Supporting Information

Anderson and Yang 10.1073/pnas.0705570105

Supporting Text

Cloning. All clones were first constructed in the high copy number plasmid pLOI2403 (obtained from L. Ingram) (1). The region encompassing *cro-*O_R*-cI* was amplified with the primers SalI-cro-R and XbaI-cI-L, using N99 cells lysogenic for λ PaPa (obtained from R. Calendar) as template. The *rexA-rexB-*O_L region was amplified from purified λ DNA (Fermentas) with the primers EcoRI-rexA-L and NheI-OL-L. The Venus fluorescent protein (obtained from the YRC Microscopy, University of Washington, Seattle, WA) was amplified by the primers VspI-Venus-F and EcoRI-Venus-R and ligated into *Nhe*I- and *Eco*RI-digested pET21 (Novagen) to make pET21Venus, which places the *venus* gene directly after a strong rbs. The *Sal*I-*cro-*O_R*-cI-Xba*I and *Eco*RI-*rexA-rexB-*O_L*-Nhe*I PCR products and the *Xba*I-*rbs-venus-Eco*RI fragment of pET21Venus were simultaneously ligated into *Nhe*I- and *Sal*I-digested pLOI2403 to make pLOIVenusWT. Mutations in O_L and O_R were introduced using quickchange mutagenesis. For the O_L3-4 mutation (2), two rounds of mutagenesis were used to introduce the four base pair changes. For the first round, QC-OL3A-F and QC-OL3A-R were used, followed by QC-OL3B-F and QC-OL3B-R. The O_R3-r1 mutation (3) was made with OR3-r1-F and OR3-r1-R. For the noO_L mutant, the primers QC-NheI-noOL-F and QC-NheI-noOL-R were used to add a second *Nhe*I site (italicized) after the t_{imm} terminator, immediately before O_L3. After mutagenesis, the plasmid was digested with NheI and religated to remove O_L. Plasmids containing mutations in both O_L and O_R were constructed by subcloning. Each high-copy clone was verified by sequencing, then the *Nhe*I/*Sal*I fragment was subcloned into the low-copy-number plasmid pKLJ12 (4). For testing the effect of a loss of the t_{imm} terminator, a PCR-based mutagenesis method was used (5) with the primers QC-timm-F and QC-timm-R. Site-directed mutagenesis with the primers QC-truncCI-F and QC-truncCI-R was used to construct a control plasmid that does not produce functional CI, which we used as a control for quantitative PCR. All primer sequences are listed below. Restriction sites and operators are underlined. Mutated bases are in bold.

Primer Sequences. SalI-cro-R, GTGGTCGACCGGGGTTATTTATGCTGTTG; XbaI-cI-L, TAATTCTAGACATTTACTATGTTATGTTCTGAGGG; EcoRI-rexA-L, GGGAATTCGGCAAGGTGTTCTGGTCGG; NheI-OL-L, CACACAGCTAGCAGCCTTCTGCTTTGAATGCTGCC; VspI-Venus-F, GCATTAATGAGTAAAGGAGAAGAACTTTTCACTGG; EcoRI-Venus-R, CGATGAATTCCTATTTGTATAGTTCATCCATGCC; QC-OL3A-F, GCCAGAGATAATTTATCACCG**G**AG**T**TGGTTATCTGTATG; QC-OL3A-R, CATACAGATAACCAA**C**T**C**CGGTGATAAATTATCTCTGGC; QC-OL3B-F, CCGCCAGAGATAATTTATCAC**TA**GAGTTGGTTATCTG; QC-OL3B-R, CAGATAACCAACTC**TA**GTGATAAATTATCTCTGGCGG; OR3-r1-F, GCTCATACGTTAAATCTATCAC**TG**CAAGGGATAAATATCTAACACC; OR3-r1-R, GGTGTTAGATATTTATCCCTTGCAGTGATAGATTTAACGTATGAGC; QC-NheI-noOL-F, TAATTTATCACCGCAGATGGCTAGCTGTATGTTTTTTATATGA; QC-NheI-noOL-R, TCATATAAAAAACATACAGCTAGCCATCTGCGGTGATAAATTA; QC-timm-F, GAATAAC**GGTTA**CTCAGCGCCAGG; QC-timm-R, CTGAG**TAAGG**GTTATTCTTGTTCTCTGG; QC-truncCI-F, CAATCGCCAGAGAAATCTA**GA**AGATGTATGAAGCGGTTAG; and QC-truncCI-R, CTAACCGCTTCATACATCT**TC**TAGATTTCTCTGGCGATTG.

Flow Cytometry Detector Settings. FL1 was measured at maximum PMT voltage (9.99) and amplifier gain (9.99). FSC was collected at setting of E02 to amplify the signal 100-fold. FSC and SSC intensity for the cells were near the detection limit, so thresholds were set to the minimum between the peaks representing noise and *Escherichia coli* cells in the raw histograms. Fluorescence histograms for the low intensity calibration beads B and C (from the LinearFlow green flow cytometry low intensity calibration kit, 2.5 μm; Molecular Probes) were collected at the maximum PMT voltage and amp gain for the FL1 channel. The amp gain was reduced from 9.99 to 1.00 to collect additional FL1 data for beads C and D. Datasets of 10,000 beads were collected, and, before fitting, bead multimers were removed by filtering out data points that differed from the mean FSC and SSC by >20%.

Distribution Widths and Variation in Cell Size. We examined the effect of variation in cellular fluorescence due to growth-related changes by selecting the time point that had the smallest differences between the five samples for the FSC and SSC histograms, computed as the standard deviation for the cumulative sum of the probability histograms. Histograms from the ≈200-min time point (OD_{600} = 0.3) showed the least variation between samples. The histograms of forward scatter, a measure of cell volume, were nearly identical for the five samples (Fig. S9). These histograms include contributions from cells of various sizes, ranging from small newly divided cells to large cells in the final stages of cell division. To select a narrow range of cell volume for the fluorescence histograms, we recalculated the histograms, using only cells that fall within 10% of the mean FSC value. Depending on the size and shape of the cells, the range of angles they present to the laser beam, and the geometry of the cytometer optics and detectors, the volume of *E. coli* cells typically are proportional to $FSC^{0.75}$ to $FSC^{0.5}$ (6), so the set of ≈17,000 cells should differ in volume by 5 to 7.5%. These histograms are more symmetric, but are broadened due to instrument response. This was deconvolved out by using calibration data from standard green fluorescent beads.

Deconvolution. To calibrate instrument response, we used the LinearFlow green flow cytometry low intensity calibration kit, 2.5 μm (Molecular Probes). The fluorescence histogram for each calibration bead sample was fit to a truncated Gaussian to determine the mean and the width of each fluorescence intensity distribution. The response function of the instrument was then calculated empirically as log(width) = 0.547 log(mean) + 1.03. This response function was used to calculate an instrument response matrix over the range of fluorescence intensity values present in the dataset. For deconvolution, the experimental data were assumed to take the form of a gamma distribution, $f(x;k,\theta) = x^{k-1}(e^{-x/\theta})/(\theta^k\,\Gamma(k))$. For each sample, the shape (k) and scale (θ) parameters were adjusted to minimize the difference between the raw cytometry histogram and the convolution of the instrument response matrix and the gamma distribution. Gamma distributions generated from the minimized parameters provided a good approximation to the

deconvolved fluorescence data (Fig. S10). Parameters for the gamma fit for the dataset collected at the ≈200-min time point, with the histogram filtered by FSC intensity, are as shown in Table S1.

Comparison of Unfiltered Histograms and Deconvolved Data. For our data analysis, we used the mean fluorescence of each mutant construct normalized by the WT construct mean. To determine how sensitive these values are to the method of determining the mean, we compared values calculated from a deconvolved dataset representing cell volumes that differ by only 5–7.5% from the time point that showed the least variation in FSC and SSC to raw histograms collected at various time points between 3 and 5 h after inoculation. The normalized values are not significantly different for the two very different approaches (uncertanties are one standard deviation—68% confidence interval—and n indicates the number of time points in the dataset) (see Table S2).

mRNA Production and CI Concentration. To arrive at Eq. **2** in the main text, we consider the production and degradation of protein and mRNA averaged over a cell division period. We use the arguments put forward by Guido *et al.* (7). Suppose $\bar{m}(t)$ and $\bar{n}_{\mathrm{CI}}(t)$ are the mean numbers of mRNAs and CI proteins at time t. Then the time-dependent changes in them between cell divisions can be expressed by

$$\frac{d\bar{m}(t)}{dt} = \gamma_m P_{\mathrm{activated}} - \delta_m \bar{m}(t) \qquad [1]$$

$$\frac{d\bar{n}_{\mathrm{CI}}(t)}{dt} = \gamma_{\mathrm{CI}}\, \bar{m}(t) - \delta_{CI}\, \bar{n}_{\mathrm{CI}}(t) \qquad [2]$$

where γ_{m} is the activated transcription rate and has a unit of number·s^{-1}, $P_{\mathrm{activated}}$ is the unitless probability that the promoter is activated, δ_{m} is the rate constant for mRNA degradation in s^{-1}, γ_{CI} is the protein synthesis rate constant, and δ_{CI} is the rate constant for CI degradation. Both the quantities in $\bar{m}(t)$ and $\bar{n}_{\mathrm{CI}}(t)$ experience a discontinuous decrease by one-half when the cell divides. Taking this into account, the solution to Eq. **1** is

$$\bar{m}(t) = \left(\bar{m}^* - \frac{\gamma_m P_{\mathrm{activated}}}{\delta_m}\right) e^{-\delta_m t} + \frac{\gamma_m P_{\mathrm{activated}}}{\delta_m},$$

where

$$\bar{m}^* = \frac{\gamma_m P_{\mathrm{activated}}}{\delta_m} \frac{1 - e^{-\delta_m}}{2 - e^{-\delta_m}}$$

is the stationary asymptotic number of mRNA's immediately after a cell division. Replacing the above solution into Eq. **2** and solving the equation, the mean number of CI per cell after averaging over one division cycle is given by,

$$\bar{n}_{\mathrm{CI}} = \gamma_{\mathrm{CI}} \gamma_m \Phi P_{\mathrm{activated}},$$

where

$$\Phi = \frac{\delta_m(e^{-\delta_{\mathrm{CI}}} - 1) + \delta_{\mathrm{CI}}\left(1 - e^{-\delta_m} + \left(\frac{1 - e^{-\delta_m}}{2 - e^{-\delta_m}}\right)(e^{-\delta_m} - e^{-\delta_{\mathrm{CI}}})\right)}{\delta_{\mathrm{CI}}\delta_m(\delta_{\mathrm{CI}} - \delta_m)(2 - e^{-\delta_{\mathrm{CI}}})} \frac{(1 - e^{\delta_m})}{\delta_m}$$
$$+ \frac{\delta_m\left(\frac{1 - e^{\delta_m}}{\delta_m} - 1\right) + \delta_{\mathrm{CI}}\left(1 - \frac{1(-e^{\delta_{\mathrm{CI}}})}{\delta_{\mathrm{CI}}} + \left(\frac{1 - e^{-\delta_m}}{2 - e^{-\delta_m}}\right)\left(\frac{(1 - e^{\delta_{\mathrm{CI}}})}{\delta_{\mathrm{CI}}} - \frac{(1 - e^{\delta_m})}{\delta_m}\right)\right)}{\delta_{\mathrm{CI}}\delta_m(\delta_{\mathrm{CI}} - \delta_m)}$$

The concentration of CI in a cell is then

$$[\mathrm{CI}] = \frac{\gamma_{\mathrm{CI}}\gamma_m\Phi}{\bar{v}} P_{\mathrm{activated}},$$

where $\bar{v}$ is the volume per cell. If two additional activation states are also possible, with relative activation levels $A2$ and $A3$, then we have

$$[\mathrm{CI}] = \frac{\gamma_{\mathrm{CI}}\gamma_m\Phi}{\bar{v}} [P(\mathrm{act1}) + A_2 P(\mathrm{act2}) + A_3 P(\mathrm{act3})]. \qquad [3]$$

The above equation can be written as,

$$[\mathrm{CI}] = B[P(\mathrm{act1}) + A_2 P(\mathrm{act2}) + A_3 P(\mathrm{act3})], \qquad [4]$$

where

$$B = \frac{\gamma_{\mathrm{CI}}\gamma_m\Phi}{\bar{v}}.$$

Therefore, B is a proportionality constant and takes into account the basal rate of mRNA production, the degradation and dilution (due to cell division) of mRNA and protein, the protein production, and cellular volume.

Quantitative-PCR Analysis. We used quantitative PCR to compare transcript levels of *cI* and *cro* in our constructs to levels in λ lysogens. In addition to the five samples used in our experiments, we included four controls. K12λ^-F^- carrying the pKLJ12 with no insert, and thus no immunity region, was the negative control for both transcripts. N99λ^+ was the lambda lysogen that was used as template for PCR reactions to make the constructs. K12 λ^+F^+ the same strain used as the host for the pKLJ12 plasmids, except that it is a lambda lysogen and carries a full-sized WT F-plasmid. We also used a construct identical to the pKLJ12 Venus WT plasmid except that it contains a mutation that truncates CI so that it cannot bind cooperatively to DNA (primers used for mutagenesis are included in the table in the 'cloning' section above). This construct contains a functional Venus gene, but in flow cytometry experiments the fluorescence level of this strain is below the detection limit. We grew samples in EZRDM exactly as for cytometry and harvested them at OD_{600} ≈0.3. Doubling times were as follows: K12λ^-F^- containing various pKLJ12 constructs ≈27 min, N99λ^+ ≈33 min, K12λ^+F^+ ≈23 min. We used RNAprotect Bacteria Reagent (Qiagen) to stabilize mRNA at the time of harvest and then froze them at −80°C before shipping on dry ice to ACGT (Wheeling, IL) for analysis.

At ACGT, the total RNA was extracted and purified from the samples with the Qiagen RNAeasy purification kit. The samples were treated with RNase-free DNase (Qiagen) to eliminate plasmid and genomic DNA and with Proteinase K (Qiagen). The total RNA was resuspended in nuclease-free water. The quantity and quality of the total extracted and purified from each sample was evaluated by measuring its absorbance at 230 nm, 260 nm, and 280 nm with a NanoDrop 1000-ND Spectrophotometer and by analyzing it on an agarose gel. The total RNA was used to generate cDNA, using SuperScript III First Strand Synthesis kit (Invitrogen) with random hexamers (Invitrogen), and this product was used for the real-time PCR analysis. PCR primers and probes for the TaqMan assay were designed by using Primer Express, Version 2.0.0 (Applied Biosystems). Primer sequences are as shown in Table S3.

The cDNA was diluted 1:20 for the analysis of all of the genes except for the 16S RNA, where it was diluted 1:100. For real-time PCR analysis, diluted cDNA for each sample was PCR amplified in duplicate, using the TaqMan Gene Expression Master Mix (Applied Biosystems) in a thermocycler at 50°C for 2 min., 95°C for 10 min. and 40 cycles of 95°C for 15 s, and 60°C for 1 min. The 7900HT Sequence Detection System (Applied Biosystems) was used for analysis. Each assay included water and RNA only negative controls. Instrument software was used to determine the Ct value for each gene, which was averaged and normalized to the 16S RNA, assuming ≈50,000 ribosomes per cell (8). Estimates of the number of *cro* and *cI* transcripts per cell, which are averages of two runs (a total of four PCR reactions per sample), are shown in Table S4.

The pKLJ12 sample does not contain *cro* or *cI* genes, so it sets the threshold for meaningful results. The results are less quantitative than the fluorescence histograms measured with flow cytometry, but they do allow us to compare the *cI* transcript levels in our constructs with transcript levels in lysogenic cells. The estimated *cI* transcript level obtained for our WT construct is indeed similar to that of lysogenic *E. coli* cells (N99 λ^+ and K12 λ^+ F^+). The estimated *cro* transcript levels for our constructs are also quite low, confirming that P_R is substantially repressed. The truncCI construct has very low *cI* transcript levels, but fairly high *cro* transcript levels, indicating that P_{RM} is repressed and P_R is active.

1. Martinez-Morales F, Borges AC, Martinez K, Shanmugam KT, Ingram LO (1999) Chromosomal integration of heterologous DNA in it *Escherichia coli* with precise removal of markers and replicons used during construction. *J Bacteriol* 181:7143–7148.
2. Dodd IB, Shearwin KE, Perkins AJ, Burr T, Hochschild A, Egan JB (2004) Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev* 18:344–354.
3. Dodd IB, Perkins AJ, Tsemitsidis D, Egan JB (2001) Octamerization of lambda CI repressor is needed for effective repression of P-RM and efficient switching from lysogeny. *Genes Dev* 15:3013–3022.
4. Jones KL, Keasling JD (1998) Construction and characterization of F plasmid-based expression vectors. *Biotechnol Bioeng* 59:659–665.
5. Zheng L, Baumann U, Reymond JL (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* 32:e115.
6. Koch AL, Robertson BR, Button DK (1996) Deduction of the cell volume and mass from forward scatter intensity of bacteria analyzed by flow cytometry. *J Microbiol Methods* 27:49–61.
7. Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ (2006) A bottom-up approach to gene regulation. *Nature* 439:856–860.
8. Bremer H, Dennis PP (1987) *Escherichia coli and Salmonella typhimurium*, ed Neidhardt F (American Society for Microbiology, Washington, DC) Vol 2, pp 1527–1542.
9. Dodd IB, Perkins AJ, Tsemitsidis D, Egan JB (2001) Octamerization of lambda CI repressor is needed for effective repression of P-RM and efficient switching from lysogeny. *Genes Dev* 15:3013–3022.

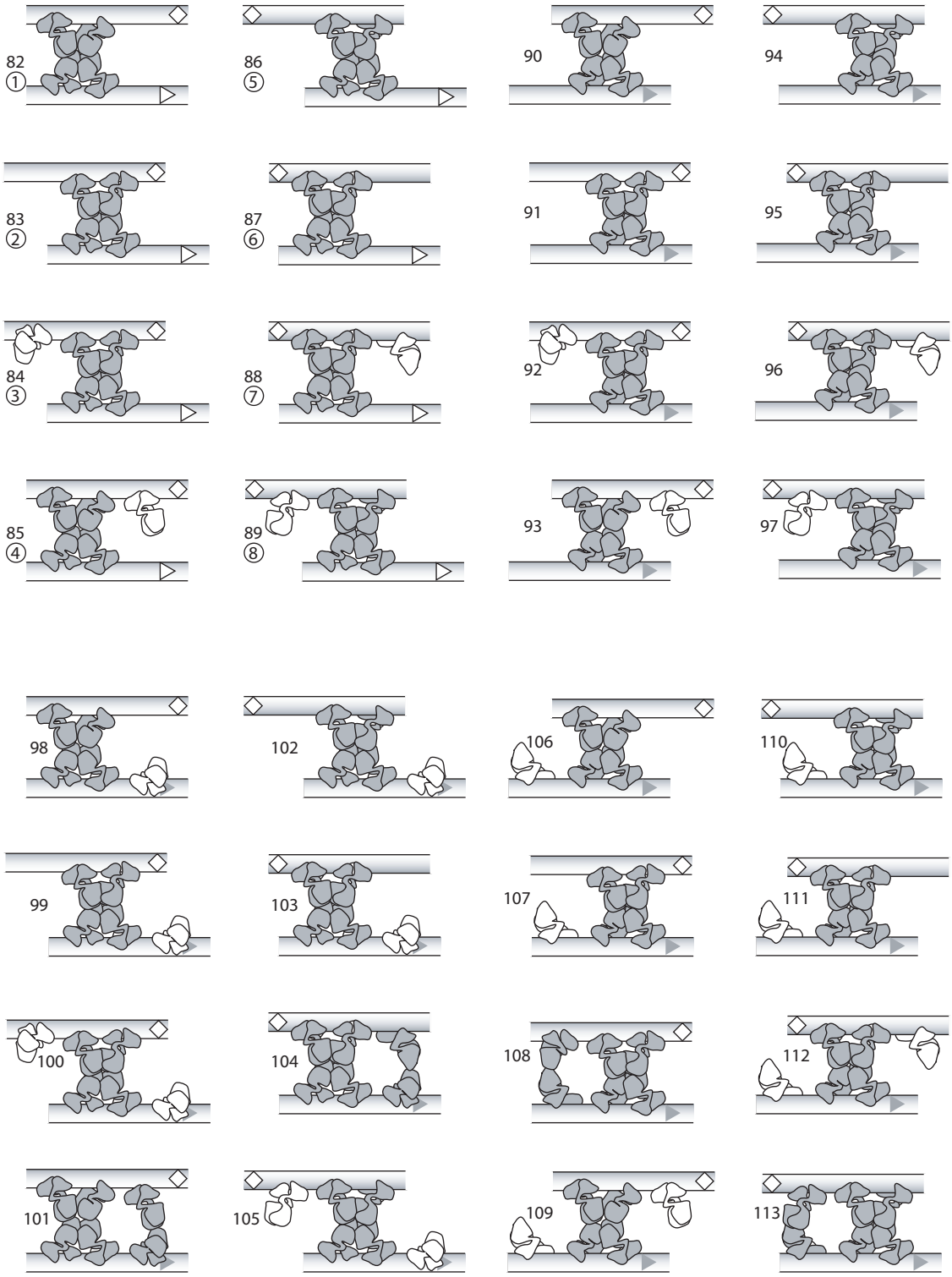

Fig. S1. Looped operator configurations. The set of all looped configurations examined in this study. The UP element adjacent to OL3 is indicated by a white diamond, PRM is shown as a triangle (white if activated, gray if repressed). The first eight configurations have OR2 occupied and OR3 free, and are considered activated to some degree. All others have OR3 occupied and are thus repressed. CI dimers involved in additional cooperative interactions are in gray. Unlooped configurations (Dataset S1) are numbered 1–81, so, to avoid confusion, the numbering for looped configurations begins with 82. The circled numbers correspond to the numbering system used in Fig. 4.

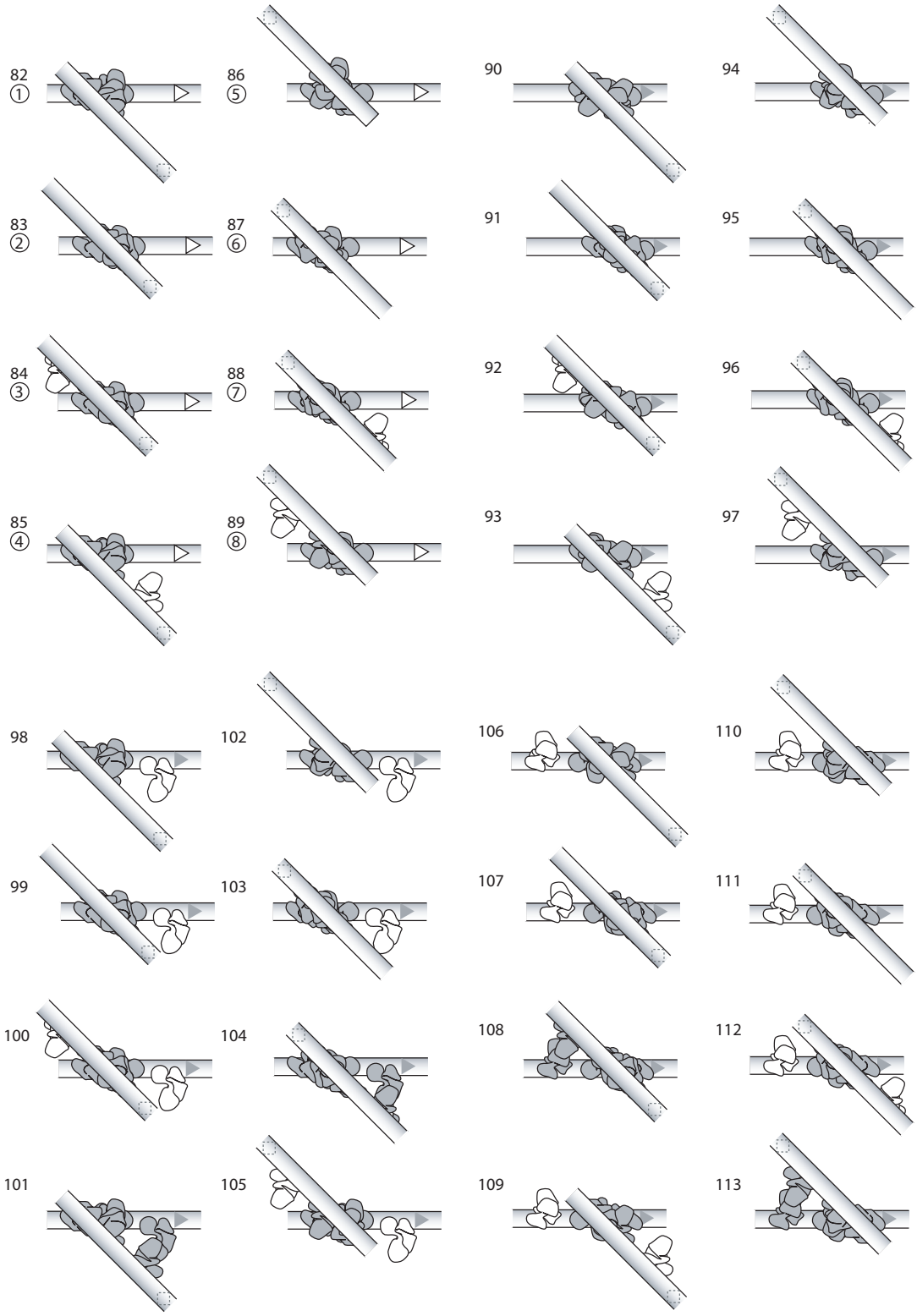

Fig. S2. Alternate view of looped operator configurations. The same configurations of Fig. S3, showing another possible way the CI dimers could interact, resulting in DNA strands that are skew rather than parallel. For each loop, the strand in the foreground represents OL, with the UP element on the bottom face indicated by a gray diamond, and the strand in the background represents OR, with PRM shown as a triangle (white if activated, gray if repressed). Numbering is as in Fig. S3. An important aspect of looping that we do not attempt to represent in any of our figures is intrinsic or CI-dependent curvature of operator DNA.

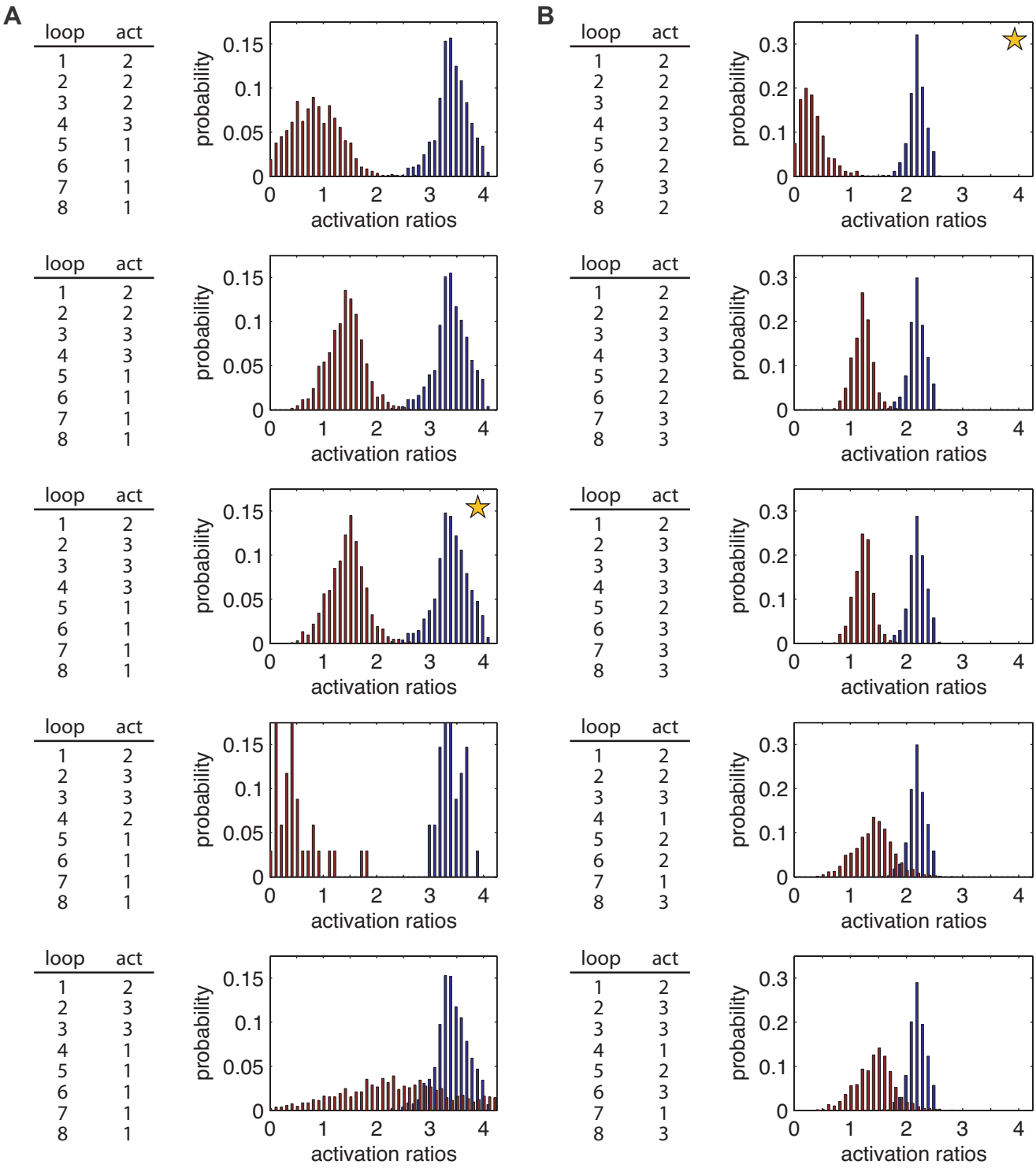

Fig. S3. Comparison of activation state assignments for UP element and looping activation models. Histograms of *A*2 (blue) and *A*3 (red), the activation levels of act2 and act3 relative to act1, shown for different ways of grouping the eight looped activated configurations into activation states based on the model that the UP element adjacent to OL3 is responsible for the observed activation enhancement (*A*) or the model that the looping interaction is itself responsible for the observed activation enhancement (*B*). Looped configuration assignments are shown to the left of each plot, with the left column indicating the configuration number (from Fig. 4) and the right column indicating the activation state assignment. Starred datasets are presented in the main text.

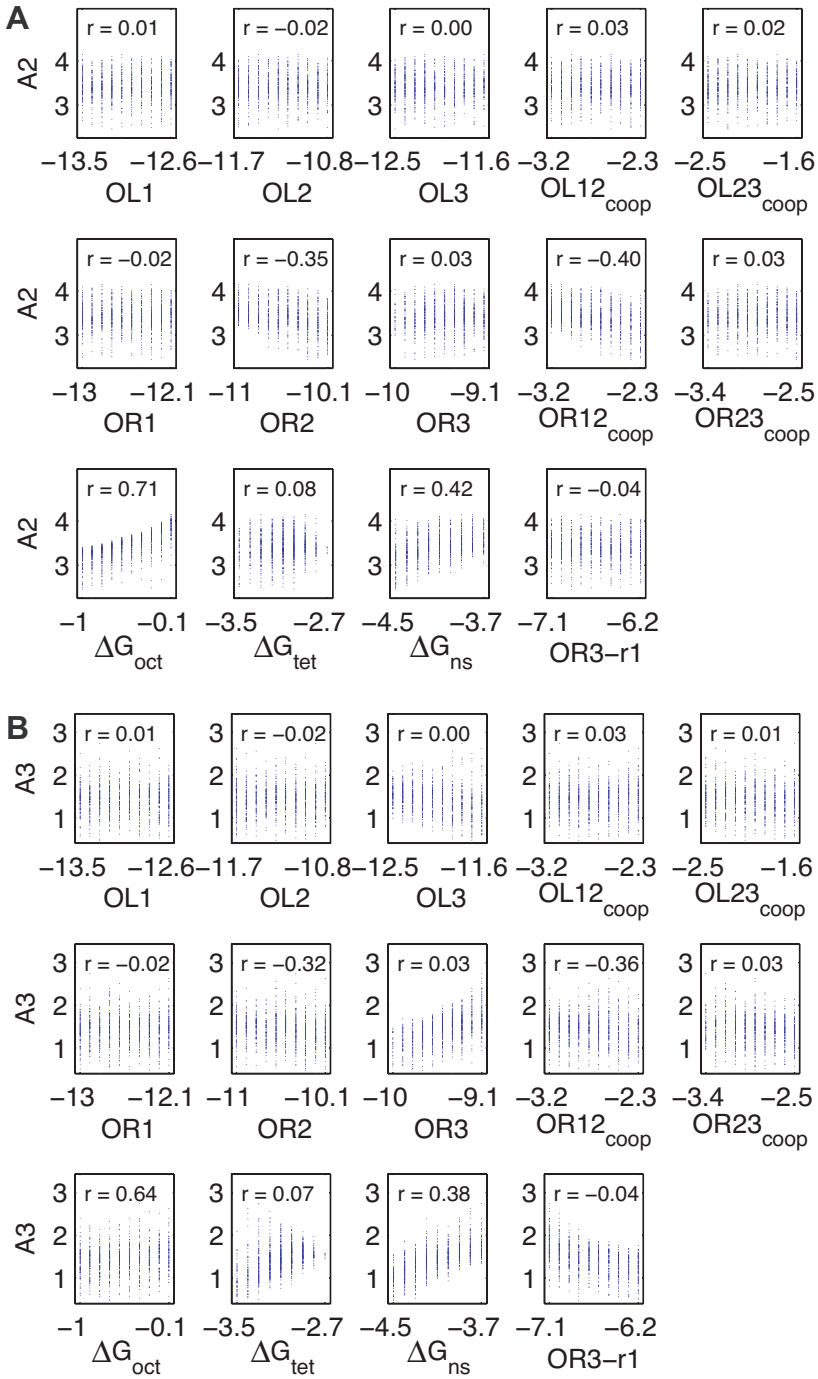

Fig. S4. Sensitivity of activation ratios for the UP model. Data are from the subset of parameter sets that have mean-squared error less than or equal to the experimental uncertainty. All free energies are in kcal/mol. (*A*) Plots and linear correlation coefficients of *A*2, act2 relative to act1, and each binding free energy parameter assuming the UP element model. (*B*) Data for *A*3, the activation strength of act3 relative to act1 and the same set of free energy parameters.

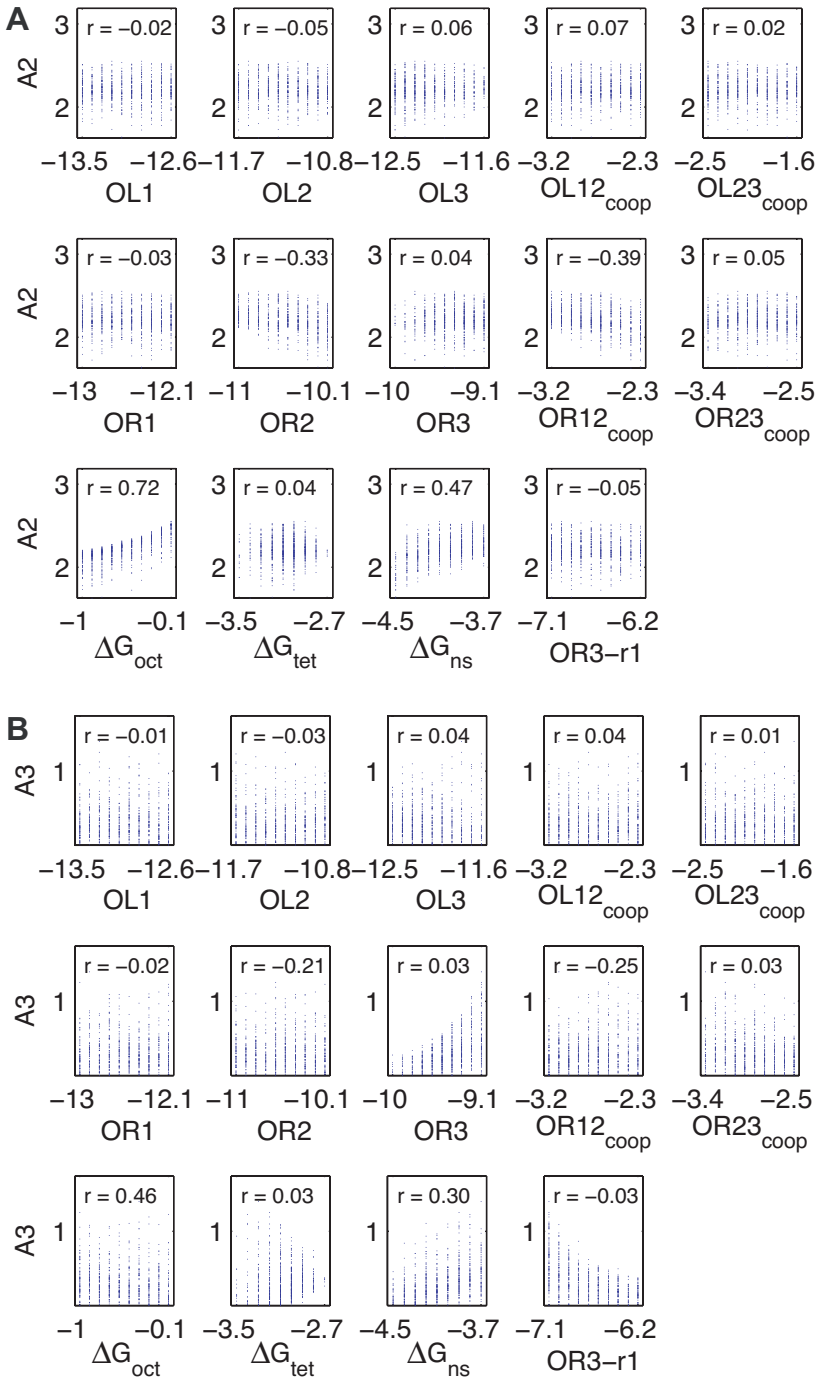

Fig. S5. Sensitivity of activation ratios for the looping activation model. Data are from the subset of parameter sets that have mean-squared error less than or equal to the experimental uncertainty. All free energies are in kcal/mol. (*A*) Plots and linear correlation coefficients of *A*2, act2 relative to act1, and each binding free energy parameter assuming looping interaction activation. (*B*) Data for *A*3, the activation strength of act3 relative to act1 and the same set of free energy parameters.

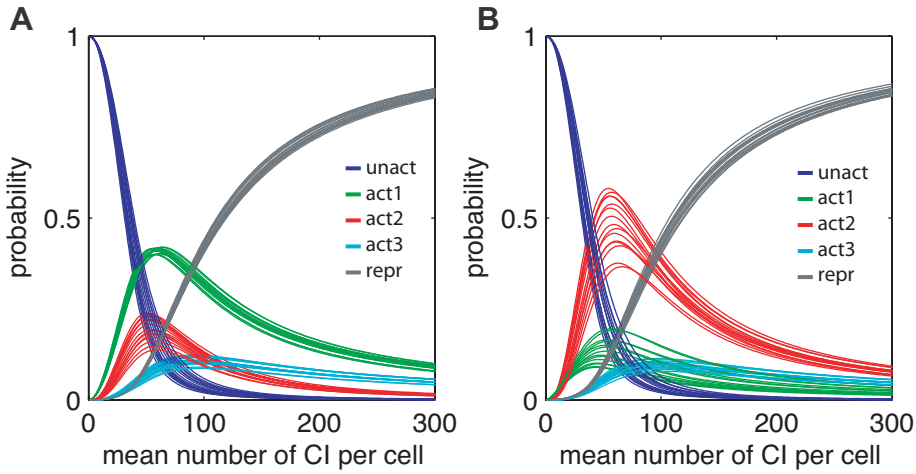

Fig. S6. Activation state probabilities. The predicted CI dependence of each activation state in a WT construct for the UP element (*A*) and the looped octamer activation (*B*) models. The data shown correspond to the operator state assignments described in the main text.

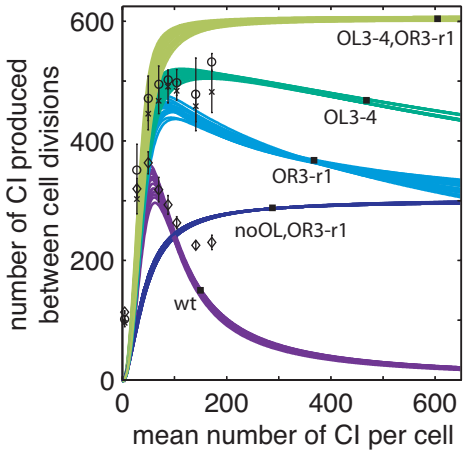

Fig. S7. Comparison of experimental and theoretical activation curves. The activation curves of Fig. 5*C* are overlaid with activation curve data provided by I.B. Dodd, from figure 2 of ref. 7. Diamonds, WT; x, OR3-r1; circles, OL3-4. The *x* axis of the experimental data has been rescaled so that 150 CI per cell corresponds to the P_{RM} activation level observed for the WT construct containing OL when lysogenic levels of CI were provided by a λatt80 prophage (9), and the *y* axis was rescaled to match the maxima for the WT construct. The steady-state mean CI levels measured in this study are indicated by filled black squares.

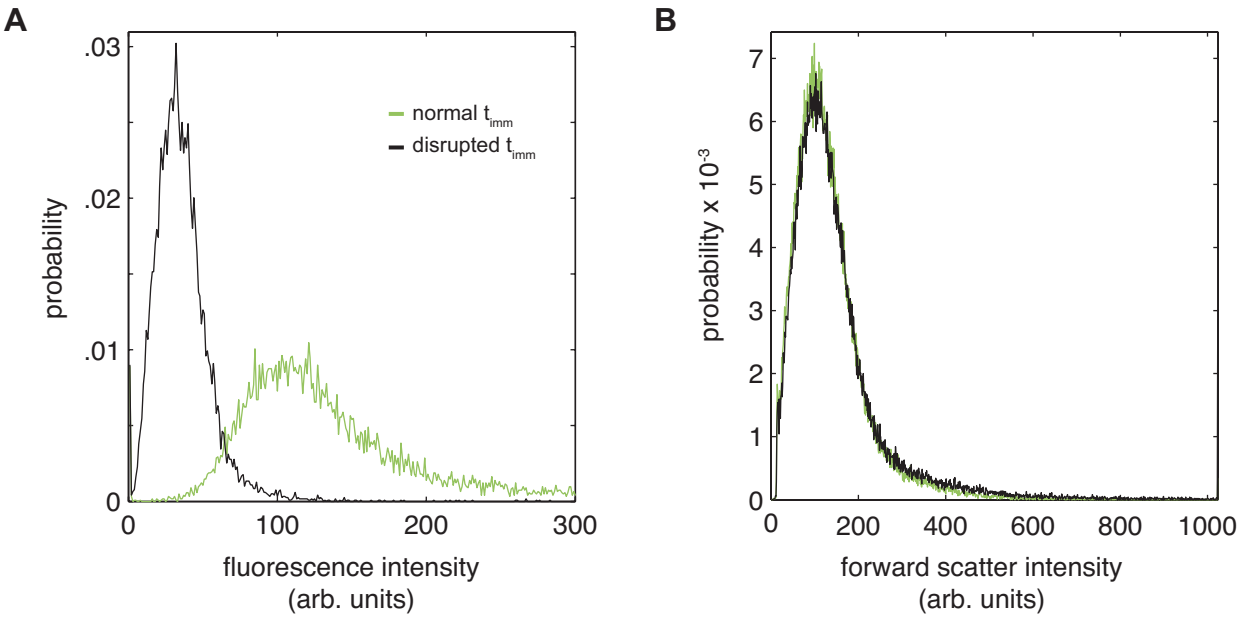

Fig. S8. Termination at O_L. (*A*) Fluorescence histograms from DH5α cells carrying the high copy plasmid pLOI2403 OL3-4,OR3-r1 construct with the normal t_{imm} termination site and for an equivalent plasmid after disrupting the hairpin loop of the terminator. The histograms represent cells within 10% of the mean FSC, collected at $OD_{600}\approx0.4$. The high copy plasmids have substantially higher fluorescence intensity than the low copy plasmids and were collected at a lower voltage and amplifier gain settings. (*B*) The two samples have nearly identical forward scattering, indicating that the relative fluorescence intensities are proportional to the relative concentrations of CI and Venus.

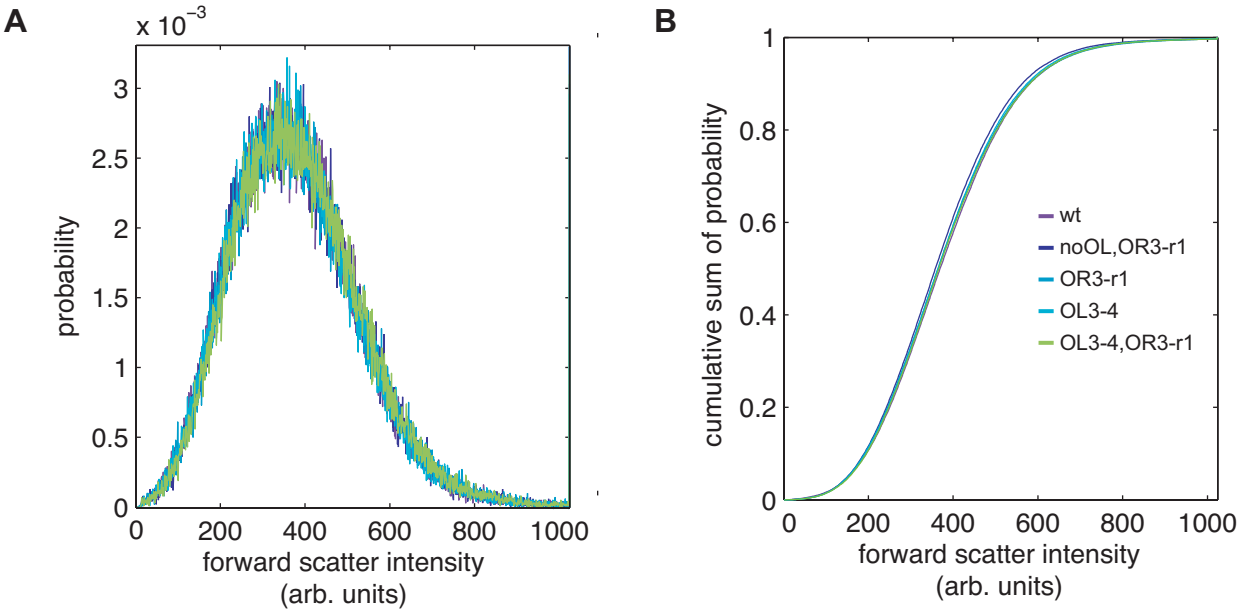

Fig. S9. Forward scatter. (*A*) Histograms for the five constructs. (*B*) Cumulative sums for the histograms. The data from this time point had the smallest variation between samples for forward scatter.

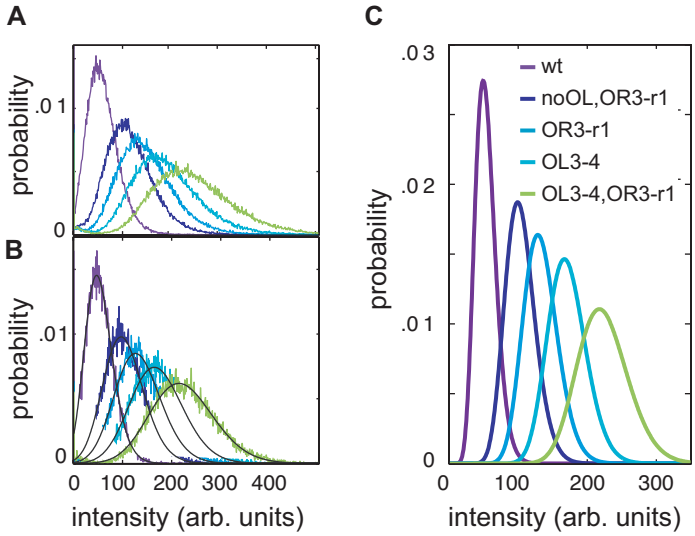

Fig. S10. Flow cytometry histograms. (*A*) Raw flow cytometry fluorescence histograms for each construct calculated from the entire dataset of 100,000 cells. (*B*) Fluorescence histograms of cells that have forward scattering intensity that falls within 10% of the most probable value (colored traces). The black lines are the reconvolution of the instrument response with the deconvolved gamma distributions. (*C*) Deconvolved distributions.

Table S1. Gamma distribution parameters from deconvolution

Parameter	WT	noOL,OR3r1	OR3r1	OL3-4	OL3-4,OR3-r1
θ	4.22	4.54	4.58	4.43	5.96
k	12.7	22.9	29.1	38.7	37.5

Table S2. Relative fluorescence for raw and deconvolved data

Dataset	WT	noOL,OR3r1	OR3r1	OL3-4	OL3-4,OR3-r1	*n*
Expt 1 (raw data)	1	1.92 ± 0.07	2.46 ± 0.11	3.13 ± 0.07	4.01 ± 0.08	5
Expt 2 (raw data)	1	1.92 ± 0.02	2.44 ± 0.02	3.12 ± 0.04	4.04 ± 0.04	7
Mean	1	1.92 ± 0.03	2.45 ± 0.06	3.12 ± 0.04	4.03 ± 0.04	
Best sample (deconvolved)	1	1.93	2.46	3.14	4.05	

Table S3. Primers and probes for quantitative PCR

Gene	Forward primer	Reverse primer	MGB
16s rRNA	CCAGCAGCCGCGGTAAT	CGCTTTACGCCCAGTAATTCC	CGGAGGGTGCAAGC
cI	TGGCTTGAGGTTGAAGGTAATTC	TCAGGAAAGCTTGGCTTGGA	ACCGCACCAACAGG
cro	AGACAGCTAAAGATCTCGGCGTAT	TCGGCCTGCATGAATGG	TCAAAGCGCGATCAA

MGB, minor grove binder labeled with FAM.

Table S4. Quantitative PCR estimates of transcript number per cell

Gene	pKLJ12	truncCI	N99λ$^+$	WT	K12λ$^+$F$^+$	noOL,OR3r1	OR3r1	OL3-4	OL3-4,OR3-r1
cI	0.1	2	44	54	63	140	170	240	320
cro	0.6	69	3.4	1.6	5.3	2.2	2.6	1.7	1.6

Other Supporting Information Files

Dataset S1
Dataset S2