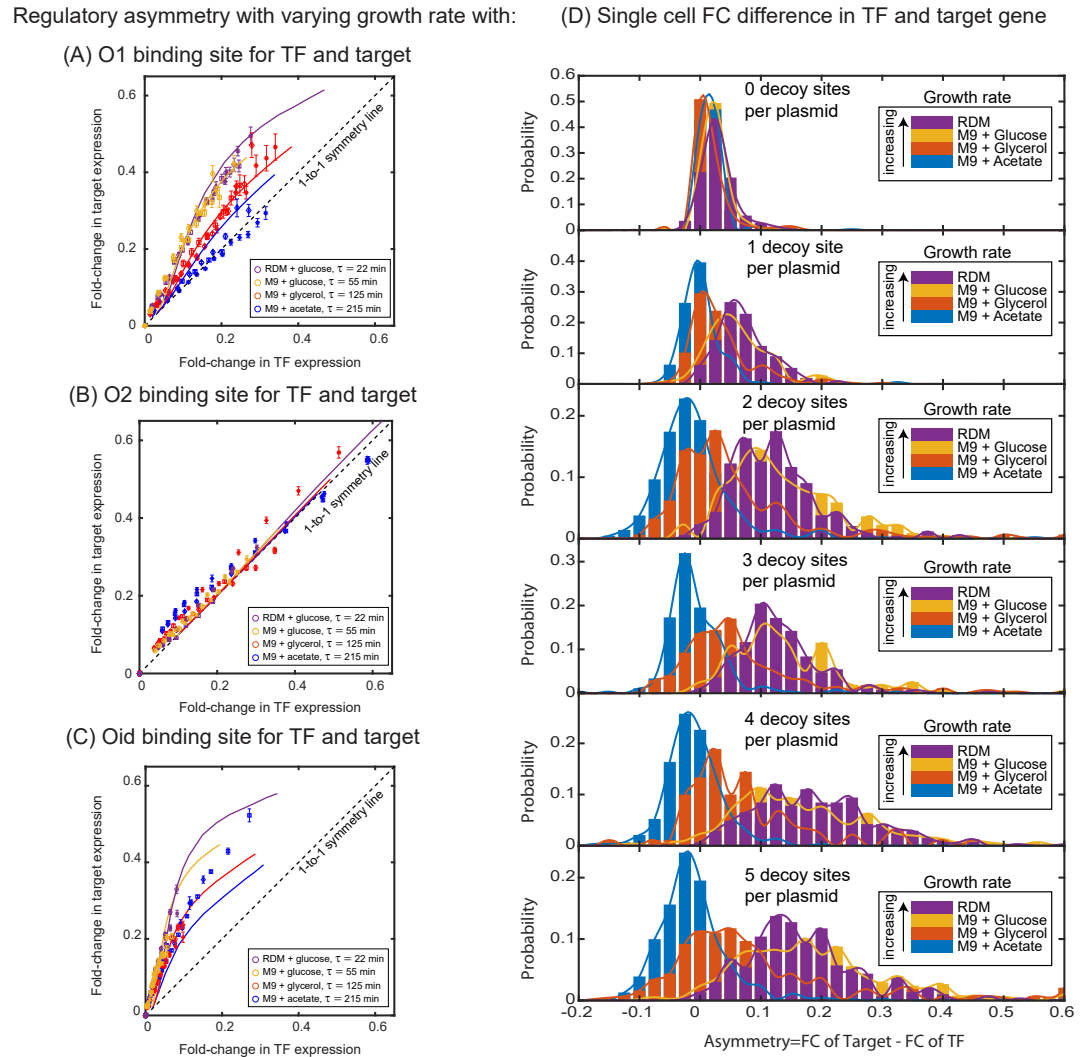
Figure 5. Dependence of regulatory asymmetry on growth rate. Measurement of asymmetry in different media as a function of TF binding energy: O1 (A), O2 (B), Oid (C). The division time (τ) is varied between 22 minutes up to 215 minutes. (A) For O1, the asymmetry decreases with slower division rates and agrees well with the simulation predictions. (B) For the weak O2 site, no asymmetry is seen at any growth rate. (C) For the strongest site Oid asymmetry is present at every growth rate although the magnitude of asymmetry still orders roughly by growth rate. Different symbols represent independent biological replicates and simulation data are shown as solid curves. (D) Histograms of single-cell asymmetry in expression of the TF and target gene regulated by O1 binding site in these 4 growth rates. Solid lines represent the interpolated distributions for better visualization of the histograms. Panels from top to bottom represent increasing the level of competition for the TF.

Table 1. Primers used in this study are listed below. Primers for the chromosomal integration of TF and the target are the same as described in (*Brewster et al.* (*2014*)). Primers to mutate the binding sites from O1 to Oid, O2 or NoO1V1 is listed below with the binding sites highlighted in yellow. Primers to introduce the degradation tags to LacI mCherry fusion protein is listed below with tag sequence highlighted in red.

	Mutagenesis Primer
Oid_mutagenesis_FP	CCGGCTCGTATAATGTGTGG AATTGTGAGCGCTCACAATT GAATTCATTAAAGAG
Oid_mutagenesis_RP	CTCTTTAATGAATTC AATTGTGAGCGCTCACAATT CCACACATT ATACGAGCCGG
O2_mutagenesis_FP	GTGAGCGAGTAACAACC GAATTCATTAAAGAGGAGAAAGGTAC
O2_mutagenesis_RP	TTGTTACTCGCTCACATTT CCACACATTATACGAGCC
NoO1V1_mutagenesis_FP	GATTGTTAGCGGAGAAGAATT GAATTCATTAAAGA GGAGAAAGGTACC
NoO1V1_mutagenesis_RP	AATTCTTCTCCGCTAACAATC CCACACATTATACGAGCCGGAAG
	Primers to introduce tags
ssrA_WT_FP	GC AGCAAACGACGAAAACTACGCTTTAGCAGCT TAAGCTTAA TTAGCTGAGTCTAGAGGC
ssrA_WT_RP	AGCTGCTAAAGCGTAGTTTTCGTCGTTTGCT GCTTTGTA CAGCTCATCCATGC
DAS_FP	C AGCAAACGACGAAAACTACGCTGATGCATCT TAAGCTTAAT TAGCTGAGTCTAGAGGC
DAS_RP	AGATGCATCAGCGTAGTTTTCGTCGTTTGCT GCTTTGTAC AGCTCATCCATGC
DASplus4_FP	GCAGCAAACGACGAAAACTACTCTGAAAATTATGCTGATGCATCT TAAGCTTAATTAGCTGAGTCTAGAGGC
DASplus4_RP	AGATGCATCAGCATAATTTTCAGAGTAGTTTTCGTCGTTTGCT GCTTTGTACAGCTCATCCATGC
	qPCR primers
qPCR_FP	GCATTTATCAGGGTTATTGTCTCAT
qPCR_RP	GGGAAATGTGCGCGGAAC

Table 2. Kinetic rates used in the simulations

Rates	Symbols	Value	Reference
Growth rate	$\ln2/\gamma$	25 min (RDM) 55 min (Glucose) 125 min (Glycerol) 225 min (Acetate)	Measured experimentally
Binding of TF	k_{on}	0.0015 $TF^{-1}s^{-1}$	Obtained from fit
Unbinding of TF	k_{off}	0.00042 s^{-1} (Oid) 0.00149 s^{-1} (O1) 0.0167 s^{-1} (O2)	Eqn. 1
mRNA degradation	γ_m	0.033 s^{-1}	Obtained from fit
mRNA production	β	0.1 s^{-1}	Obtained from fit
Translation rate	α	0.03-0.2 s^{-1}	Obtained from fit
Dimerization	k_p	1.38s^{-1}	*Stamatakis and Zygourakis* (*2011*)
Monomerization	k_m	0.000002s^{-1}	*Stamatakis and Zygourakis* (*2011*)

786 **Appendix 1**

787 ## Sensitivity in choosing the background values

788 The local background of each image is subtracted from individual cells of that image, rather
789 than using a global average over every position. Getting a precise quantitative measurement
790 of fluorescence values is important especially for the tagged strains as their mCherry signal
791 can be only several counts above autofluorescence. The background fluorescence can be
792 influenced by factors such as the local thickness of the agarose pad and positional effects due
793 to the glass dish (which can have small local defects). As shown in **Appendix 1 Figure 1**, a
794 no fluorescent strain corrected using the local fluorescence (calculated by making an inverse
795 mask of each frame, excluding regions with cell, and calculating the mean intensity of the
796 background) of each frame produces a tight, symmetric distribution of cell fluorescence with
797 the mean centered near 0 when compared to using the mean value of no fluorescent strain.
798 In other words, many of the YFP or mCherry signals that appear high in the autofluorescence
799 samples also have higher than average backgrounds and thus accounting for this image to
800 image difference is important. Hence, for all experiments we have used the local background
801 fluorescence of each frame to correct for the autofluorescence of cells in the corresponding
802 frame and excluding frames with too high variation in the background fluorescence.
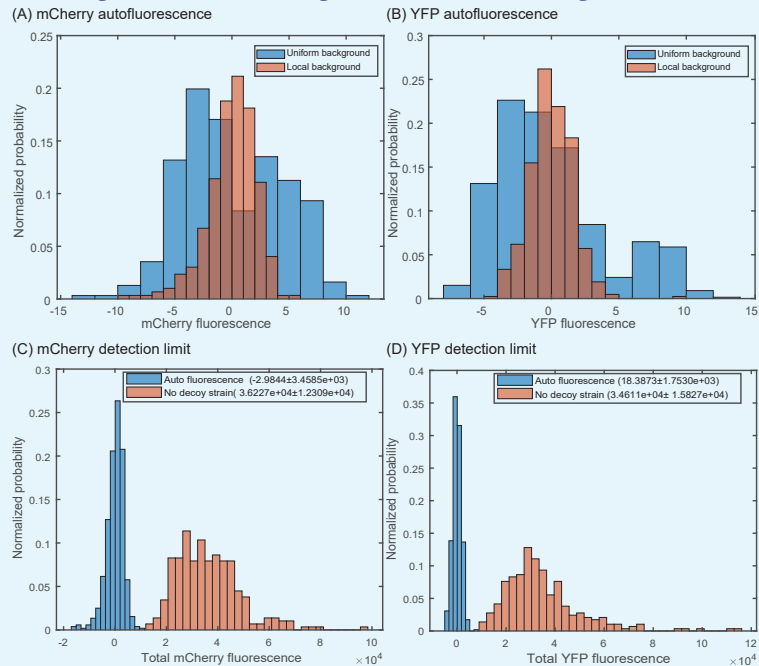
803
804 **Appendix 1 Figure 1. Accounting for local variation in background fluorescence**. Histogram of
805 single-cell autofluorescence levels of (**A**) mCherry or (**B**) YFP fluorescence in a strain without the YFP and
806 mCherry casettes. The blue bars are calculated as the fluorescence level subtracted from the average
807 across the entire sample (9 different fields of view). The red bars are calculated by first removing the
808 local background fluorescence from cells at each position before subtracting the remaining signal from
809 the average. The wide distribution seen in the blue bars is owed largely to local differences in
810 background fluorescence and is removed by accounting for position-to-position variability. (**C,D**)
811 Histogram showing the minimal detection limit (in a no decoy strain) for mCherry (**C**) and YFP (**D**)
813 compared to an autofluorescence strain.

814 **Appendix 2**

Cell growth rate in different media

815
816 Cell growth rate is measured in strain HG105 growing in a 50 mL flask at 37°C and at 250 rpm.
817 Samples are collected at precise time points and OD600 is measured (see **Appendix 2 Figure**
818 **1C**). Doubling time is calculated by first interpolating the intermediate time points from the
819 measurements of OD600 and with the single exponential robust fit function in Matlab (see
820 **Appendix 2 Figure 1A**). **Appendix 2 Figure 1B** shows the scaling in cell area (measured in
821 pixel units) in different media in accordance with the previous literature *Jun et al.* (*2018*).
822 Interestingly, the strain with 5X decoy plasmid has a strikingly different area (from other
823 strains) in glucose minimal media possibly indicating sickness due to the presence of multiple
824 arrays of Oid binding site. Hence, results of 5X decoy strain is excluded from the data set for
825 glucose minimal media.
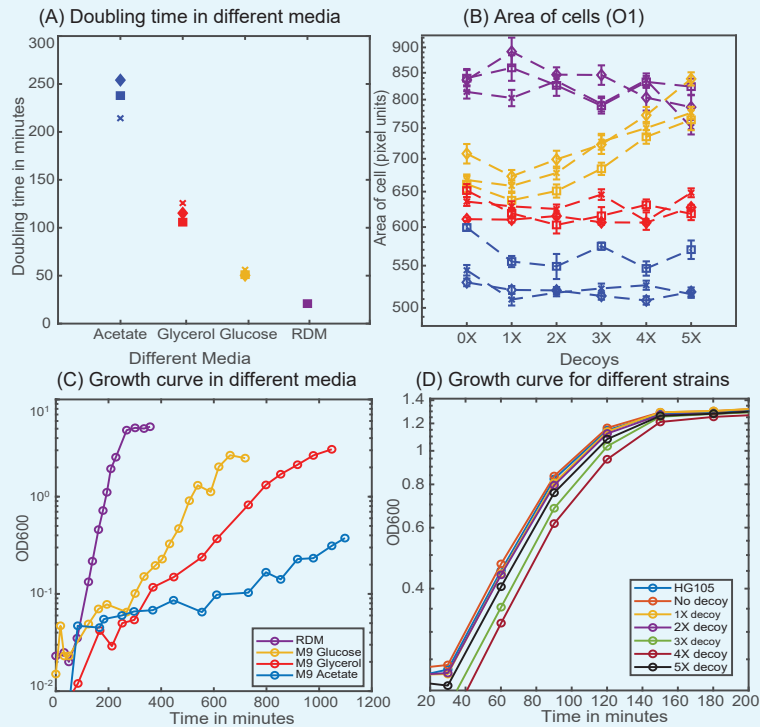
826
827 **Appendix 2 Figure 1. Cellular physiology in different media**. (**A**) Doubling time of HG105 in
828 different media used in this study. (**B**) Consistent with the literature there is a scaling of cell area in
829 different media in accordance with their growth rate. Strains with 4X and 5X decoys growing in glucose
830 minimal media have a drastically different cell area. (**C**) Plot showing the growth curves for the strain
831 HG105 grown in M9-minimal media with glucose, glycerol and acetate or in rich-defined media. (**D**) Plot
832 showing the growth curves in rich-defined media for strains carrying in different decoy plasmid. Cells
833 are grown in TECAN machine (maintained at 37°C) in a 96-well plate with constant shaking and
835 measurements are made every 30 minutes.

Appendix 3

Quantification of plasmid copy number

Five different variants of Oid decoy arrays (carrying 1, 2, 3, 4 and 5 binding sites for Oid, respectively) are inserted in the intergenic region between the origin of replication and ampicillin cassette of the pZE plasmid. Plasmid copy number is quantified in qPCR measurements using primers that targets a 90 bp-intergenic region in the plasmid backbone immediately upstream of the site of insertion of our decoy array. The total number of decoys can then be estimated by multiplying the measured copy number of pZE plasmid backbone with the number of binding sites in the decoy array. As shown in **Appendix 3 Figure 1A**, pZE plasmid backbone had similar copy number in strains with different decoy arrays except for strains carrying the 5X decoy array plasmid. Copy number of 5X-decoy array plasmid is significantly higher when compared to strains carrying other decoy array plasmids. This difference is primarily due to a reduced CFU/mL obtained (see **Appendix 3 Figure 1C**) for strains carrying the 5X decoy arrays; the number of molecules of plasmid per reaction is uniform across different strains (see **Appendix 3 Figure 1B**). It is not clear if this is due to this sample actually containing less cells or if it is due to a reduced ability to recover and separate these cells (which tend to clump and stick more in microscopy imaging) in the plating assay. This may lead to over-prediction of the copy number of 5X decoy plasmid. Hence, we excluded the 5X-decoy plasmid data in Figure 2A. The average ($\pm$ standard deviation) number of decoy binding arrays in different strains are: 39$\pm$8, 96$\pm$17, 134$\pm$25, 245$\pm$40, and 607$\pm$47, respectively.
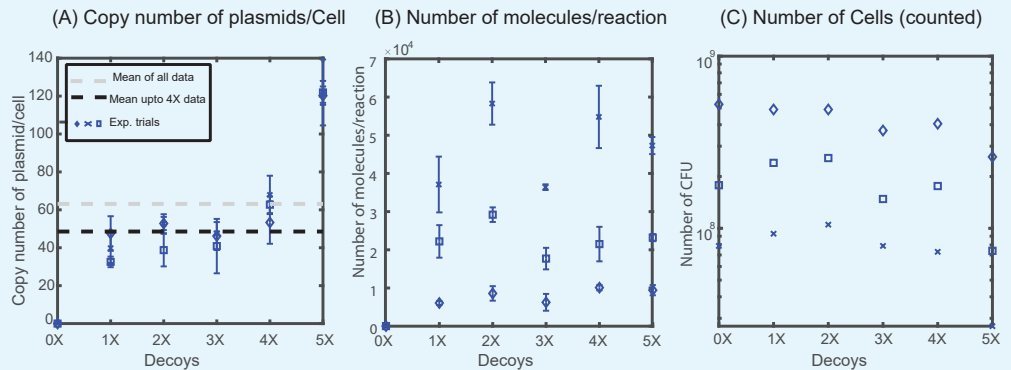
Appendix 3 Figure 1. Quantification of plasmid copy number. (**A**) Copy number of decoy array plasmids measured in M9-Glucose minimal media. (**B**) Number of molecules obtained per qPCR reaction remains constant across different decoy strains (1X, 2X, 3X, 4X, 5X). (**C**) Number of Colony Forming Units (CFU) per mL used to normalize the number of molecules to account for the copy number of plasmids per cell.

Appendix 4

Constitutive values for the autoregulatory gene

To compare expression levels between the TF and the target genes, we wish to compare fold-change as an "apples-to-apples" comparison of the regulation of each gene. To calculate fold-change we must know the constitutive expression of the gene, *i.e.* how much expression is seen in the absence of regulation by TF. In simulation, this is simple to calculate because we can remove any reactions that include TF binding. Experimentally, calculating constitutive expression for the target gene is also relatively straight-forward; we delete the gene express-ing LacI-mCherry and measure the same construct in the absence of TF. However, measuring constitutive expression experimentally for an autoregulating gene was more challenging. There are many possible strategies, but all of them come with some complication. In short, we attempted 3 different strategies which included: 1) IPTG induction (with or without the addition of decoys), 2) mutated LacI to ablate specific binding, 3) mutated binding site se-quences (which has the complication that the site is centered at +11 and thus is both close to the promoter and present on the transcript, see **Appendix 4 Figure 1A**). In the end, we identified one mutated site (NoO1V1) which faithfully preserved constitutive expression of the target gene in all media studied. Unfortunately, we were not able to find corresponding mutated sites that reproduced expression of promoters bearing O2 or Oid binding sites. As such, for data using those binding sites on the TF gene we have an unknown scaling factor between the x- and y-axis in the fold-change versus fold-change plots which we determine by fitting the glucose data to our simulations (and then hold constant for all other data sets). In the following sections we discuss techniques we tried.

Allosteric induction with IPTG to achieve constitutive expression

One way to obtain the constitutive values is to exploit the property of the LacI to become less active when bound to small molecules like IPTG. Previous studies indicate that even with the use of IPTG, expression from a stronger binding site (like Oid) cannot be fully rescued when the repressor copy number is high ***Razo-Mejia et al. (2018)***. In our experiments, we observed this phenomenon as well. As shown in **Appendix 4 Figure 1C-E**, for most strains expressing the TF, the expression of the target could not be fully rescued with 2.5mM IPTG and decoys. Further increase in IPTG concentration (to up to 10mM) did not help in increasing the target expression. Hence, allosteric induction with IPTG could not serve as a right constitutive value for our system.

### *Use of LacI with mutated DNA recognition domains*

We constructed a mutant protein by deleting 10 amino acids (from amino acid 60 to amino acid 70) in the DNA binding domain of LacI. This mutant helped to completely restore the target expression. However, the mCherry level of strains with the mutated LacI-mCherry were significantly lower than the mCherry level of strains with the functional LacI-mCherry. Since, we would expect the expression of the non-functional TF to be higher than the functional, we reason that this did not provide an accurate estimate of the constitutive mCherry level in the LacI-mCherry strain. This discrepancy may originate from many possible sources such as a change to the stability of the mRNA/protein or a possible alteration to the spectral property of mCherry (which is directly fused to LacI). In the end we were unable to find a suitable LacI mutant without this feature.

Use of binding sequence insensitive to LacI

Oehler *et al.* 1994 has reported inactivated O1 site (NoO1V1) that has close consensus to O1 binding sequence but does not allow LacI binding. We verified that the expression of YFP from the promoter with NoO1V1 is comparable to the expression of YFP from O1 regulated promoter (in the absence of any LacI) but is lower than the expression from O2 and Oid regulated promoters (**Appendix 4 Figure 1B**). Although expression alone does not guarantee that all intermediate steps are precisely the same, we believe this construct gives accurate measurements of constitutive expression for the TF and target genes. We used TF and target with NoO1V1 binding sequence as our constitutive strain to normalize expression from any O1 regulated genes in our experiments. We also tried other forms of mutations on the NoO1V1 binding site (**Appendix 4 Figure 1A**) in order to obtain mutants that relieves *lacI* repression and restore expression of Oid or O2 sequence but with no success.
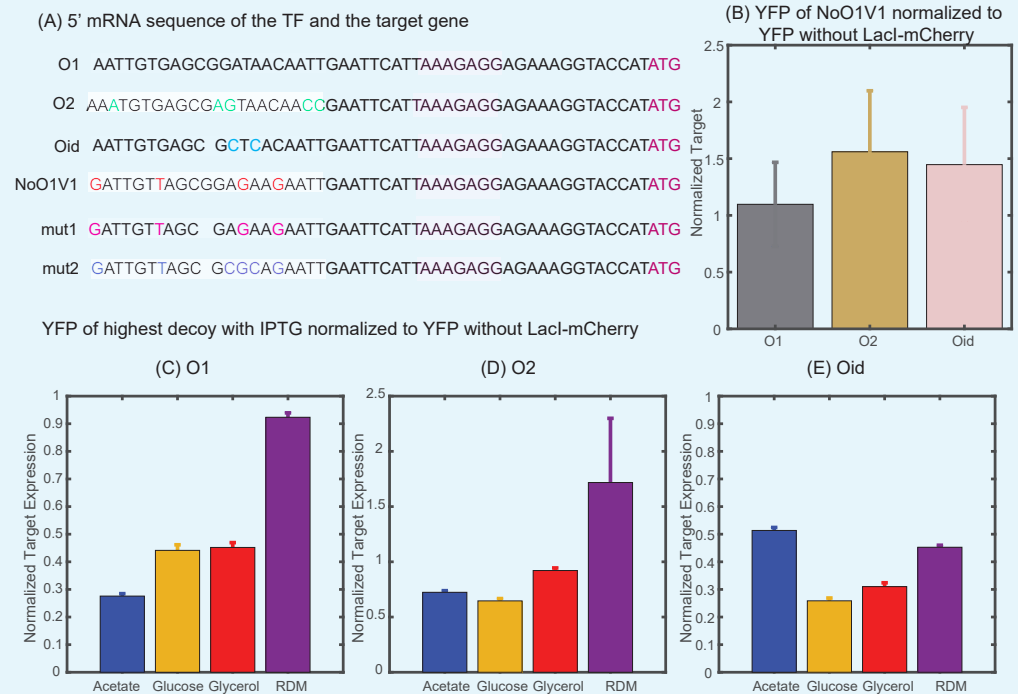
Appendix 4 Figure 1. Determining constitutive expression of YFP and mCherry. (**A**) 5′ mRNA sequence of the TF and the target genes. The binding site for the TF is carried in the mRNA sequence and is highlighted in shaded dark grey boxes with base changes for different binding sites coded in multicolor. mut1 and mut2 are the two variant binding sites that are designed with mutations similar to NoO1V1 but with Oid site length. However, such changes do not achieve constitutive unregulated expression similar to O2 or Oid. (**B**) Plot showing YFP expressed from NoO1V1 regulated promoter normalized to YFP expressed from promoter regulated with O1, O2 or Oid. (**C-E**) Plot showing the effect of 2.5mM IPTG in relieving YFP expression form O1 (**C**), O2 (**D**) or Oid. (**E**) regulated promoter and with 5X decoy plasmids. As indicated in the plot IPTG is not sufficient to restore complete expression of YFP in different media and hence cannot be used as a measure of constitutive expression.
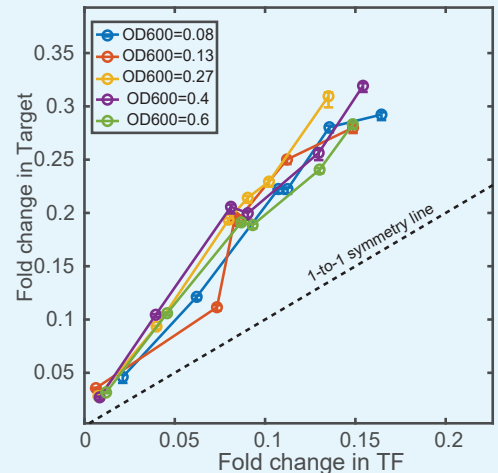
931 **Appendix 5**

932 ### Copy number difference and Diffusion limitation of TF

933 Copy number variation of genes along the long axis of the chromosome and the diffusion
934 limitation of LacI-mCherry could be suggested as a significant contributor to the asymmetry
935 between TF and the target. *E. coli* can initiate multiple replication events (depending on
936 the division rate in the given media) and hence different genes along the chromosome will
937 experience a different copy number in a given time. For instance, *E. coli* growing in RDM
938 (with a division rate of 22 minutes) will have a copy number of 4 at the *ybcN* locus (where the
939 TF gene is integrated) and a copy number of 3.6 at the *galK* locus (where the target gene is
940 integrated) as described by Cooper *et al.* ***Cooper and Helmstetter* (*1968*)**. We believe that
941 the use of fold-change as the measurement of expression helps to reduce the influence of
942 copy number effects (since both the regulated and unregulated measurements have the
943 same copy number). However, the effects may not be linear and LacI has been shown to
944 suffer from diffusion limitation from its origin of synthesis ***Kuhlman and Cox* (*2012*)**. Hence,
945 we tested our system by placing the TF and the target genes integrated next to each other
946 at the *gspI* locus. As evident from **Appendix 5 Figure 1**, there is no significant contribution
947 of the copy number difference between TF and target or diffusion limitation of TF on the
948 phenomenon of asymmetry observed in our negatively-autoregulated SIM motif.
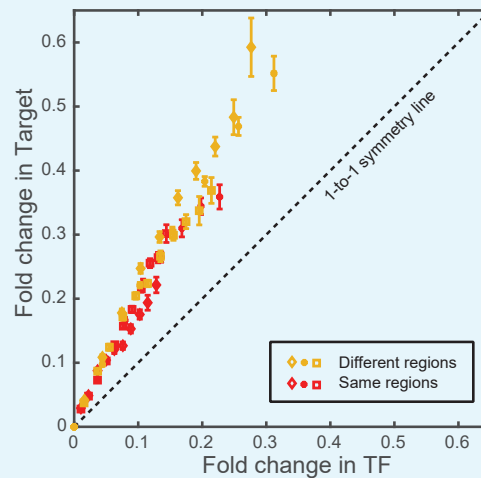
949
950 **Appendix 5 Figure 1. Effect of copy number difference on asymmetry.** Comparison of asymmetry
951 in strain where the TF and the target genes are located either at two different regions of the
952 chromosome (*ybcN* for TF and *galK* for target, shown in yellow data points)) or when it is present
953 together in the chromosome (at the *gspI* locus, shown in red data points). **(B)** Plot showing the
954 measurement of asymmetry in glucose-minimal media at different optical density (OD600) in the
956 exponential phase.

957 **Appendix 6**

958 ### Deterministic solution
959 Using the assumptions of equilibrium mass-action kinetics, the deterministic counterpart of
960 the negative autoregulation system described in the main text and **Appendix 6 Figure 2A**
961 can be written as

$$\frac{dX}{dt} = \alpha m_x - \gamma X + 2k_m R - 2k_p X^2,$$

$$\frac{dR}{dt} = -k_m R + k_p X^2 - \gamma R - k_{on} R P_{fx} - k_{on} R P_{fy} - k_{on} R N_f + k_{off,x}(1 - P_{fx})$$

$$+ k_{off,y}(1 - P_{fy}) + k_{off,d}(N - N_f),$$

962
$$\frac{dY}{dt} = \alpha m_y - \gamma Y,$$

963
$$\frac{dP_{fx}}{dt} = -k_{on} R P_{fx} + (k_{off,x} + \gamma)(1 - P_{fx}),$$

964 $$\tag{A6-1}$$

965
$$\frac{dP_{fy}}{dt} = -k_{on} R P_{fy} + (k_{off,y} + \gamma)(1 - P_{fy}),$$

$$\frac{dN_f}{dt} = -k_{on} R N_f + (k_{off,d} + \gamma)(N - N_f),$$

$$\frac{dm_x}{dt} = \beta P_{fx} - \gamma_m m_x,$$

$$\frac{dm_y}{dt} = \beta P_{fy} - \gamma_m m_y.$$

966   Here, X is the concentration of free TF monomer, Y is the concentration of target protein,
967 and R is the concentration of TF dimer. m_x, m_y, $P_{fx}(P_{ox})$, $P_{fy}(P_{oy})$, N, and $N_f(N_o)$ are TF mRNA,
968 target mRNA, free (bound) TF-promoter, free (bound) target-promoter, total concentration
969 of decoy sites, and concentration of free (bound) decoy sites, respectively. Inherent in the
970 equations are the assumptions of the conservation for the concentration of binding sites ,
971 i.e. $P_{fx} + P_{ox} = 1$, $P_{fy} + P_{oy} = 1$, and $N_f + N_o = N$. The right hand side of the equations can be
972 set to zero to obtain the steady state values for all the components.

$$P_{fx} = \frac{k_{off,x} + \gamma}{k_{on} R + k_{off,x} + \gamma} = \frac{1}{1 + \sigma_1 R},$$

$$P_{fy} = \frac{k_{off,y} + \gamma}{k_{on} R + k_{off,y} + \gamma} = \frac{1}{1 + \sigma_2 R},$$

973
$$N_f = \frac{N(k_{off,d} + \gamma)}{k_{on} R + k_{off,d} + \gamma} = \frac{N}{1 + \sigma_3 R},$$

974
975 $$m_x = \frac{\beta}{\gamma_m} P_{fx}, \tag{A6-2}$$

976
$$m_y = \frac{\beta}{\gamma_m} P_{fy},$$

$$0 = \alpha m_x - \gamma X + 2k_m R - 2k_p X^2,$$

$$0 = -k_m R + k_p X^2 - \gamma(R + P_{ox} + P_{oy} + N_o),$$

$$Y = \frac{\alpha \beta}{\gamma \gamma_m} P_{fy} = \frac{\alpha \beta}{\gamma \gamma_m} \frac{1}{1 + \sigma_2 R},$$

977   where $\sigma_i = k_{on}/(k_{off,i} + \gamma)$. The concentration of total TF protein can be expressed as a
978 sum of free TF monomer, TF dimer bound to each promoter, and TF dimers bound to the
979 decoys sites

$$
\begin{aligned}
X_{\text{Total}} &= X + 2(R + P_{\text{ox}} + P_{\text{oy}} + N_{\text{o}}), \\
&= \frac{\alpha}{\gamma} m_{\text{x}}, \\
&= \frac{\alpha\beta}{\gamma\gamma_{\text{m}}} \frac{1}{1 + \sigma_1 R}.
\end{aligned}
\tag{A6-3}
$$

The fold-change of the TF and target expression, thus can be obtained by dividing X_{Total} and Y with the constitutive expression, i.e., $\alpha\beta/\gamma\gamma_{\text{m}}$ which yields,

$$
\text{FC}_{\text{TF}} = \frac{1}{1 + \sigma_1 R} = \frac{1}{1 + \frac{k_{\text{on}}}{k_{\text{off,x}} + \gamma} R},
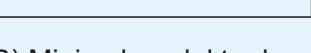\tag{A6-4}
$$
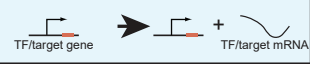
$$
\text{FC}_{\text{Target}} = \frac{1}{1 + \sigma_2 R} = \frac{1}{1 + \frac{k_{\text{on}}}{k_{\text{off,y}} + \gamma} R}.
\tag{A6-5}
$$

It is worth noting that both TF and target protein follows $1/(1 + R^*)$, where $R^* = R k_{\text{on}}/(k_{\text{off}} + \gamma)$ is the reduced free TF concentration, which is equivalent to the thermodynamic solution *Weinert et al.* (*2014*). When the unbinding rates of TF and target are identical, each of them follow the same fold-change curve irrespective of the competition from other decoy sites. In **Appendix 6 Figure 1**C, we plot the fold-change for TF and target with $k_{\text{off,x}}$ corresponding to O1 binding site and $k_{\text{off,y}}$ corresponding to O1 (yellow), O2 (purple), and Oid (blue). It can be seen from the figure that when the off-rates are identical the fold-change curve follows one-to-one line showing no asymmetry which is in contrast with the results obtained using stochastic simulations and experimental results. Furthermore, both the transient and steady state behavior of mean fold-change of TF and target obtained from deterministic solution deviate from the stochastic behavior (see **Appendix 6 Figure 1**B). Importantly, when autoregulation is removed from the simulation, the deterministic and stochastic solutions agree precisely (**Appendix 6 Figure 1**B inset).
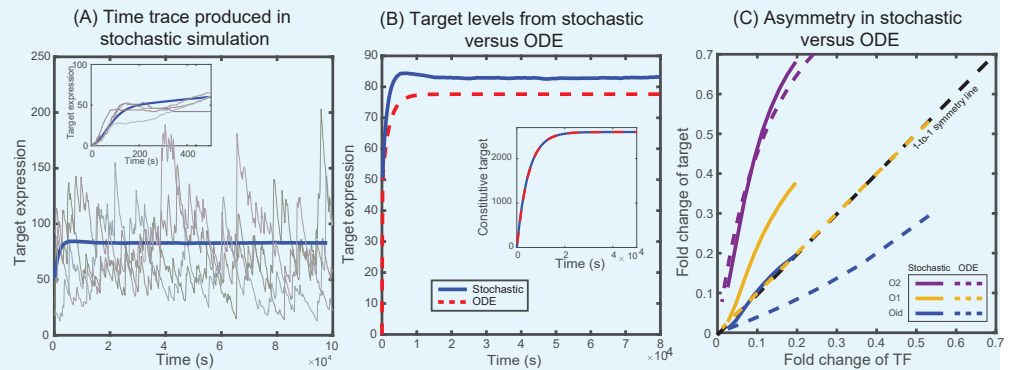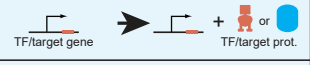

Appendix 6 Figure 1. Solutions from stochastic simulation and from deterministic ODEs. (**A**) Representative time traces of target expression in individual cells (grey shades) from stochastic simulations. Blue solid line represents the mean behavior averaged over 5×10^4 iterations. Inset shows the transient behavior. (**B**) Plot showing the average target expression in the negative SIM motif from stochastic simulations (solid line) and from solving deterministic ODEs (dashed line). Inset shows that when regulation is removed the average levels are identical for stochastic and deterministic models. (**C**) Plot showing the asymmetry between TF and target expression from using either stochastic simulation (solid lines) or solving deterministic ODEs (dashed lines). The TF is always regulated by O1 binding site whereas the target is regulated by O1 (yellow), O2 (purple) or Oid (blue) binding sites. The black dashed line represents line of no asymmetry.

(A) Full model

Schematic reaction	Stochastic chemical reaction	Deterministic rate
TF/target gene → TF/target mRNA	$m_{\text{TF(Target)}} \xrightarrow{\beta} m_{\text{TF(Target)}} + 1$	$\beta P_{\text{fx(fy)}}$
TF mRNA → TF monomer	$m_{\text{TF}} \xrightarrow{\alpha} m_{\text{TF}} + X$	αm_x
target mRNA → target protein	$m_{\text{Target}} \xrightarrow{\alpha} m_{\text{Target}} + Y$	αm_y
+ → TF dimer	$X + X \xrightarrow{k_p} Z$	$k_p X^2$
→ +	$Z \xrightarrow{k_m} X + X$	$k_m Z$
+ → repressed gene	$Z + P_{\text{TF(Target)}} \xrightarrow{k_{\text{on}}} ZP_{\text{TF(Target)}}$	$k_{\text{on}} ZP_{\text{fx(fy)}}$
→ +	$ZP_{\text{TF(Target)}} \xrightarrow{k_{\text{off,TF(Target)}}} Z + P_{\text{TF(Target)}}$	$k_{\text{off,x(off, y)}}(1 - P_{\text{fx(fy)}})$
decoy site + → bound decoy	$Z + N \xrightarrow{k_{\text{on}}} ZN$	$k_{\text{on}} ZN_f$
→ +	$ZN \xrightarrow{k_{\text{off,Decoy}}} Z + N$	$k_{\text{off,d}}(N - N_f)$
→ decayed mRNA	$m_{\text{TF(Target)}} \xrightarrow{\gamma_m} m_{\text{TF(Target)}}$	$\gamma_m m_{x(y)}$
or or → decayed protein	$X, Y, Z \xrightarrow{\gamma} \phi$	γX or γY or γZ
or → or	$ZP_{\text{TF(Target)}} \xrightarrow{\gamma} P_{\text{TF(Target)}}$ or $ZN \xrightarrow{\gamma} N$	$\gamma(1 - P_{\text{fx(fy)}})$ or $\gamma(N - N_f)$

(B) Minimal model to demonstrate asymmetry

| Schematic reaction	Stochastic chemical reaction	Deterministic rate
TF/target gene → TF/target prot.	$P_{\text{TF(Target)}} \xrightarrow{\alpha} P_{\text{TF(Target)}} + X(Y)$	$\alpha P_{\text{fx(fy)}}$
+ → repressed gene	$X + P_{\text{TF(Target)}} \xrightarrow{k_{\text{on}}} XP_{\text{TF(Target)}}$	$k_{\text{on}} XP_{\text{fx(fy)}}$
→ +	$XP_{\text{TF(Target)}} \xrightarrow{k_{\text{off}}} X + P_{\text{TF(Target)}}$	$k_{\text{off}}(1 - P_{\text{fx(fy)}})$
or → decayed protein	$X, Y \xrightarrow{\gamma} \phi$	γX or γY
→	$XP_{\text{TF(Target)}} \xrightarrow{\gamma} P_{\text{TF(Target)}}$	$\gamma(1 - P_{\text{fx(fy)}})$

Appendix 6 Figure 2. List of reactions used in the (A) stochastic model and (B) in the minimal model.

Appendix 7

Maximum asymmetry

The asymmetry in regulation (defined as $\mathrm{FC}_{TF} - \mathrm{FC}_{Target}$) is a function of all the rates describing the system and number of decoy binding sites. For a given set of rates (k_{on}, k_{off}, γ, γ_m) as the decoy number is varied the asymmetry first increases, attains a maximum and then approaches zero for infinite number of decoy binding sites (see **Appendix 7 Figure 1**). The maximum asymmetry for a given set of rates is this peak asymmetry observed as decoy number is varied. In the manuscript we show a heatmap (Fig. **3E**) to emphasize how this maximum asymmetry depends on the two crucial rate parameters, off-rate of the binding sites (k_{off} or equivalently binding affinity, since in our model k_{on} is kept constant) and the degradation of TF molecules (γ).
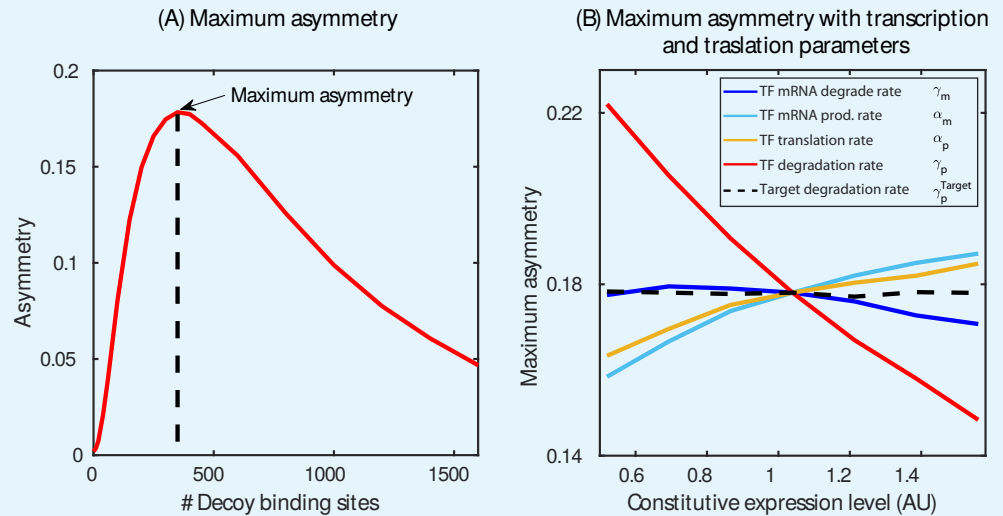
Appendix 7 Figure 1. Determination of maximum asymmetry. (A) Maximum asymmetry in simulation is computed by plotting the asymmetry, difference in fold-change between target and TF, versus number of decoy binding sites in SIM motif. The peak of this asymmetry corresponds to the maximum asymmetry. (B) Exploring the model parameters of the TF (mRNA production and degradation; protein production and degradation) that could influence the asymmetry between the TF and the target. Tuning the protein degradation rate (red line) has the maximum influence on the asymmetry between the TF and its target gene.

Appendix 8

A minimal model of an autoregulatory gene and a single target gene

The full model in **Appendix 6** contains many reactions that are included to more faithfully mirror the biological system we are modeling. However, not all of these reactions are necessary to observe the phenomenon of asymmetry which we describe in this manuscript. In this section, we present a reduced model of the extended model of transcription described in Materials and methods to show that the asymmetry in TF and target expression stems from the network architecture and not due to the intermediate steps of transcription and the presence of excess decoy binding sites. We consider an autoregulatory gene whose protein product X inhibits its own expression and also represses a single target gene with protein product Y. To reduce the complexity, the protein is made directly from the gene with no intermediates (eliminating translation rates and mRNA decay rates). In this system the TF, X, acts as a monomer and binds to its own gene with rate k_{on} and unbinds with rate k_{off}. Similarly, the TF (X) binds and unbinds from the target gene with the same rates. Both the TF gene and target gene in free state (not bound with TF) produces their protein with rate α which degrades with rate γ (dilution through cell division). The reactions describing this reduced model are listed in **Appendix 6 Figure 2B**. We implement the simulations using stochastic simulation algorithms as described in Materials and Methods section.

Next, we write a set of deterministic coupled ODEs corresponding to the reactions described above which is given by

$$
\begin{aligned}
\frac{dX}{dt} &= \alpha P_{fx} - \gamma X - k_{on} X P_{fx} - k_{on} X P_{fy} + k_{off}(1 - P_{fx}) + k_{off}(1 - P_{fy}), \\
\frac{dY}{dt} &= \alpha P_{fy} - \gamma Y, \\
\frac{dP_{fx}}{dt} &= -k_{on} X P_{fx} + (k_{off} + \gamma)(1 - P_{fx}), \\
\frac{dP_{fy}}{dt} &= -k_{on} X P_{fy} + (k_{off} + \gamma)(1 - P_{fy}).
\end{aligned}
\tag{A8-1}
$$

Here, X is the concentration of free TF and Y is the concentration of target protein. $P_{fx}(P_{ox})$ and $P_{fy}(P_{oy})$ are free (bound) TF-promoter, free (bound) target-promoter, respectively. Inherent in the equations are the assumptions of the conservation for the concentration of binding sites, i.e. $P_{fx} + P_{ox} = 1$, $P_{fy} + P_{oy} = 1$. To obtain the steady state values of TF and target expression the right hand side of the equations is set to zero which yield

$$
\begin{aligned}
P_{fx} &= \frac{k_{off} + \gamma}{k_{on} X + k_{off} + \gamma} = \frac{1}{1 + \sigma X}, \\
P_{fy} &= \frac{k_{off} + \gamma}{k_{on} X + k_{off} + \gamma} = \frac{1}{1 + \sigma X}, \\
X &= \frac{\alpha}{\gamma} P_{fx} - P_{ox} - P_{oy}, \\
Y &= \frac{\alpha}{\gamma} P_{fy} = \frac{\alpha}{\gamma} \frac{1}{1 + \sigma X},
\end{aligned}
\tag{A8-2}
$$

where $\sigma = k_{on}/(k_{off} + \gamma)$. Total TF concentration, X_{Total}, can be expressed as the sum of

free TF and TFs bound to each promoter

$$X_{\text{Total}} = X + P_{\text{ox}} + P_{\text{oy}}$$
$$= \frac{\alpha}{\gamma} P_{\text{fx}}$$
$$= \frac{\alpha}{\gamma} \frac{1}{1 + \sigma X}.$$

(A8-3)

The fold-change of the TF and target expression, thus can be obtained by dividing X_{Total} and Y by the constitutive expression, *i.e.* without any regulation, $C_0 = \alpha/\gamma$ which yields,

$$\text{FC}_{\text{TF}} \quad = \frac{1}{1 + \sigma X} = \frac{1}{1 + \frac{k_{\text{on}}}{k_{\text{off}} + \gamma} X},$$

(A8-4)

$$\text{FC}_{\text{Target}} \quad = \frac{1}{1 + \sigma X} = \frac{1}{1 + \frac{k_{\text{on}}}{k_{\text{off}} + \gamma} X}.$$

(A8-5)

As was shown previously in section **Appendix 6**, both TF and target protein follows $1/(1 + \sigma X)$ and show no asymmetry in regulation.

Furthermore, solving Eqn. A8-2 we get the free TF expression as,

$$X = \frac{-1 - 2\sigma + \sqrt{(1 + 2\sigma)^2 + 4C_0\sigma}}{2\sigma}.$$

(A8-6)

1091 **Appendix 9**

1092 ## Chemical master equation (CME) for the minimal model

1093 The chemical master equation governing the dynamics of the expression for TF and target
1094 gene for the minimal model discussed in **Appendix 8** (also shown in **Appendix 6 Figure 2B**)
1095 is given by

$$\frac{dP_{00}(n,m,t)}{dt} = \alpha\Big[P_{00}(n-1,m,t) - P_{00}(n,m,t) + P_{00}(n,m-1,t) - P_{00}(n,m,t)\Big]$$
$$+ \gamma\Big[(n+1)P_{00}(n+1,m,t) - nP_{00}(n,m,t) + (m+1)P_{00}(n,m+1,t)$$
$$- mP_{00}(n,m,t) + P_{01}(n,m,t) + P_{10}(n,m,t)\Big] + k_{\text{off}}\Big[P_{01}(n-1,m,t)$$
$$+ P_{10}(n-1,m,t)\Big] - 2k_{\text{on}}nP_{00}(n,m,t),$$

$$\frac{dP_{01}(n,m,t)}{dt} = \alpha\Big[P_{01}(n-1,m,t) - P_{01}(n,m,t)\Big] + \gamma\Big[(n+1)P_{01}(n+1,m,t)$$
$$- nP_{01}(n,m,t) + (m+1)P_{01}(n,m+1,t) - mP_{01}(n,m,t)$$
1096
1097 $$+ P_{11}(n,m,t) - P_{01}(n,m,t)\Big] + k_{\text{off}}\Big[P_{11}(n-1,m,t) - P_{01}(n,m,t)\Big]$$
1098 $$+ k_{\text{on}}\Big[(n+1)P_{00}(n+1,m,t) - nP_{01}(n,m,t)\Big], \tag{A9-1}$$
1099
$$\frac{dP_{10}(n,m,t)}{dt} = \alpha\Big[P_{10}(n,m-1,t) - P_{10}(n,m,t)\Big] + \gamma\Big[(n+1)P_{10}(n+1,m,t)$$
$$- nP_{10}(n,m,t) + (m+1)P_{10}(n,m+1,t) - mP_{10}(n,m,t)$$
$$+ P_{11}(n,m,t) - P_{10}(n,m,t)\Big] + k_{\text{off}}\Big[P_{11}(n-1,m,t) - P_{10}(n,m,t)\Big]$$
$$+ k_{\text{on}}\Big[(n+1)P_{00}(n+1,m,t) - nP_{10}(n,m,t)\Big],$$

$$\frac{dP_{11}(n,m,t)}{dt} = \gamma\Big[(n+1)P_{11}(n+1,m,t) - nP_{11}(n,m,t) + (m+1)P_{11}(n,m+1,t)$$
$$- mP_{11}(n,m,t) - 2P_{11}(n,m,t)\Big] - 2k_{\text{off}}P_{11}(n,m,t)+$$
$$k_{\text{on}}\Big[(n+1)P_{01}(n+1,m,t) + (n+1)P_{10}(n+1,m,t)\Big]$$

1100 Here $P_{ij}(n,m,t)$ is the probability of having n TF protein and m target protein at any instant
1101 of time t in the state (i,j). i and j denotes the occupancy of the TF promoter and target
1102 promoter, respectively. A value of 0 indicates that the promoter of TF/target gene is occupied
1103 by a TF. A value of 1, similarly indicates a promoter which is free to express.
1104 Summing Eqn. A9-1 over all values of (m,n) we get the rate equation for occupancy
1105 defined as $S_{ij} = \sum_{n,m=0}^{\infty} P_{ij}$ in each state

1106 $$\frac{dS_{00}}{dt} = (\gamma + k_{\text{off}})(S_{01} + S_{10}) - 2k_{\text{on}}\langle n\rangle_{00},$$
1107 $$\frac{dS_{01}}{dt} = (\gamma + k_{\text{off}})(S_{11} - S_{01}) + k_{\text{on}}\Big[\langle n\rangle_{00} - \langle n\rangle_{01}\Big],$$
1108 $$\tag{A9-2}$$
1109 $$\frac{dS_{10}}{dt} = (\gamma + k_{\text{off}})(S_{11} - S_{10}) + k_{\text{on}}\Big[\langle n\rangle_{00} - \langle n\rangle_{10}\Big],$$
$$\frac{dS_{11}}{dt} = -2(\gamma + k_{\text{off}})S_{11} + k_{\text{on}}\Big[\langle n\rangle_{01} + \langle n\rangle_{10}\Big].$$

Multiplying both sides of Eqn. A9-1 by n and summing over all values of (m, n) we get the time evolution of free TF protein in each state ($\langle n \rangle_{ij} = \sum_{m,n} n P_{i,j}(n, m, t)$)

$$\frac{d\langle n \rangle_{00}}{dt} = \alpha S_{00} - \gamma \langle n \rangle_{00} + \gamma \left[\langle n \rangle_{01} + \langle n \rangle_{10} \right] + k_{\text{off}} \left[\langle n+1 \rangle_{01} + \langle n+1 \rangle_{10} \right] - 2k_{\text{on}} \langle n^2 \rangle_{00}$$

$$\frac{d\langle n \rangle_{01}}{dt} = \alpha S_{01} - \gamma \langle n \rangle_{01} + \gamma \left[\langle n \rangle_{11} - \langle n \rangle_{01} \right] + k_{\text{off}} \left[\langle n+1 \rangle_{11} - \langle n \rangle_{01} \right] + k_{\text{on}} \left[\langle n(n-1) \rangle_{00} - \langle n^2 \rangle_{01} \right]$$

$$\frac{d\langle n \rangle_{10}}{dt} = -\gamma \langle n \rangle_{10} + \gamma \left[\langle n \rangle_{11} - \langle n \rangle_{10} \right] + k_{\text{off}} \left[\langle n+1 \rangle_{11} - \langle n \rangle_{10} \right] + k_{\text{on}} \left[\langle n(n-1) \rangle_{00} - \langle n^2 \rangle_{10} \right]$$

$$\frac{d\langle n \rangle_{11}}{dt} = -\gamma \langle n \rangle_{11} - 2\gamma \langle n \rangle_{11} - 2k_{\text{off}} \langle n \rangle_{11} + k_{\text{on}} \left[\langle n(n-1) \rangle_{01} + \langle n(n-1) \rangle_{10} \right].$$

(A9-3)

Similarly, multiplying both sides of Eqn. A9-1 by m and summing over all values of (m, n) we obtain the time evolution of target protein in each state ($\langle m \rangle_{ij} = \sum_{m,n} m P_{i,j}(n, m, t)$)

$$\frac{d\langle m \rangle_{00}}{dt} = \alpha S_{00} - \gamma \langle m \rangle_{00} + \gamma \left[\langle m \rangle_{01} + \langle m \rangle_{10} \right] + k_{\text{off}} \left[\langle m \rangle_{01} + \langle m \rangle_{10} \right] - 2k_{\text{on}} \langle mn \rangle_{00}$$

$$\frac{d\langle m \rangle_{01}}{dt} = -\gamma \langle m \rangle_{01} + \gamma \left[\langle m \rangle_{11} - \langle m \rangle_{01} \right] + k_{\text{off}} \left[\langle m \rangle_{11} - \langle m \rangle_{01} \right] + k_{\text{on}} \left[\langle mn \rangle_{00} - \langle mn \rangle_{01} \right]$$

$$\frac{d\langle m \rangle_{10}}{dt} = \alpha S_{10} - \gamma \langle m \rangle_{01} + \gamma \left[\langle m \rangle_{11} - \langle m \rangle_{10} \right] + k_{\text{off}} \left[\langle m \rangle_{11} - \langle m \rangle_{10} \right] + k_{\text{on}} \left[\langle mn \rangle_{00} - \langle mn \rangle_{01} \right]$$

$$\frac{d\langle m \rangle_{11}}{dt} = -\gamma \langle m \rangle_{11} - 2\gamma \langle m \rangle_{11} - 2k_{\text{off}} \langle m \rangle_{11} + k_{\text{on}} \left[\langle mn \rangle_{01} + \langle mn \rangle_{10} \right].$$

(A9-4)

The rate equation for total number of TF (sum of the free TFs in each state and the bound TFs in state 2, 3, and 4) and the total target protein can be written as

$$\frac{d\langle n \rangle}{dt} = \frac{d}{dt} \left[\langle n \rangle_{00} + \langle n \rangle_{01} + \langle n \rangle_{10} + \langle n \rangle_{11} + S_{01} + S_{10} + 2S_{11} \right]$$

$$= \alpha(S_{00} + S_{01}) - \gamma \langle n \rangle$$

$$\frac{d\langle m \rangle}{dt} = \frac{d}{dt} \left[\langle m \rangle_{00} + \langle m \rangle_{01} + \langle m \rangle_{10} + \langle m \rangle_{11} \right]$$

$$= \alpha(S_{00} + S_{10}) - \gamma \langle m \rangle.$$

(A9-5)

The steady state expression for total TF and target can be obtained by setting Eqn. A9-5 to zero which yields

$$\langle n \rangle_{\text{ss}} = \frac{\alpha}{\gamma}(S_{00} + S_{01}) = C_0(S_{00} + S_{01}),$$

$$\langle m \rangle_{\text{ss}} = \frac{\alpha}{\gamma}(S_{00} + S_{10}) = C_0(S_{00} + S_{10}),$$

(A9-6)

where $C_0 = \alpha/\gamma$ is the constitutive protein expression. The asymmetry defined as the difference of fold change in expression of target and TF gene expression is given by

$$\text{Asymmetry} = \text{FC}_{\text{Target}} - \text{FC}_{\text{TF}},$$

$$= \frac{\langle m \rangle_{\text{ss}}}{C_0} - \frac{\langle n \rangle_{\text{ss}}}{C_0},$$

$$= S_{10} - S_{01}.$$

(A9-7)

The asymmetry in TF and target regulation simply depends on the difference of occupancy in the states where TF gene is bound and where the target gene is bound. Furthermore, the steady state occupancies are given by (setting Eqns. A9-2 to zero)

$$S_{01} = S_{11} + \frac{k_{\text{on}}}{(\gamma + k_{\text{off}})} \Big[\langle n \rangle_{00} - \langle n \rangle_{01} \Big],$$

$$S_{10} = S_{11} + \frac{k_{\text{on}}}{(\gamma + k_{\text{off}})} \Big[\langle n \rangle_{00} - \langle n \rangle_{10} \Big], \tag{A9-8}$$

$$S_{11} = \frac{k_{\text{on}}}{2(\gamma + k_{\text{off}})} \Big[\langle n \rangle_{01} + \langle n \rangle_{10} \Big].$$

The asymmetry using Eqns. A9-7 and A9-8 is then given by

$$\text{Asymmetry} = \frac{k_{\text{on}}}{(\gamma + k_{\text{off}})} \Big[\langle n \rangle_{01} - \langle n \rangle_{10} \Big]. \tag{A9-9}$$

Eqn. A9-9 clearly demonstrates that the asymmetry in TF and target expression arises from the difference in the free TF concentration in state 2 (only target gene bound) and state 3 (only TF gene bound). Analytical expression for free TFs in different state cannot be determined explicitly as it can be seen from the Eqns. A9-3 and A9-4 that the mean protein ($\langle n \rangle, \langle m \rangle$) depends on the higher order moments ($\langle n^2 \rangle, \langle mn \rangle$) which then depends on the next higher order moments and so on.

1158 **Appendix 10**

1159 ## Modified ODEs for the minimal model

1160 The asymmetry, as explained in the main text and evident from Eqn. A9-9, appears due
1161 to the difference in the TF concentration when only the TF gene is occupied and when
1162 only the target gene is occupied. The general deterministic approach does not capture
1163 this asymmetry due to the mean field assumption of uniform TF concentration in all the
1164 states. To incorporate the difference in TF concentration in the deterministic model we now
1165 specifically assume the four state model; 1) both the TF gene and target gene are free to
1166 express, 2) TF gene is bound by TF, 3) target gene is bound by TF, and 4) both the genes are
1167 bound by TF. The number of cells in each state are $S_1, S_2, S_3,$ and S_4 and the total population
1168 (S) is constant. The free TF and total target protein number in each states are $(n_1, m_1), (n_2, m_2),$
1169 $(n_3, m_3),$ and (n_4, m_4) such that the average free TFs in each cell is $\langle n \rangle_i = n_i/S_i$ and average
1170 target protein in each cell is $\langle m \rangle_i = m_i/S_i$. State 1 switches to state 2 and 3 when a free TF
1171 binds to the free promoter of TF gene or target gene. State 2 and state 3 switch to state 1
1172 when a bound TF unbinds or degrade from the gene. State 2 and state 3 also switch to state
1173 4 due to TF binding. Finally, state 4 switches to state 2 and state 3 when a bound TF unbinds
1174 or degrade from the gene.

1175 Change in cell number due to the reactions that switch the cells from state i to state j
1176 causing an increase(state j) or decrease (state i) in the cell population per unit time are

1177
$$\text{Binding}: \quad k_{\text{on}} \langle n_i \rangle S_i = k_{\text{on}} \frac{n_i}{S_i} S_i = k_{\text{on}} n_i$$
1178
$$\text{Unbinding}: \quad k_{\text{off}} S_i \tag{A10-1}$$
1179
1180
$$\text{Degradation of TF from gene}: \quad \gamma S_i$$

1181 When a TF binds to a promoter of TF gene or target gene in state i switching the cells
1182 to state j the number of free TF of the cells in state j increases by the $(\langle n \rangle_i - 1)$ times the
1183 number of cell switched $(k_{\text{on}} n_i)$ and the number of target protein increases by $\langle m \rangle_i k_{\text{on}} n_i$. It
1184 is to be noted that a binding event decreases the average free TF pool by one in the cells
1185 which switch from state i to state j. In the process the cells in state i loses $\langle n \rangle_i k_{\text{on}} n_i$ number
1186 of free TFs and $\langle m \rangle_i k_{\text{on}} n_i$ number of target. Similarly, when a TF unbinds from a promoter
1187 switching state i to state j the number of free TFs of cells in state j increases by $(\langle n \rangle_i + 1)$
1188 times the number of cell switched $(k_{\text{off}} S_i)$ and the number of free TFs of each cell in state i
1189 goes down by $\langle n \rangle_i$ times the number of cell switched. The target protein number of cells in
1190 state i goes down by $\langle m \rangle_i k_{\text{off}} n_i$ and increase by the same amount in state j. Degradation of
1191 bound TF changes the free TF number by $\langle n \rangle_i \gamma S_i$ and target protein number by $\langle m \rangle_i \gamma S_i$. The
1192 change in protein number for all the reactions are listed in **Appendix 10 Table 1**.

1193 The set of ODEs describing the time evolution of the cell populations (S_i) in each state is
1194 then given by

1195
$$\frac{dS_1}{dt} = -2k_{\text{on}} n_1 + (k_{\text{off}} + \gamma)(S_2 + S_3),$$
1196
$$\frac{dS_2}{dt} = k_{\text{on}} n_1 - k_{\text{on}} n_2 + (k_{\text{off}} + \gamma)(S_4 - S_2),$$
1197
$$\tag{A10-2}$$
1198
$$\frac{dS_3}{dt} = k_{\text{on}} n_1 - k_{\text{on}} n_3 + (k_{\text{off}} + \gamma)(S_4 - S_3),$$
$$\frac{dS_4}{dt} = k_{\text{on}} n_2 + k_{\text{on}} n_3 - 2(k_{\text{off}} + \gamma)S_4.$$

The rate equations for free TF number can be written as

$$\frac{dn_1}{dt} = \alpha S_1 - \gamma n_1 + k_{off}(n_2 + S_2) + k_{off}(n_3 + S_3) - 2k_{on}\frac{n_1^2}{S_1} + \gamma(n_2 + n_3),$$

$$\frac{dn_2}{dt} = \alpha S_2 - \gamma n_2 + k_{off}(n_4 + S_4) - k_{off}n_2 + k_{on}\frac{n_1(n_1 - S_1)}{S_1} - k_{on}\frac{n_2^2}{S_2} + \gamma(n_4 - n_2),$$

$$\frac{dn_3}{dt} = -\gamma n_3 + k_{off}(n_4 + S_4) - k_{off}n_3 + k_{on}\frac{n_1(n_1 - S_1)}{S_1} - k_{on}\frac{n_3^2}{S_3} + \gamma(n_4 - n_3),$$

$$\frac{dn_4}{dt} = -\gamma n_4 - 2k_{off}n_4 + k_{on}\frac{n_2(n_2 - S_2)}{S_2} + k_{on}\frac{n_3(n_3 - S_3)}{S_3} - 2\gamma n_4,$$

(A10-3)

and the rate equations for target protein number is given by

$$\frac{dm_1}{dt} = \alpha S_1 - \gamma m_1 + k_{off}m_2 + k_{off}m_3 - 2k_{on}\frac{m_1 n_1}{S_1} + \gamma(m_2 + m_3),$$

$$\frac{dm_2}{dt} = -\gamma m_2 + k_{off}m_4 - k_{off}m_2 + k_{on}\frac{m_1 n_1}{S_1} - k_{on}\frac{m_2 n_2}{S_2} + \gamma(m_4 - m_2),$$

$$\frac{dm_3}{dt} = \alpha S_3 - \gamma m_3 + k_{off}m_4 - k_{off}m_3 + k_{on}\frac{m_1 n_1}{S_1} - k_{on}\frac{m_3 n_3}{S_3} + \gamma(m_4 - m_3),$$

$$\frac{dm_4}{dt} = -\gamma m_4 - 2k_{off}m_4 + k_{on}\frac{m_2 n_2}{S_2} + k_{on}\frac{m_3 n_3}{S_3} - 2\gamma m_4.$$

(A10-4)

Using Eqns. A10-2-A10-4, the rate equations for total TF and target protein can be written as

$$\frac{dn}{dt} = \frac{d}{dt}(n_1 + n_2 + n_3 + n_4 + S_2 + S_3 + 2S_4),$$

$$= \alpha(S_1 + S_2) - \gamma n,$$

$$\frac{dm}{dt} = \frac{d}{dt}(m_1 + m_2 + m_3 + m_4),$$

$$= \alpha(S_1 + S_3) - \gamma m.$$

(A10-5)

The steady state concentration for total TF and target protein is obtained by setting Eqn. A10-5 to zero which gives

$$n_{ss} = \frac{\alpha}{\gamma}(S_{1,ss} + S_{2,ss}) = C_0(S_{1,ss} + S_{2,ss}),$$

$$m_{ss} = \frac{\alpha}{\gamma}(S_{1,ss} + S_{3,ss}) = C_0(S_{1,ss} + S_{3,ss}).$$

(A10-6)

Here, $C_0 = \alpha/\gamma$ is the protein number of unregulated gene (constitutive expression). The steady state number of cells in states in terms of free TF number can be obtained by setting Eqn. A10-2 to zero and is given by

$$S_{1,ss} = S - (S_{2,ss} + S_{3,ss} + S_{4,ss}) = S - \frac{k_{on}}{2(k_{off} + \gamma)}(4n_{1,ss} + n_{2,ss} + n_{3,ss}),$$

$$S_{2,ss} = \frac{k_{on}}{2(k_{off} + \gamma)}(2n_{1,ss} - n_{2,ss} + n_{3,ss})$$

$$S_{3,ss} = \frac{k_{on}}{2(k_{off} + \gamma)}(2n_{1,ss} + n_{2,ss} - n_{3,ss})$$

$$S_{4,ss} = \frac{k_{on}}{2(k_{off} + \gamma)}(n_{2,ss} + n_{3,ss}).$$

(A10-7)

Setting $S = 1$ converts the number of cells (S_i) to occupancy of the cell in each state and n_i, m_i to fractional average of free TF and target protein per cell, *i.e.* $n_i = n_{ss}S_i$ and $m_i = m_{ss}S_i$. The asymmetry defined as the difference of fold change in expression of target and TF gene expression is given by

$$
\begin{aligned}
\text{Asymmetry} &= \text{FC}_{\text{Target}} - \text{FC}_{\text{TF}}, \\
&= \frac{m_{ss}}{C_0} - \frac{n_{ss}}{C_0}, \\
&= S_{3,ss} - S_{2,ss}, \\
&= \frac{k_{on}}{k_{off} + \gamma}(n_{2,ss} - n_{3,ss}).
\end{aligned}
\tag{A10-8}
$$

It is important to note that the same set of ODEs (Eqns. A10-2-A10-4) can be derived from CME by setting the variance and covariance of protein number in each state to zero. This modified ODEs predicts asymmetry between the TF and target expressions as shown in **Appendix 10 Figure 1A**, however, the predicted asymmetry doesn't match quantitatively with the CME predictions (see **Appendix 10 Figure 1B**). This discrepancy arises because of the fluctuations in the protein number in each state which is not considered in the modified ODEs.
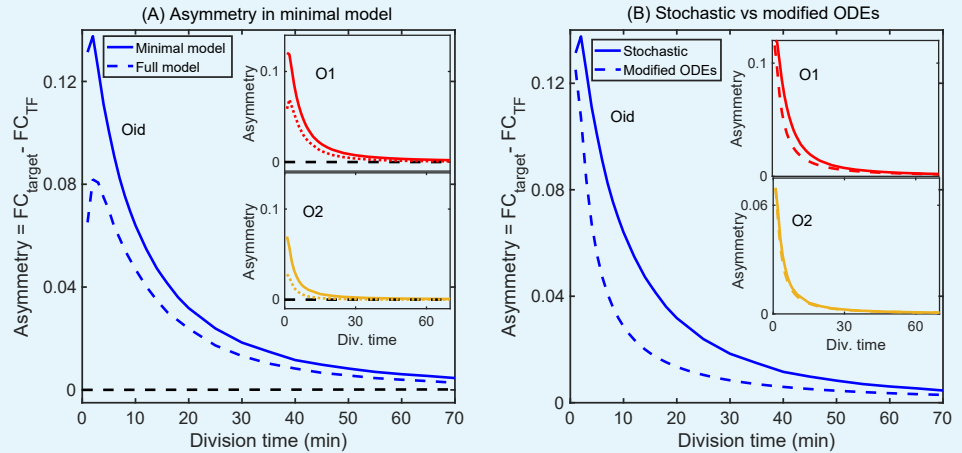
Appendix 10 Figure 1. Minimal model of autoregulation. (**A**) Asymmetry predicted from a minimal model without intermediate transcription steps and decoy binding sites using stochastic simulations (solid lines in blue, red and yellow for Oid, O1 and O2 binding sites, respectively). The asymmetry follows similar trend as predicted in the complete stochastic model(shown as dashed lines). Stronger binding site (Oid, shown in solid blue line) shows higher asymmetry than a weak binding site (O2, shown in solid yellow line). Also, asymmetry decreases as the growth rate is increased. Black dashed line corresponds to the deterministic counterpart of the stochastic reaction systems. Again, we do not find any asymmetry in TF and target regulation from the deterministic solution. (B) Modified ODEs with the inclusion of four states each having a different TF concentration predict asymmetry (dashed lines for different binding sites Oid (blue), O1 (red), and O2(yellow)). However, the quantitative values disagrees from the stochastic simulations of minimal model shown as solid lines.

Reaction	Increase in free TF	Decrease in free TF
Production in active state	αS_i	-
Degradation of free TF	-	γn_i
Binding	$\left(\frac{n_i}{S_i} - 1\right) k_{\text{on}} n_i = k_{\text{on}} \frac{n_i(n_i - S_i)}{S_i}$	$\frac{n_i}{S_i} k_{\text{on}} n_i = k_{\text{on}} \frac{n_i^2}{S_i}$
Unbinding	$\left(\frac{n_i}{S_i} + 1\right) k_{\text{off}} S_i = k_{\text{off}}(n_i + S_i)$	$\left(\frac{n_i}{S_i} + 1\right) k_{\text{off}} S_i = k_{\text{off}}(n_i + S_i)$
Degradation of TF from gene	$\frac{n_i}{S_i} \gamma S_i = \gamma n_i$	$\frac{n_i}{S_i} \gamma S_i = \gamma n_i$
	Increase in target	**Decrease in target**
Production in active state	αS_i	-
Degradation of target	-	γm_i
Binding	$\frac{m_i}{S_i} k_{\text{on}} n_i = k_{\text{on}} \frac{m_i n_i}{S_i}$	$\frac{m_i}{S_i} k_{\text{on}} n_i = k_{\text{on}} \frac{m_i n_i}{S_i}$
Unbinding	$\frac{m_i}{S_i} k_{\text{off}} S_i = k_{\text{off}} m_i$	$\frac{m_i}{S_i} k_{\text{off}} S_i = k_{\text{off}} m_i$
Degradation of TF from gene	$\frac{m_i}{S_i} \gamma S_i = \gamma m_i$	$\frac{m_i}{S_i} \gamma S_i = \gamma m_i$

Appendix 10 Table 1. Change in free TF and target protein number for the reactions describing the minimal model
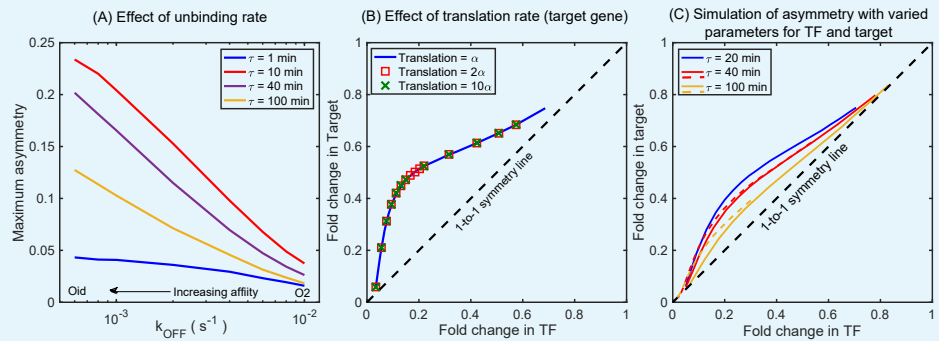
Appendix 11


Appendix 11 Figure 1. Simulations showing the effects of rate parameters on asymmetry. (**A**) Effect of TF unbinding rate (k_{OFF}) on asymmetry. Irrespective of the k_{OFF}, the maximum asymmetry decreases monotonically. (**B**) Asymmetry is not affected by difference in translation rate between the TF gene and the target gene. Blue solid curve represents asymmetry obtained from simulations where the translation rate of TF gene and the target gene is exactly same. The data points are generated with a translation rate of target gene twice (red square) and ten times (green cross) that of the TF gene and fall exactly on the blue curve showing no deviation. (**C**) Asymmetry for different growth rate (τ) with varying transcription rate, translation rate, and mRNA stability. Stochastic simulation performed using the kinetic parameters listed in *Bremer and Dennis* (*2008*) for τ being 20 (blue line), 40 (red line), and 100 (yellow line) minutes. Dashed lines show the asymmetry for τ = 40 min and 100 min for the rate parameters same as τ = 20 min. The qualitative ordering and features of the asymmetry curve is not impacted by the changes in the kinetic parameters such as transcription rate, translation rate, and mRNA stability due to change in growth rates.

1277 **Appendix 12**
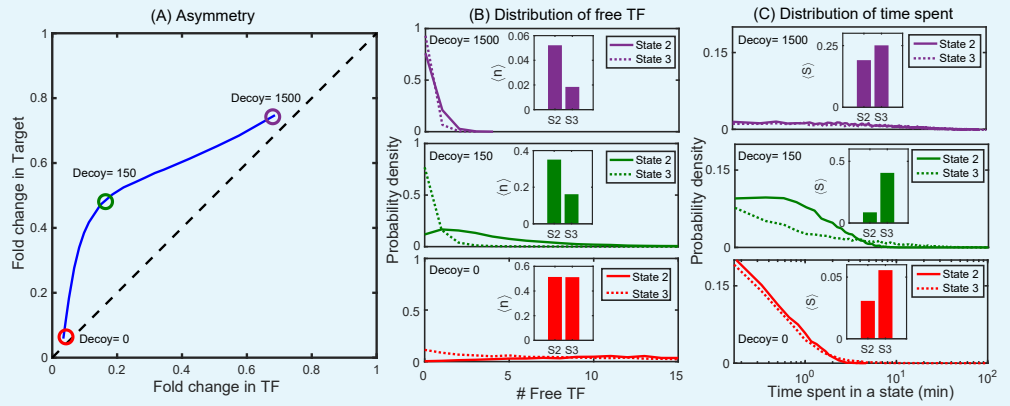
1278

1279 **Appendix 12 Figure 1. Distributions of free TFs and time spent in different promoter states. (A)**
1280 Typical asymmetry plot obtained from simulations for Oid binding site with division time τ = 25 min.
1281 (B-C) Distribution of free TFs and time spent in state 2 (S2) and state 3 (S3) for varying level of
1282 asymmetry corresponding to different decoy number as shown in panel (A). The plots in red, green and
1283 purple correspond to no decoys (low asymmetry), 150 decoy (maximum asymmetry) and 1500 decoy
1284 (low asymmetry). Insets in (B) are steady state fractional average of free TFs in state 2 and state 3
1285 obtained from stochastic simulations using equation $\langle n \rangle = \sum_{m,n} n P_{i,j}$, where $P_{i,j}$ is the probability of
1286 having m target protein and n free TF in the promoter state (i,j); see Appendix 9. Insets in (C) are
1288 TF-occupancy in state 2 and state 3 defined as $\langle n \rangle = \sum_{m,n} P_{i,j}$.