

**A peer-reviewed version of this preprint was published in PeerJ on 29 July 2014.**

[View the peer-reviewed version](https://peerj.com/articles/498) (peerj.com/articles/498), which is the preferred citable publication unless you specifically need to cite this preprint.

Sochor MA. 2014. *In vitro* transcription accurately predicts lac repressor phenotype *in vivo* in *Escherichia coli*. PeerJ 2:e498  
<https://doi.org/10.7717/peerj.498>

1 **Title**

2 *In vitro* transcription accurately predicts lac repressor phenotype *in vivo* in *Escherichia coli*

3

4 **Author**

5 Matthew Almond Sochor

6

7 **Affiliation**

8 Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine,  
9 37th and Hamilton Walk, Philadelphia, PA 19104-6059, USA

10

11 **Abstract**

12 A multitude of studies have looked at the *in vivo* and *in vitro* behavior of the lac repressor  
13 binding to DNA and effector molecules in order to study transcriptional repression, however  
14 these studies are not always reconcilable. Here we use *in vitro* transcription to directly mimic the  
15 *in vivo* system in order to build a self consistent set of experiments to directly compare *in vivo*  
16 and *in vitro* genetic repression. A thermodynamic model of the lac repressor binding to operator  
17 DNA and effector is used to link DNA occupancy to either normalized *in vitro* mRNA product or  
18 normalized *in vivo* fluorescence of a regulated gene, YFP. An accurate measurement of repressor,  
19 DNA and effector concentrations were made both *in vivo* and *in vitro* allowing for direct  
20 modeling of the entire thermodynamic equilibrium. *In vivo* repression profiles are accurately  
21 predicted from the given *in vitro* parameters when molecular crowding is considered.  
22 Interestingly, our measured repressor-operator DNA affinity differs significantly from previous  
23 *in vitro* measurements. The literature values are unable to replicate *in vivo* binding data. We  
24 therefore conclude that the repressor-DNA affinity is much weaker than previously thought. This  
25 finding would suggest that *in vitro* techniques that are specifically designed to mimic the *in vivo*  
26 process may be necessary to replicate the native system.

## 13 Introduction

14

15 The lac genetic switch consists of the lac repressor, a short “operator” DNA sequence,  
16 and effector molecules (Swint-Kruse & Matthews, 2009). The minimal functional lac repressor is  
17 homo-dimeric and includes an N-terminal DNA binding domain and two effector binding sites  
18 (one per monomer). Repressor binds to operator DNA preventing RNA polymerase from  
19 transcribing downstream genes. Effector molecules bind to each effector binding site causing an  
20 allosteric transition wherein repressor dissociates from operator DNA allowing transcription to  
21 proceed (Lewis, 2005). Previously our lab has used a standard Monod, Wyman, and Changeux  
22 (MWC) model of thermodynamic equilibrium to model the behavior of the lac genetic switch  
23 (Fig. 1) (Monod, Wyman, & Changeux, 1965).

24 While the underpinnings of the lac genetic switch have been well characterized, it is less  
25 well understood how to utilize this information to achieve practical goals. How do we reduce the  
26 background leakiness of the repressor? Can you do so without compromising maximal  
27 inducibility? Can you target certain phenotypic properties through directed mutation? Will novel  
28 genetic switches developed in *E. coli* perform the same in different cell types? Significant  
29 advancement has been made in recent years towards answering these more complex questions.

30 Daber, et al. examined the number of effector molecules necessary to induce transcription  
31 (Daber, Sharp, & Lewis, 2009). Hetero-dimeric lac repressors were created that bound either 0, 1  
32 or 2 effector molecules and the *in vivo* regulation of a fluorescent gene was measured. An  
33 analytical solution of a simplified MWC equilibria allowed for direct measurements of  
34 dimensionless bulk parameters comprised of combinations of thermodynamic binding constants  
35 and species concentrations. While these parameters were useful in showing that two effector  
36 molecules are required for fully inducing the genetic switch, they were unable to measure the  
37 thermodynamic constants themselves.

38 Daber, et al. next sought to link distinct perturbations of the lac genetic switch to changes  
39 in thermodynamic parameters (Daber, Sochor, & Lewis, 2011). Mutations were made in the  
40 DNA binding domain and effector binding pocket of the repressor. They were able to measure  
41 the repressor-effector binding affinities; however they still could only measure a dimensionless  
42 constant which contained repressor concentration and repressor-DNA affinity. Mutations in the  
43 DNA binding domain of the lac repressor were linked to changes in the repressor-DNA affinity.  
44 Alternatively, changes in the repressor concentration could also account for the phenotype.  
45 Mutations in the effector binding domain did alter the effector binding affinities. Interestingly,  
46 effector binding domain mutations were also linked to changes in the conformational equilibrium  
47 of the repressor, but once again changes in the repressor concentration could account for the  
48 phenotype. These results were encouraging evidence that directed mutations lead to directed  
49 phenotypes, but the question of repressor concentration clouded the picture.

50 A study by Poelwijk, et al. looked for unique phenotypes through random mutagenesis of  
51 the lac repressor (Poelwijk, de Vos, & Tans, 2011). Mutants were identified which exhibited an  
52 inverted repression behavior; a phenotype also found by Daber, et al. by mutating the effector  
53 binding domain (Daber, Sochor, & Lewis, 2011). Interestingly, Poelwijk’s mutations were in

54 regions physically distinct from either the DNA or effector binding domains. One potential  
55 explanation is that the mutations destabilize the folded form of the repressor, altering the  
56 conformational landscape. Mutagenesis of the repressor can result in more than just predictable  
57 changes of thermodynamic binding constants.

58 Central to all of these studies is the use of *in vivo* data to understand the behavior of  
59 genetic switches. It has been pointed out that a lack of corroborating *in vitro* evidence prevents  
60 the identification of other processes which may significantly play into the equilibrium, such as  
61 non-specific DNA binding or effector uptake (Tungtur et al., 2011). They attempted to measure  
62 the thermodynamic binding of a LacI/GalR hybrid repressor both *in vitro* and *in vivo*. Notably, a  
63 DNA pull down assay was used to quantify the *in vivo* concentration of their hybrid repressor.  
64 Unfortunately, they were unable to rectify a greater than 25-fold difference between their two  
65 data sets. This indicates that they are missing a significant contributor to the genetic switch by  
66 only analyzing *in vivo* data.

67 Here we sought to overcome the limitations of past studies three ways: 1.) measure the *in*  
68 *vivo* concentration of the lac repressor, 2.) measure the *in vitro* transcription of purified lac  
69 genetic switch, and 3.) use an assumption free solution to the MWC equilibrium to model both *in*  
70 *vitro* and *in vivo* data.

71 We were able to measure lac repressor concentration *in vivo* and use *in vitro* transcription  
72 to assess the purified lac genetic switch. Furthermore, we found excellent agreement between *in*  
73 *vitro* and *in vivo* data when molecular crowding was taken into consideration. We do however  
74 find that the repressor-DNA affinity is much lower than has previously been measured *in vitro*.  
75 Additional concerns, such as effector uptake and non-specific DNA binding do not appear to play  
76 significant roles.

## 77 **Materials and methods**

### 78 *In vivo measurement of lac genetic switch*

79 Reporter plasmid was made as previously reported (Daber & Lewis, 2009) with the O1  
80 operator sequence (5'-AA TT GTG AGC G GAT AAC AA TT-3') followed by YFP and  
81 providing ampicillin (AMP) resistance. Lac repressor was expressed on a second plasmid as  
82 previously described (Daber & Lewis, 2009) providing chloramphenicol (CAM) resistance. A C-  
83 terminal mCherry tag was added to the Lac repressor gene after an 11bp linker to create the Lac-  
84 mCherry construct.

85 We double transformed reporter and repressor plasmids into EPB229 cells ( $F^- \Delta(\text{lacI-}$   
86  $\text{lacA})::\text{frt}$ ). These cells were derived from the MG1655 "wild type" line. Colonies were picked in  
87 triplicate into MOPS minimal media with 0.4% glucose, AMP and CAM and grown overnight at  
88 37C with shaking. 50 $\mu$ L of the overnight culture was used to inoculate 1mL fresh MOPS  
89 minimal media supplanted with varying amounts of IPTG. We measured optical density at  
90 600nm (OD<sub>600</sub>), YFP fluorescence (excite: 510 nm emit: 535 nm), and mCherry fluorescence  
91 (excite: 585 nm emit: 610 nm) for all wells at 1 hour intervals over a 12 hour period using a  
92 TECAN M1000 plate reader in 384 well optical bottom plates (Corning).

### 93 *Purification of Lac-mCherry*

94 Lac-mCherry was cloned into the pBAD-DEST49 expression vector (Clontech). A 6xHis  
95 C-terminal tag was added to aid in purification. BL21(DE3) cells were transformed and grown to  
96 mid-log at 37C with shaking in 2xYT media. At mid-log growth, expression of Lac-mCherry was  
97 induced with the addition of arabinose 0.1% (v/v) and the temperature was reduced to 15C and  
98 cells were allowed to grow overnight (approximately 12-16 hours). Cell extract was purified with  
99 Ni-NTA beads (Clontech) and a sizing column (HiLoad 16/60 Superdex 75 Prep Grade with  
100 AKTA Prime FPLC) and purified Lac-mCherry was equilibrated into GF buffer (200 mM Tris  
101 pH 7.4, 200 mM KCl, 10mM EDTA, 3mM DTT).

#### 102 *Measuring in vivo concentrations of the lac repressor*

103 EPB229 cells were co-transformed with Lac-mCherry and O1 YFP reporter. An  
104 individual colony was picked into MOPS minimal media with 0.4% glucose, AMP and CAM and  
105 grown overnight at 37C with shaking. 50  $\mu$ L was inoculated into 1 mL fresh media and grown  
106 to mid-log phase.

107 Purified Lac-mCherry was quantitated with both a BCA Assay Kit (Pierce) and optical  
108  $A_{280}$  measurements using a NanoDrop 2000 Spectrometer (Thermo Scientific). Dilutions were  
109 made over 8 orders of magnitude and 50  $\mu$ L was loaded into clear bottom 384 well plates in  
110 triplicate. mCherry fluorescent measurements (excite: 585 nm emit: 610 nm) were made using  
111 various gains to establish linear regimes for the instrument (TECAN M1000).

112 We established a raw cell count by plating dilutions of a culture of EPB229 cells. Serial  
113 dilutions were made over 10 orders of magnitude and each dilution had  $OD_{600}$  measured  
114 (TECAN M1000 and Ultrospec 2100 pro) and 100  $\mu$ L plated onto LB agar with AMP and CAM.  
115 We found  $1.92 \times 10^6$  cells/ $\mu$ L at mid-log growth phase which is about two-fold higher than  
116 standard estimates of  $1 \times 10^6$  cells/ $\mu$ L for *E. coli*. Aliquots of known cell counts were then used to  
117 establish a linear relationship with  $OD_{600}$  on our plate reader. Similarly, purified Lac-mCherry of  
118 known concentration was used to establish a linear relationship with mCherry fluorescence on  
119 our plate reader at a fixed gain.

120 EPB229 cells were co-transformed with plasmid constitutively expressing Lac-mCherry  
121 and a reporter plasmid which has YFP under the control of the natural operator O1. We measured  
122 mCherry fluorescence at a fixed gain and  $OD_{600}$  from which we calculated the concentration of  
123 Lac-mCherry in the well and the number of cells in the well. The approximate volume of *E. coli*  
124 was estimated to be  $1 \times 10^{-15}$  L (Kubitschek & Friske, 1986). Multiplying volume of *E. coli* by  
125 number of cells allows us to estimate what fraction of the well volume is intracellular.

126 Calibration of raw mCherry fluorescent signal and  $OD_{600}$  was converted to intracellular  
127 repressor concentration.

#### 128 *Fluorescent data processing*

129 *In vivo* data was normalized for growth by measuring cells in triplicate as they were  
130 growing. All data points collected were then fit to a 2<sup>nd</sup> order polynomial to obtain a curve which  
131 is fluorescence as a function of  $OD_{600}$ . Positive control was established by co-transforming  
132 EPB229 cells with O1 YFP reporter and a CAM plasmid without Lac-mCherry (pABD34). YFP  
133 signal was normalized to the polynomial fit from the positive control. Final values for fitting  
134 were calculated for cells at approximately mid-log growth phase ( $OD_{600} = 0.4$ ).

135 *Measuring in vitro transcription*

136 A reporter plasmid was made with the O1 operator after a T7 promoter. Reporter was  
137 linearized to 450bp and purified by spin column purification (Clontech).

138 MaxiScript T7 kit (Ambion) was used to perform *in vitro* transcription. CTP[ $\alpha$ - $^{32}$ P] was  
139 incorporated into mRNA transcripts and the water fraction of the standard reaction was  
140 supplanted with varying concentrations of Lac-mCherry and IPTG. Transcription was allowed to  
141 proceed for 30 minutes at 37C until halted by boiling. Samples were loaded onto polyacrylamide  
142 gels and electrophoresis was used to separate free CTP[ $\alpha$ - $^{32}$ P] from that incorporated into  
143 mRNA. Gels were dried and exposed to radiological plates. Plates were imaged on a Typhoon  
144 scanner and bands were quantitated using ImageJ (NIH).

145 *Modeling*

146 Experimentally, we would like to measure the output from a promoter regulated by the  
147 lac genetic switch. It is assumed that transcription by RNA polymerase from the promoter is  
148 linearly related to the occupancy of the DNA operator within the promoter by the lac repressor,

$$\text{transcription} \propto \frac{[O]}{[O]_{\text{tot}}} \quad (1)$$

149 In order to model experimental data, we need to compute the occupancy of the DNA  
150 operator in terms of the thermodynamic constants ( $K_{RR^*}$ ,  $K_{RE}$ ,  $K_{R^*E}$ ,  $K_{RO}$ , and  $K_{R^*O}$ ) and the total  
151 concentration of repressor, effector and operator ( $[R]_{\text{tot}}$ ,  $[E]_{\text{tot}}$ , and  $[O]_{\text{tot}}$ ).

152 Start by defining the following affinity constants in equilibrium:

$$K_{RR^*} = \frac{[R^*]}{[R]} \quad (2)$$

$$K_{RE} = \frac{[RE]}{[R][E]} \quad (3)$$

$$K_{[R^*E]} = \frac{[R^*E]}{[R][E]} \quad (4)$$

$$K_{RO} = \frac{[RO]}{[R][O]} \quad (5)$$

$$K_{R^*O} = \frac{[R^*O]}{[R^*][O]} \quad (6)$$

153 We also need to define the total concentrations of operator, effector and repressor in terms  
 154 of the individual bound and conformational states,

$$[O]_{\text{tot}} = [O] + [RO] + 2[REO] + [RE_2O] + [R^*O] + 2[R^*EO] + [R^*E_2O] \quad (7)$$

$$[E]_{\text{tot}} = [E] + 2[RE] + 2[RE_2] + 2[R^*E] + 2[R^*E_2] + 2[ROE] + 2[ROE_2] + 2[R^*OE] + 2[R^*OE_2] \quad (8)$$

$$[R]_{\text{tot}} = [R] + 2[RE] + [RE_2] + [R^*] + 2[R^*E] + [R^*E_2] + [RO] + 2[REO] + [RE_2O] + [R^*O] + 2[R^*EO] + [R^*E_2O] \quad (9)$$

155 Of note are the various coefficients of 2. All of the singly bound effector species are  
 156 degenerate since the effector can bind to either the left or right effector site, which gives rise to  
 157 the statistical mass balancer 2. For Equation 8, the doubly bound effector species have two  
 158 effector molecules bound and hence are doubled.

159 The strategy is to write all of the equations in terms of the free species concentrations  
 160 ( $[R]$ ,  $[E]$ ,  $[O]$ ) and the equilibrium constants in Equations 2-6. Then we try to rearrange such that  
 161 we can make polynomials of just  $[E]$ . The reasons will become apparent after we have done the  
 162 above operations.

163 Starting with Equation 9, we re-write using only free species and constants,

164

$$[RO] = K_{RO} [R][O] \quad (10)$$

$$[REO] = K_{RE} K_{RO} [R][E][O] \quad (11)$$

$$[RE_2O] = K_{RE}^2 K_{RO} [R][E]^2 [O] \quad (12)$$

$$[R^*O] = K_{RR^*} K_{R^*O} [R][O] \quad (13)$$

$$[R^*EO] = K_{RR^*} K_{R^*E} K_{R^*O} [R][E][O] \quad (14)$$

$$[R^*E_2O] = K_{RR^*} K_{R^*E}^2 K_{R^*O} [R][E]^2 [O] \quad (15)$$

$$\begin{aligned}
 [R]_{\text{tot}} = & [R] + 2[R][E]K_{RE} + [R][E]^2K_{RE}^2 \\
 & + [R]K_{RR^*} + 2[R][E]K_{RR^*}K_{R^*E} + [R][E]^2K_{RR^*}K_{R^*E}^2 \\
 & + [R][O]K_{RO} + 2[R][O][E]K_{RO}K_{RE} + [R][O][E]^2K_{RO}K_{RE}^2 \\
 & + [R][O]K_{RR^*}K_{R^*O} + 2[R][O][E]K_{RR^*}K_{R^*O}K_{R^*E} + [R][O][E]^2K_{RR^*}K_{R^*O}K_{R^*E}^2
 \end{aligned} \quad (16)$$

165 We then make the following definitions,

$$\alpha_1 = 1 + K_{RR^*} \quad (17)$$

$$\beta_1 = 2K_{RE} + 2K_{RR^*}K_{R^*E} \quad (18)$$

$$\gamma_1 = K_{RE}^2 + K_{RR^*}K_{R^*E}^2 \quad (19)$$

$$\gamma_2 = 2K_{RO}K_{RE} \quad (20)$$

$$\delta_1 = K_{RO}K_{RE}^2 \quad (21)$$

$$\beta_2 = K_{RR^*}K_{R^*O} \quad (22)$$

$$\gamma_3 = 2K_{RR^*}K_{R^*O}K_{R^*E} \quad (23)$$



$$\delta_2 = K_{RR} * K_{R*O} K_{R*E}^2 \quad (24)$$

166 Substituting into Equation 16 and re-arranging to isolate [R],

$$[R] = \frac{[R]_{tot}}{\alpha_1 + [E]\beta_1 + [E]^2 \gamma_1 + [O](K_{RO} + [E]\gamma_2 + [E]^2 \delta_1 + \beta_2 + [E]\gamma_3 + [E]^2 \delta_2)} \quad (25)$$

167 The equation has been organized such that polynomials in [E] are apparent. As long as  
 168 we only add and multiply polynomials, they can trivially be treated as symbolic functions for  
 169 further simplification. We define the following polynomials,

$$B_1 = \alpha_1 + [E]\beta_1 + [E]^2 \gamma_1 \quad (26)$$

$$B_2 = K_{RO} + \beta_2 + [E](\gamma_2 + \gamma_3) + [E]^2 (\delta_1 + \delta_2) \quad (27)$$

170 Now substituting back into Equation 25,

$$[R] = \frac{[R]_{tot}}{B_1 + [O]B_2} \quad (28)$$

171 We next want to follow the same path for [E] and [O]. Inspection of Equations 7-9 show  
 172 that we have already done the most complicated case. We can then quickly arrive at,

$$[O] = \frac{[O]_{tot}}{1 + [R]B_2} \quad (29)$$

173 The effector equation is similar, but it has a few extra coefficients of two within its  
 174 equations. We define two more polynomials,

$$A_1 = \beta_1 + 2[E]\gamma_1 \quad (30)$$

$$A_2 = \gamma_2 + \gamma_3 + 2[E](\delta_1 + \delta_2) \quad (31)$$

175 Substituting into Equation 8,

$$[E]_{\text{tot}} = [E] + [R][E]A_1 + [R][E][O]A_3 \quad (32)$$

176 We can then eliminate [O] by substituting Equation 29 into Equations 28 and 32. Since  
 177 we can only multiply and add polynomials, we multiply the denominator of Equation 29 on both  
 178 sides. Substituting into Equation 28,

$$[R]_{\text{tot}} + [R]B_2[R]_{\text{tot}} = [R]B_1 + [R]^2B_1B_2 + [R]B_2[O]_{\text{tot}} \quad (33)$$

179 We then define the following polynomials,

$$\varphi_1 = B_1B_2 \quad (34)$$

$$\varphi_2 = B_1 + B_2([O]_{\text{tot}} - [R]_{\text{tot}}) \quad (35)$$

180 Substituting into Equation 33,

$$[R]^2\varphi_1 + [R]\varphi_2 = [R]_{\text{tot}} \quad (36)$$

181 The substitution of Equation 29 into Equation 32 requires the following definitions,

$$\psi_1 = [E]A_1B_2 \quad (37)$$

$$\psi_2 = [E](B_2 + A_1 + A_2[O]_{\text{tot}}) - B_2[E]_{\text{tot}} \quad (38)$$

182 We then arrive at,

$$[R]^2\psi_1 + [R]\psi_2 = [E]_{\text{tot}} - [E] \quad (39)$$

183 We now have two equations (Eqn. 36 and 39) with two unknowns ([R] and [E]). In  
 184 principal we can get this down to a single equation, but in order to do so the final polynomial  
 185 becomes of a much higher order which prevents accurate computational solutions.

186 The strategy is then to guess at the free effector concentration to calculate Equations 34,  
 187 35, 37, and 38. Equations 36 and 39 can then be solved for [R] by looking for the roots to the  
 188 equation. When the correct free effector concentration ([E]) is found the roots of Equation 36 and

189 Equation 39 will converge. By minimizing the difference between the roots a solution can be  
190 reached. All other concentrations are then trivial to calculate once [R] and [E] are known.  
191 Custom Matlab (Mathworks) software was written to numerically solve the MWC equilibria  
192 (Matlab File Exchange ID #40602).

193 The accuracy of the solution is easily checked by using the bound and free species  
194 concentrations to calculate the total species concentrations and thermodynamic parameters.  
195 Calculated values should agree with input values.

196 Five independent thermodynamic parameters ( $K_{RE}$ ,  $K_{R^*E}$ ,  $K_{RO}$ ,  $K_{R^*O}$ , and  $K_{RR^*}$ ) were used  
197 for each model and all data points were simultaneously fit using a standard non-linear least  
198 squares algorithm in Matlab.

199 A Monte Carlo approach was used to estimate error in the fit parameters. The known  
200 error of the experiment was used to generate data sets with random error. 100 such data sets were  
201 generated and a non-linear least squares fitting algorithm was used to fit the thermodynamic  
202 parameters. Standard deviation of these fit thermodynamic parameters was used as the error of  
203 the best fit for the actual data set.

## 204 **Results and Discussion**

### 205 *Measuring the In Vivo Concentration of the Lac Repressor*

206 We sought a method where we could simultaneously measure lac repressor concentration  
207 and transcriptional regulation and thus chose to fluorescently tag the repressor. The fluorescent  
208 protein mCherry was chosen due to minimal auto-fluorescence from MOPS minimal media and  
209 minimal spectral overlap with our reporter gene YFP. Furthermore, a dimeric Lac-mCherry  
210 fusion construct is known to be functional *in vivo* (Lau et al., 2004). The goal is to measure raw  
211 mCherry fluorescence and OD<sub>600</sub> in growing *E. coli* cells and convert those measurements to an  
212 intracellular concentration of lac repressor (Fig. 2).

213 A linear relationship was established for OD<sub>600</sub> and cell count. We estimate the volume of  
214 *E. coli* growing in glucose supplemented minimal media to be  $1 \times 10^{-15}$  L (Kubitschek & Friske,  
215 1986). We then measured OD<sub>600</sub>, calculated the number of cells and multiplied by volume of the  
216 cell to calculate the fraction of the well that is intracellular. A linear relationship was also  
217 established for purified Lac-mCherry fluorescence and concentration of Lac-mCherry.

218 We assume all of the Lac-mCherry is intracellular; therefore we divided the Lac-mCherry  
219 concentration by the fraction of volume that is intracellular. Using this method, we can quickly  
220 and accurately measure *in vivo* Lac-mCherry concentrations.

### 221 *Intracellular Lac-mCherry concentration in EPB229 cells*

222 EPB229 cells expressing Lac-mCherry and the reporter plasmid were grown in varying  
223 concentrations of the inducer IPTG. Intracellular concentration of Lac-mCherry was calculated  
224 from mCherry fluorescence and OD<sub>600</sub> and found to be  $664 \pm 90$  nM at mid-log growth phase  
225 (OD<sub>600</sub> = 0.6). As expected for a constitutively expressed gene, minimal variation was seen with  
226 IPTG and cell growth (Fig. 3A).

227 We then converted to molecules per cell,

$$6.6 \times 10^{-7} \text{ M} * 1 \times 10^{-15} \frac{\text{L}}{\text{cell}} * 6.022 \times 10^{23} \frac{\text{molecules}}{\text{mole}} = 397 \frac{\text{molecules}}{\text{cell}} \quad (40)$$

228 We have previously estimated the copy number of our plasmid to be ~10-20 plasmids/cell  
 229 (Daber, Sharp, & Lewis, 2009). This corresponds to approximately 20-40 Lac-mCherry dimers  
 230 per plasmid which agrees well with previous estimates of ~40 Lac repressor dimers per plasmid  
 231 for our promoter (Oehler et al., 1994).

### 232 *Measuring the In Vivo Regulation of YFP*

233 In addition to mCherry fluorescence and OD<sub>600</sub> measurements, YFP fluorescence was  
 234 measured in cells as a function of IPTG. Unregulated expression was established by measuring  
 235 OD<sub>600</sub> and YFP in cells co-transformed with O1 YFP reporter and a plasmid which does not  
 236 contain any repressor (pABD34). These positive control cells were grown in tandem with cells  
 237 containing both reporter and repressor and grown in a variety of IPTG concentrations.

238 Positive controls showed no IPTG dependence as expected, so data from every sample  
 239 was combined to determine an overall positive control polynomial fit. YFP fluorescence is seen  
 240 to increase as cells grow as would be expected due to the increased number of cells per  $\mu\text{L}$ . We  
 241 remove this bias and normalize regulated YFP expression by dividing by the positive control fit  
 242 curve.

243 Normalized YFP expression was then measured as a function of OD<sub>600</sub> and IPTG (Fig.  
 244 3B. Almost no OD<sub>600</sub> dependence can be noted in the induction profile. The YFP signal is  
 245 repressed without IPTG and is approximately  $1.7 \pm 0.2\%$  of unregulated expression. Upon  
 246 induction with saturating IPTG we see a robust YFP increase to approximately  $61 \pm 5\%$  of the  
 247 unregulated expression.

### 248 *Measuring the In Vitro Regulation of mRNA*

249 While the *in vivo* experiment measures translation product (fluorescing YFP) we know  
 250 the lac repressor actually regulates mRNA production. Previously, our lab has determined a  
 251 linear relationship between mRNA and fluorescence protein signal allowing us to use  
 252 fluorescence as a proxy for mRNA regulation *in vivo* (Daber & Lewis, 2009). The situation *in*  
 253 *vitro* is reversed; it is much easier to measure mRNA production.

254 We used the Maxiscript T7 *in vitro* transcription kit (Ambion) which produces mRNA  
 255 from linearized DNA with a T7 promoter. We then measured incorporation of radioactive  
 256 labeled CTP into mRNA. The T7 promoter was modified to add an O1 operator DNA site and  
 257 we were able to modulate Lac-mCherry and IPTG concentrations. A positive control of  
 258 constitutive mRNA production is established by not adding any Lac-mCherry.

259 We first established that radioactively labeled mRNA was linearly observable by  
 260 constitutively producing mRNA and loading various dilutions onto polyacrylamide gels and  
 261 established that mRNA concentration was linearly related to the concentration of mRNA loaded  
 262 on the gel. Positive controls were included for every experiment and were used for  
 263 normalization.

264 The additional benefit of *in vitro* transcription is the flexibility in dosing not only IPTG,  
265 but also Lac-mCherry. We exploited this flexibility by first titrating in Lac-mCherry without  
266 IPTG present and with saturating IPTG (1mM) (Fig. 4A). As expected, increasing concentration  
267 of Lac-mCherry decreases mRNA production. Furthermore, addition of IPTG returns mRNA  
268 signal to near constitutive levels.

269 We then titrated IPTG at a fixed Lac-mCherry concentration (Fig. 4B). The induction of  
270 mRNA is seen to very closely resemble that of the *in vivo* data, but it is noticeably leakier.  
271 Maximal repression was about  $7.8 \pm 1.3\%$  and maximal induction was approximately  $88 \pm 9\%$ .

## 272 *Modeling Using MWC Thermodynamic Equilibrium*

273 Finally, we sought to simultaneously model the *in vivo* and *in vitro* data using the Monod,  
274 Wyman, and Changeux (MWC) model of thermodynamic equilibrium. Previously, we have  
275 relied upon approximate solutions of the lac genetic switch equilibrium to model *in vivo*  
276 induction profiles. This solution assumes that the total repressor concentration greatly exceeds  
277 operator concentration ( $[R]_{tot} \gg [O]_{tot}$ ). This condition does not hold for our *in vitro* experiment  
278 where we titrated in Lac-mCherry nor would it necessarily be true in all *in vivo* systems.  
279 Therefore, we sought a solution to the equilibrium that held for every potential input. An  
280 assumption free solution to the MWC model was found and is solved in detail in the methods.

### 281 *Using the assumption-free solution to measure thermodynamic parameters*

282 Experimentally we know the total concentrations ( $[R]_{tot}$ ,  $[E]_{tot}$ ,  $[O]_{tot}$ ) and normalized  
283 transcription/expression ( $[O]/[O]_{tot}$ ). We want to measure the thermodynamic constants ( $K_{RR*}$ ,  
284  $K_{RE}$ ,  $K_{R^*E}$ ,  $K_{RO}$ ,  $K_{R^*O}$ ). This leaves 5 independent constants in the MWC model to fit to the  
285 experimental data. The large number of independent constants results in a myriad of non-unique  
286 solutions to the equations. This complication was limited by the following algorithm.

287 First, since it is widely reported to be effectively zero,  $K_{R^*O}$  was set to be very, very small  
288 ( $1 \times 10^{-10} \text{ nM}^{-1}$ ). This leaves four independent parameters.

289 Next, it had been observed from previous studies that the ratio of  $K_{R^*E}$  to  $K_{RE}$  is well  
290 defined when the concentration of repressor greatly exceeds that of operator. Under this  
291 assumption, a simpler solution of the MWC equilibrium exists as previously reported (Daber,  
292 Sharp, & Lewis, 2009). We isolated a subset of the *in vitro* data where this condition was true  
293 and used a non-linear least squares fitting algorithm to measure the ratio  $X = K_{R^*E}/K_{RE}$  as a  
294 function of conformational equilibrium. The ratio was seen to asymptote at approximately  
295 13.75. This value is then used to reduce the number of independent constants to 3 ( $K_{RR*}$ ,  $K_{RE}$ , and  
296  $K_{RO}$ ).

297 We then simultaneously fit the *in vitro* data to obtain the best fit thermodynamic  
298 parameters using a non-linear least squares algorithm in Matlab (Table 1). The model accurately  
299 fits both the lac repressor (Figure 4A) and IPTG doping (Figure 4B) *in vitro* transcription  
300 experiments. The fit values agree well with values obtained in the literature with the exception  
301 of repressor-DNA affinity. The repressor-DNA affinity ( $K_{RO}$ ) was measured to be  $0.4 \pm 0.2 \text{ nM}^{-1}$ .  
302 This is significantly weaker than the  $100\text{--}3333 \text{ nM}^{-1}$  that has been measure previously (Sharp,  
303 2011). It does agree well with an estimated value of  $1 \text{ nM}^{-1}$  for lac repressor-DNA affinity that  
304 prevails under conditions within the *E. coli* cell (Müller-Hill, 1996). The thermodynamic

305 equilibrium value ( $6.3 \pm 3.3$ ) does not significantly differ from that measured previously by our  
306 group. The repressor-IPTG affinity ( $7.6 \times 10^{-4} \pm 2.5 \times 10^{-4} \text{ nM}^{-1}$  for the higher affinity  
307 conformation) was found to be slightly higher than previously published values ( $2.3 \times 10^{-4} \text{ nM}^{-1}$ )  
308 but it is generally within agreement. The ratio of affinities for the two conformations (13.7) was  
309 in good agreement with previously measured values.

310 *Using the in vitro thermodynamic parameters to predict in vivo genetic regulation*

311 The raison d'être for *in vitro* measurements is to inform what is occurring *in vivo*. One of  
312 the central difficulties in using *in vitro* measurements is the lack of a well enough defined *in vivo*  
313 system to directly compare it with. Furthermore, a model is required which can accurately  
314 function in both circumstances and provide useful predictions. We then seek to fully define our  
315 *in vivo* experiment to model it with the *in vitro* determined thermodynamic parameters.

316 We estimate the copy number of our operator reporter plasmid to be  $\sim 20$  copies per cell  
317 (Daber, Sharp, & Lewis, 2009). This then gives us,

$$[O]_{\text{tot}} = \frac{20 \text{ molecules}}{6.02 \times 10^{23} \frac{\text{molecules}}{\text{mole}}} * \frac{1}{1 \times 10^{-15} \text{ L}} * 1 \times 10^9 \frac{\text{nM}}{\text{M}} = 33 \text{ nM} \quad (41)$$

318 The strain of *E. coli* used has the lac genetic switch deleted from the genome; therefore  
319 lac permease is also deleted. It is then assumed that IPTG enters the cell through passive  
320 diffusion and has the same concentration as the media.

321 Figure 5A shows the simulated *in vivo* data (solid blue line) along with experimentally  
322 determined values (blue squares). The model predicts both higher leakiness (2.7% predicted  
323 versus  $1.7 \pm 0.2\%$  observed) and higher maximal induction (80% predicted versus  $61 \pm 5\%$   
324 observed) than is measured *in vivo*. This indicates that there are additional effects not being  
325 accounted for in the *in vitro* data. It has been postulated that non-specific DNA binding of  
326 repressors could play a significant role (Tungtur et al., 2011), however this should have the effect  
327 of decreasing the effective lac repressor concentration since the non-specific DNA will  
328 competitively bind with operator DNA for lac repressor. We see the opposite in our data; the lac  
329 repressor concentration appears higher *in vivo* than we are measuring.

330 There is a known molecular crowding effect in living cells due to the density of  
331 molecules which will increase the *effective* concentration of molecules. We can quickly model  
332 the effect of crowding by decreasing the available space for the lac repressor and estimating its  
333 effective concentration,

$$[R]_{\text{tot}}^{\text{eff}} = \frac{[R]_{\text{tot}}}{\% \text{ available space}} \quad (42)$$

334 Figure 5B shows the effect of including molecular crowding on the predicted *in vivo*  
335 induction curve. The model shows excellent agreement with experiment at a molecular  
336 crowding of 40-60% which estimates effective *in vivo* repressor concentration to be 1.1-1.6  $\mu\text{M}$

337 (Leakiness:  $1.3 \pm 0.3\%$  predicted versus  $1.7 \pm 0.2\%$  observed; Maximal expression:  $67 \pm 4\%$   
338 predicted versus  $61 \pm 5\%$  observed). Furthermore, this value agrees well with estimates of 20%-  
339 40% available space *in vivo* (Kubitschek & Friske, 1986).

340 Since there is a notable deviation in repressor-DNA affinity with previous *in vitro*  
341 measurements, the same analysis was carried out for the three curated data sets from Sharp  
342 (Sharp, 2011). Using the values from the literature, we find that they do not in any case come  
343 close to replicating our *in vivo* data (Fig. 5A, orange dashed, purple dotted line, and solid green  
344 lines). The DNA affinities are much too high for the measured DNA and repressor  
345 concentrations. At these affinities the switch is essentially completely off and cannot be induced  
346 with any concentration of IPTG. Crowding only enhances the deviation from experiment as it  
347 further increases the concentration of repressor.

348 *Simulating native in vivo lac genetic switch phenotype*

349 The thermodynamic constants from our *in vitro* data better represents our *in vivo* model  
350 system. The question then is: which set of thermodynamic parameters could effectively regulate  
351 the native lac genetic switch?

352 Essentially we have rebuilt the lac operon with the *lacZ*, *lacY* and *lacA* polycistronic  
353 message replaced by the reporter gene YFP and the dimeric lac repressor constitutively expressed  
354 by its native promoter. We have a higher copy number of both the reporter and repressor  
355 plasmids (~20 copies per cell) which increases both the operator and repressor concentrations  
356 above that normally found in the cell. A secondary deviation is the removal of the tetramerization  
357 domain and multiple operator DNA sites (O2 and O3 additionally exist on the genome) which  
358 simplifies our analysis. The cooperativity of the native tetrameric lac repressor is known to  
359 decrease leakiness approximately 10-fold, so we might expect a dimeric lac repressor with one  
360 operator (O1) to have some leakiness in its repression (Oehler et al., 1994).

361 As previously mentioned, *in vivo* lac repressor dimer concentration was measured to be  
362 ~40 dimers per cell, which gives,

$$[R]_{\text{tot}} = \frac{40 \text{ molecules}}{6.02 \times 10^{23} \frac{\text{molecules}}{\text{mole}}} * \frac{1}{1 \times 10^{-15} \text{ L}} * 1 \times 10^9 \frac{\text{nM}}{\text{M}} = 66 \text{ nM} \quad (43)$$

363 And we know there is one operator per cell,

$$[O]_{\text{tot}} = \frac{1 \text{ molecules}}{6.02 \times 10^{23} \frac{\text{molecules}}{\text{mole}}} * \frac{1}{1 \times 10^{-15} \text{ L}} * 1 \times 10^9 \frac{\text{nM}}{\text{M}} = 1.7 \text{ nM} \quad (44)$$

364 Using these values, along with the experimentally determined binding constants derived  
365 from this study and those curated by Sharp, we can simulate dimeric lac repressor induction  
366 curves at native conditions. Figure 6A shows that the values determined in this study predict a

367 leaky repressor that is maximally inducible. The much higher DNA affinities of the curated data  
368 sets all produce over-repressed curves that do not show good induction.

369 The over-repression is even more prominent as cell crowding is considered. Using the  
370 value of 40%, which gives  $R_{\text{tot}} = 66 \text{ nM} / 0.4 = 165 \text{ nM}$ , we find that the over-repression of the  
371 high affinity DNA sets all produce curves that weakly induce or do not induce at all (Fig. 6B).  
372 The predicted curve using our thermodynamic parameters again provides reasonable induction  
373 (~10% leakiness up to ~95% maximal induction). While this level of leakiness would be  
374 intolerably high for efficient regulation of the lac operon, the restoration of the tetramerization  
375 domain would significantly decrease the leakiness while minimally impairing inducibility.

376 If we consider the lowest possible concentration of lac repressor (1 molecule/cell;  $R_{\text{tot}} =$   
377  $1.7 \text{ nM}$ ; with 40% crowding  $R_{\text{tot}} = 4.25 \text{ nM}$ ) we find that second curated data set does produce  
378 reasonable induction curves, even if 40% crowding is taken into consideration (Fig 6C and Fig.  
379 6D). Unfortunately, in this regime the binding would be highly stochastic and hence noisy, which  
380 would not produce stable repression. Furthermore, this level of repressor expression does not  
381 agree with published values. While it is technically possible for these affinities to be accurate, it  
382 is highly improbable. The first and third data sets would require less than 1 molecule of dimeric  
383 lac repressor per cell to be functionally useful according to our model.

384 Given the wide range of repressor-operator DNA affinities (100nM – 3333nM) it can be  
385 reasonably concluded that these values must contain significant artifacts from the experimental  
386 techniques. Techniques such as gel shift assays, where molecular “caging” effects are known to  
387 be significant, and nitrocellulose filter binding assays, where the binding is removed from the  
388 solution phase, were used to create the curated data sets. Our measurement of repressor-DNA  
389 binding affinity did require an indirect measurement, namely transcription, but it did occur in the  
390 solution phase. We attribute the difference in values to differences in experimental setup.

## 391 Conclusions

392 We have reproduced the transcriptional regulation of the lac repressor dimer *in vitro* and  
393 shown that it accurately reproduces the *in vivo* repression of YFP under control of the lac  
394 repressor. Accurate modeling of the *in vivo* data required an estimate of 40-60% cellular  
395 crowding in the cell, which agrees with previous estimates. Non-specific DNA binding and IPTG  
396 uptake did not appear to have any significant effect. Crowding could be tested *in vitro* through  
397 crowding agents such as bovine serum albumin (BSA) or polyethylene glycol (PEG) (Ellis,  
398 2001). Alternative explanations are potentially possible such as fluctuations in the size of the *E.*  
399 *coli*. What is essentially important is that the concentration of lac repressor in the cell greatly  
400 affects the maximal induction given our thermodynamic parameters. The curve is extremely  
401 sensitive in that region to changes in repressor concentration. So only an approximately two-fold  
402 increase in repressor concentration is sufficient to replicate the *in vivo* data. Whether the lac  
403 repressor concentration is increased by molecular crowding or by decreased *E. coli* volume  
404 would have to be tested by further experiments.

405 The measured thermodynamic binding parameters match well for IPTG binding and  
406 conformational equilibrium, except there is significantly lower repressor/operator DNA affinity  
407 measured (by approximately 3-4 orders of magnitude). This discrepancy was modeled and it was

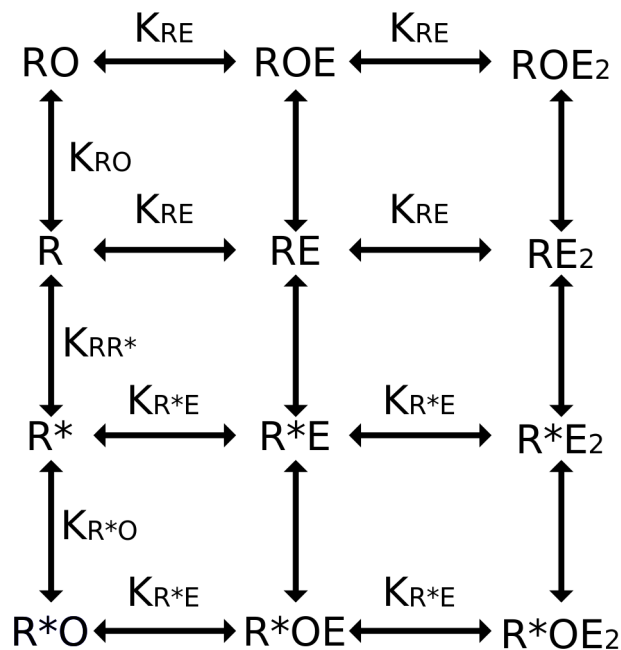


408 demonstrated that the affinity measured in this study is capable of reproducing not only the *in*  
409 *vivo* data from this study, but also can predict reasonable induction curves at concentrations of  
410 repressor and DNA that are naturally seen by *E. coli*. We therefore conclude that lac repressor  
411 DNA affinity is significantly weaker than previous *in vitro* measures and more in line with the  
412 estimates for repressor-DNA affinity at *in vivo* conditions where we do find good agreement with  
413 previously published values.

414 Finally, this study highlights the difficulty in using *in vitro* data generated from  
415 experimental techniques that are divorced from conditions closer to that of the cell. Experimental  
416 artifacts may greatly overshadow actual values, which should come as no surprise in the case of  
417 lac repressor binding to operator DNA where the published binding constant has changed 33-fold  
418 as experimental techniques have changed. The difficulty in *in vitro* measurements is well known  
419 in the field as is evidenced by the large consideration given to differences in buffer conditions  
420 (Ha et al., 1992), DNA length (Khoury et al., 1990), and even hydrostatic pressure (Royer,  
421 Chakerian, & Matthews, 1990). Techniques such as gel filtration or nitrocellulose filter binding  
422 assays are excellent at differentiating binding strength between point mutants; they are limited in  
423 comparison with *in vivo* results. Using experimental setups which more closely mimic the *in vivo*  
424 system can significantly improve the ability of the predictive capabilities of *in vitro* experiments.  
425 They do come with the caveat that the data interpretation is not as straightforward as simple  
426 binding experiments.

#### 427 **Acknowledgements**

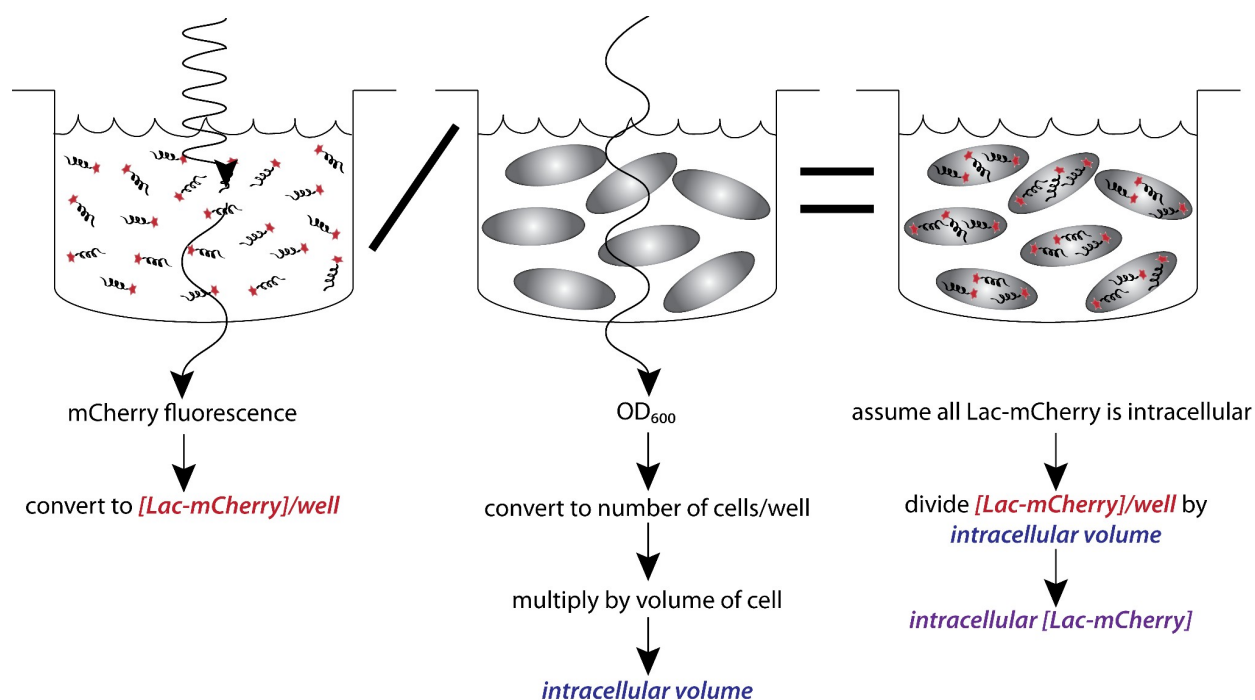
428 We would like to thank Dr. Mark Goulian for the kind gift of the YFP and mCherry genes  
429 along with the EPB229 cell line. We would like to thank Dr. Kristin Lynch, Sandya Ajith, Nicole  
430 Martinez, Chris Yarosh and Michael Mallory for guidance and assistance performing the *in vitro*  
431 transcription experiments. We would also like to thank Elizabeth Sweeny for assistance in  
432 making figures.



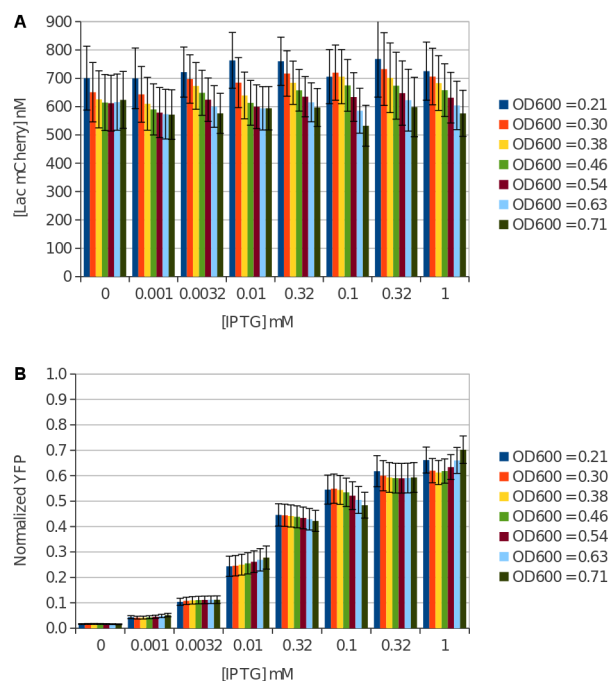
434 **Figure 1. Monod, Wyman, and Changeux (MWC) model of thermodynamic equilibrium.**

435 This model identifies two primary structural conformations of the lac repressor (R and R\*): the R  
 436 state has high operator DNA (O) affinity and the R\* state has low operator DNA affinity.

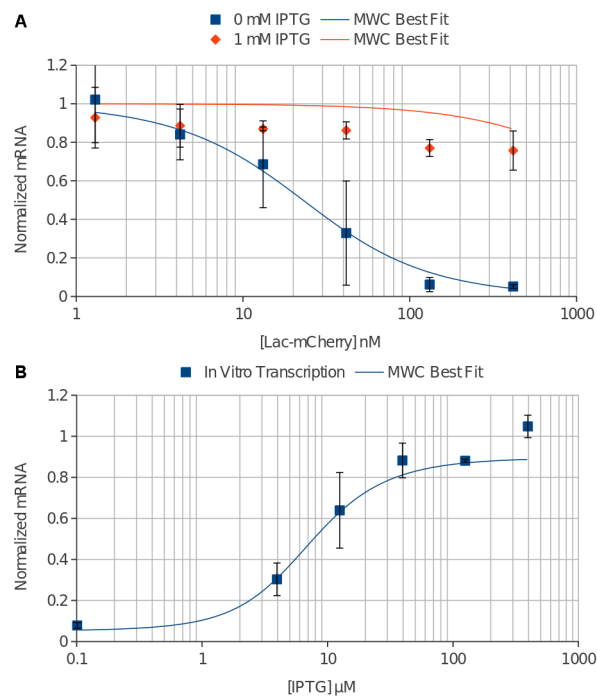
437 Addition of effector (E) alters the effective equilibrium between the two states allowing for an  
 438 increase or decrease in amount of operator DNA bound. Fraction of bound operator is considered  
 439 a proxy for transcription; unbound operator can be freely transcribed. Thermodynamic binding  
 440 and conformational equilibrium constants are fully defined in the methods.



442 **Figure 2. Measuring intracellular Lac-mCherry concentration.** Raw mCherry fluorescence  
 443 and OD<sub>600</sub> are measured on a plate reader. Calibration curves for both were established given our  
 444 experimental setup (cell line, plasmids, media, amount of sample loaded, plates and plate reader).  
 445 Raw fluorescent signal is converted to concentration of Lac-mCherry per well. Raw OD<sub>600</sub> signal  
 446 is converted to the fraction of well volume that is intracellular. Dividing Lac-mCherry well  
 447 concentration by intracellular volume fraction effectively concentrates the Lac-mCherry to be  
 448 intracellular. These two measurements, combined with the appropriate calibrations, allow a quick  
 449 and accurate measurement of intracellular Lac-mCherry concentration.



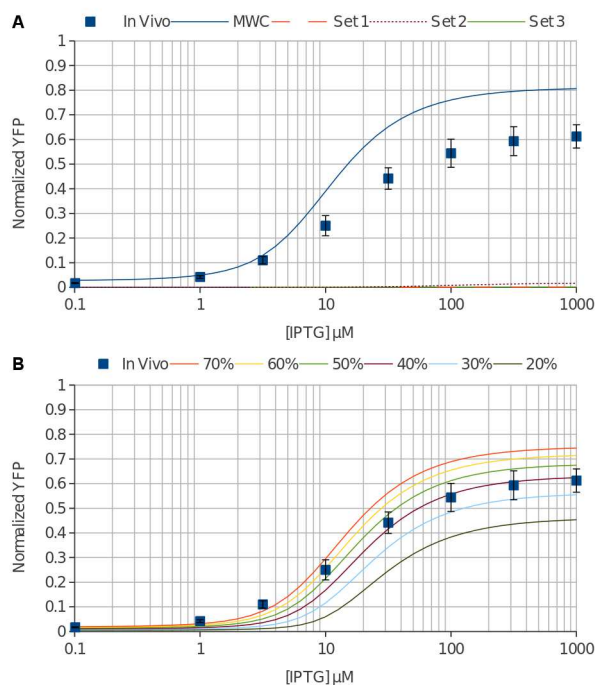
451 **Figure 3. *In vivo* Lac-mCherry and YFP regulation show no growth dependence.** (A) Lac-  
 452 mCherry was calculated for growing *E. coli* cells and found to have minimal OD<sub>600</sub> dependence.  
 453 As expected for a constitutively expressed gene, there is no change in Lac-mCherry  
 454 concentration with increasing IPTG concentration. (B) Normalized YFP was simultaneously  
 455 measured and again no OD<sub>600</sub> dependence was found throughout the exponential growth phase. In  
 456 stark contrast to the Lac-mCherry concentration, a distinct induction profile is measured for YFP  
 457 as a function of IPTG.



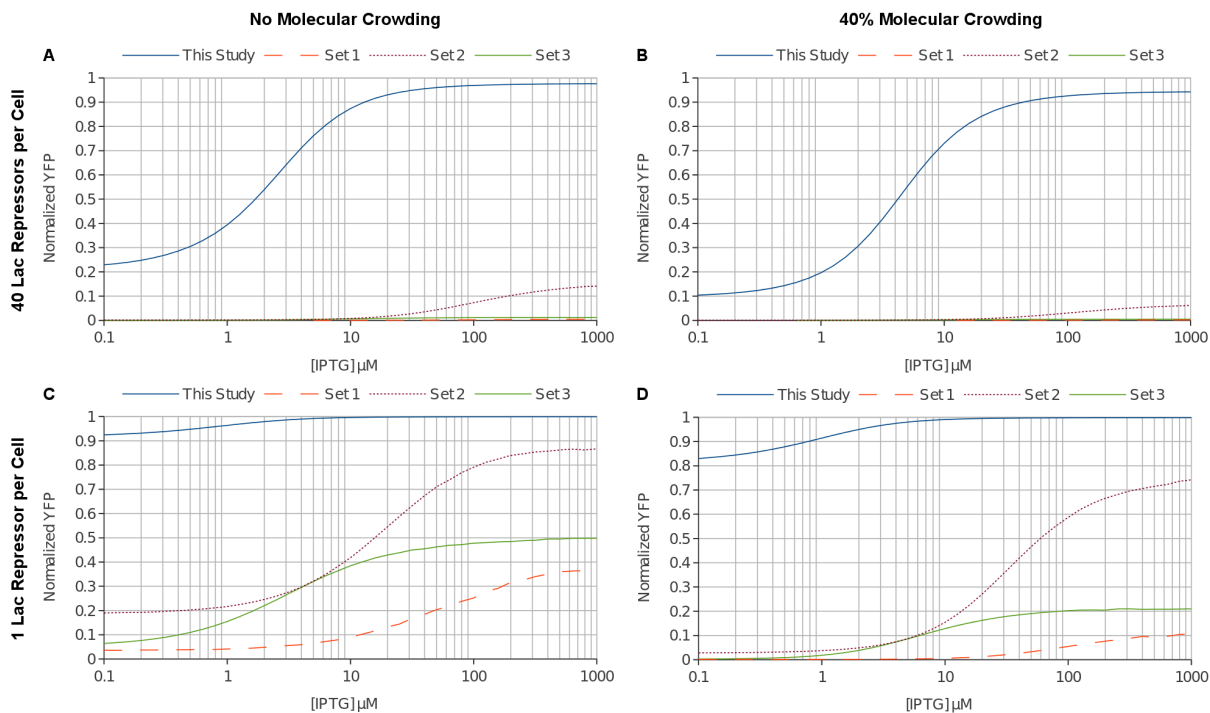
459 **Figure 4. *In vitro* transcription controlled by the lac repressor is accurately fit by the MWC**  
 460 **model. (A)** Lac-mCherry was added at varying concentrations with 11nM O1 DNA and mRNA  
 461 was quantitated (blue squares). The repression was relieved upon addition of 1mM IPTG (orange  
 462 diamonds). The data was globally fit by the MWC model and an accurate solution was found for  
 463 the Lac-mCherry titration (solid blue line). The model predicts higher induction than was  
 464 measured experimentally (solid orange line). **(B)** IPTG was added at varying concentrations with  
 465 333nM Lac-mCherry and 11nM O1 DNA and again mRNA was quantitated (blue squares). A  
 466 robust induction profile was measured showing induction up to approximately 80% of  
 467 constitutive expression. The global fit also accurately fits the IPTG titration data (solid blue line).

	This study	Daber, Sharp, and Lewis†	Daber, Sochor, and Lewis‡	Sharp, Set 1 ‡	Sharp, Set 2 ‡	Sharp, Set 3 ‡	Müller-Hill §
$K_{RR^*} = [R^*]/[R]$	$6.3 \pm 3.4$	$2 \pm 0.5$	$5.8 \pm 0.07$				
$K_{RO}$ (nM <sup>-1</sup> )	$0.42 \pm 0.21$			3330	100	1510	1
$K_{RE}$ (nM <sup>-1</sup> )	$5.6 \times 10^{-5} \pm 1.8 \times 10^{-5}$		$6 \times 10^{-5} \pm 2 \times 10^{-7}$				
$K_{R^*E}$ (nM <sup>-1</sup> )	$7.6 \times 10^{-4} \pm 2.5 \times 10^{-4}$		$5 \times 10^{-4} \pm 5 \times 10^{-6}$	$2.3 \times 10^{-4}$	$2.3 \times 10^{-4}$	$2.3 \times 10^{-4}$	
$K_{R^*O}$ (nM <sup>-1</sup> )	$1.0 \times 10^{-10}$						
$R_{tot}$ (nM) [with 40% crowding]	$664 \pm 90$ [1660 + 225]						
$r = K_{RO} * R_{tot}$ [with 40% crowding]	278 [697]	$150 \pm 50$	$150 \pm 50$				
$X = K_{R^*E} / K_{RE}$	$13.7 \pm 0.13$	$15 \pm 3$	8.28				

469 **Table 1. Fit values from the MWC models compared with literature values.** All fit  
 470 parameters agree with the exception of repressor-operator DNA affinity ( $K_{RO}$ ). †(Daber, Sharp, &  
 471 Lewis, 2009), ‡(Daber, Sochor, & Lewis, 2011), ‡(Sharp, 2011), §(Müller-Hill, 1996).



473 **Figure 5. *In vivo* regulation by the lac repressor is accurately predicted with molecular**  
 474 **crowding. (A)** YFP under control of the lac repressor was measured in *E. coli* cells at varying  
 475 concentrations of IPTG (blue squares). We used the measured intracellular concentration of the  
 476 lac repressor (660nM) and the fit values from *in vitro* transcription to predict the *in vivo*  
 477 induction curve with the MWC model (solid blue line). The model predicts more YFP signal at  
 478 all concentrations of IPTG. Our repressor-DNA affinity was much lower than previously  
 479 published values, so we also modeled three curated data sets (Sharp, 2011) (dashed orange,  
 480 dotted purple, and solid green lines). All three predict greatly over-repressed YFP expression and  
 481 do not fit the *in vivo* data. **(B)** Molecular crowding is known to play a significant role in cells. We  
 482 modeled this by estimating the available volume in percentage for our repressor and calculated  
 483 an effective repressor concentration. We modeled several percentages and 40-60% available  
 484 volume (solid purple, green and yellow lines) accurately reproduces the *in vivo* regulation from  
 485 the *in vitro* transcription derived thermodynamic constants. 40% crowding corresponds to an  
 486 effective repressor concentration of 1.6 $\mu\text{M}$ .



488 **Figure 6. Simulating a simplified lac operon from *in vitro* derived thermodynamic**  
 489 **constants.** The correct repressor-DNA affinities must be able to provide robust switching under  
 490 conditions naturally experienced by *E. coli*. With this in mind, we modeled a dimeric lac  
 491 repressor regulating a gene with a single operator sequence. **(A)** The natural lac promoter makes  
 492  $\sim 66\text{nM}$  of lac repressor dimer and one operator is at  $\sim 1.7\text{nM}$  in the cell. We modeled these  
 493 conditions for the thermodynamic parameters from this study and for the three curated data sets  
 494 of Sharp. The predicted curve from this study shows a reasonable repression and induction  
 495 profile (solid blue line). Only Set 2 from Sharp is weakly inducible (dotted purple line). **(B)**  
 496 Including molecular crowding (40% available volume) enhances the situation. The curated data  
 497 sets do not make useful switches. Alternately, the predicted induction curve from *in vitro*  
 498 transcription derived constants shows a leaky switch that induces very well (solid blue line). **(C)**  
 499 We next sought to model the minimal possible repressor to find a condition where the curated  
 500 data sets produce reasonable induction curves. 1 molecule of dimer per cell ( $\sim 1.7\text{nM}$ ) does show  
 501 good induction profiles for set 2 (dotted purple line) and set 3 (solid green line). Set 1 still shows  
 502 a switch that can marginally be induced and would likely not be useful (dashed orange line). **(D)**  
 503 Molecular crowding effects again enhance the repressor concentration and only set 2 could  
 504 reasonably regulate a gene (dotted purple line). The values from this study (solid blue line)  
 505 predict a very leaky switch. Although the second curated set could effectively regulate the gene  
 506 at this concentration, in reality a single dimer and single operator DNA binding would be  
 507 dominated by stochastic events creating an inherently unstable switch.



508 **Bibliography**

509

510 Daber, R., & Lewis, M. 2009. *Towards evolving a better repressor*. Protein engineering, design &  
511 selection : PEDS 22(11):673–83.

512 Daber, R., Sharp, K. A., & Lewis, M. 2009. *One is not enough*. Journal of molecular biology  
513 392(5):1133–1144.

514 Daber, R., Sochor, M. A., & Lewis, M. 2011. *Thermodynamic analysis of mutant lac repressors*.  
515 Journal of molecular biology 409(1):76–87.

516 Ellis, R. J. 2001. *Macromolecular crowding : an important but neglected aspect of the*  
517 *intracellular environment*. Current Opinion in Structural Biology 11:114–119.

518 Ha, J. H., Capp, M. W., Hohenwalter, M. D., Baskerville, M., & Record, M. T. 1992.  
519 *Thermodynamic stoichiometries of participation of water, cations and anions in specific and*  
520 *non-specific binding of lac repressor to DNA. Possible thermodynamic origins of the “glutamate*  
521 *effect” on protein-DNA interactions*. Journal of molecular biology 228(1):252–64.

522 Khoury, a M., Lee, H. J., Lillis, M., & Lu, P. 1990. *Lac repressor-operator interaction: DNA*  
523 *length dependence*. Biochimica et biophysica acta 1087(1):55–60.

524 Kubitschek, H. E., & Friske, J. A. 1986. *Determination of bacterial cell volume with the Coulter*  
525 *Counter*. Journal of bacteriology 168(3):1466–7.

526 Lau, I. F., Filipe, S. R., Søballe, B., Økstad, O.-A., Barre, F.-X., & Sherratt, D. J. 2004. *Spatial*  
527 *and temporal organization of replicating Escherichia coli chromosomes*. Molecular  
528 Microbiology 49(3):731–743.

529 Lewis, M. 2005. *The lac repressor*. Comptes rendus biologies 328(6):521–48.

530 Monod, J., Wyman, J., & Changeux, J.-P. 1965. *On the nature of allosteric transitions: A*  
531 *plausible model*. Journal of Molecular Biology 12(1):88–118.

532 Müller-Hill, B. 1996. *The Lac Operon: A Short History Of A Genetic Paradigm* (p. 135).

533 Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B., & Müller-Hill, B. 1994.  
534 *Quality and position of the three lac operators of E. coli define efficiency of repression*. The  
535 EMBO journal 13(14):3348–3355.

536 Poelwijk, F. J., de Vos, M. G. J., & Tans, S. J. 2011. *Tradeoffs and optimality in the evolution of*  
537 *gene regulation*. Cell 146(3):462–70.

538 Royer, C. a, Chakerian, a E., & Matthews, K. S. 1990. *Macromolecular binding equilibria in the*  
539 *lac repressor system: studies using high-pressure fluorescence spectroscopy*. Biochemistry  
540 29(20):4959–66.

541 Sharp, K. A. 2011. *Allostery in the lac operon: population selection or induced dissociation?*  
542 Biophysical chemistry 159(1):66–72.

543 Swint-Kruse, L., & Matthews, K. S. 2009. *Allostery in the LacI/GalR family: variations on a*  
544 *theme*. *Current opinion in microbiology* 12(2):129–37.

545 Tungtur, S., Skinner, H., Zhan, H., Swint-Kruse, L., & Beckett, D. 2011. *In vivo tests of*  
546 *thermodynamic models of transcription repressor function*. *Biophysical chemistry* 159(1):142–  
547 51.

548

549