Supplementary Information for "Robustness of DNA Looping Across Multiple Cell Divisions in Individual Bacteria"

Chang Chang[a,1], Mayra Garcia-Alcala[a,b], Leonor Saiz[c,d,2], Jose M.G. Vilar[e,f,2], Philippe Cluzel[a,2]

[a]Department of Molecular and Cellular Biology, Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

[b]Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México.

[c]Department of Biomedical Engineering, University of California, Davis, Davis, CA 95616, USA

[d]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[e]Biofisika Institutua (CSIC, UPV/EHU), University of the Basque Country UPV/EHU, 48080 Bilbao, Spain

[f]IKERBASQUE, Basque Foundation for Science, 48013, Bilbao, Spain

[1]Current address: Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China.

[2]Corresponding authors

Table of Contents

Strains construction

We constructed all the strains using the lambda red recombination technique (1, 2) derived from *Escherichia coli* strain MG1655 (The Coli Genetic Stock Center, Yale University; CGSC 6300), with the helper plasmid pSIM5 expressing recombineering functions (3). To construct a strain without an antibiotic marker, we used a "scarless" chromosomal engineering technique based on a counter-selection cassette (4). In the first step, we inserted a linear fragment from P_{araB}-*ccdB* cassette (gift from J. Mark Kim) into target site using kanamycin for selection. Then this region was further replaced by the final construct, using arabinose for selection. The expression of CcdB is toxic to *E. coli* in the absence of CcdA, thus, with the induction by arabinose, the fitness of a cell without P_{araB}-*ccdB* cassette (e.g., replaced by the final construct) outperforms a cell with the cassette. On the other hand, to construct a strain with an antibiotic marker, we directly integrated into the target site a linear fragment with both the final construct and the antibiotic marker, using this specific antibiotic for selection (kanamycin in our study). Most of the linear fragments were derived from plasmids (Table S2). For all strains used in mother machine experiments, the temperature-sensitive helper plasmid pSIM5 was removed (3), and their single clones were obtained for further experiments. The code and the descriptions of the strains are available in Table S1.

The default carbon source in our experiments is glycerol, under which condition the synthesis of flagella is induced. As a consequence, many of the cells in the mother machine would swim away from the channels of the mother. To disable the motility of the cells, we knocked out in all strains of this study *fliC*, the gene that encodes for the protein flagellin so that flagellar filament cannot be synthetized. We constructed CC-41 (*E. coli, str.* MG1655, Δ*fliC*) from MG1655 in a "scarless" way, which is the starting point of other derived strains. In the first step, a strain with Δ*fliC* is constructed by replacing *fliC* with a linear fragment from P_{araB}-*ccdB* cassette, using kanamycin for selection. In the second step, with the counter selection, the sequence of this integrated linear fragment is removed using a short oligo sharing homology with both ends. CC-41 is denoted as the background strain, which does not carry any fluorescence reporter gene nor any antibiotics maker.

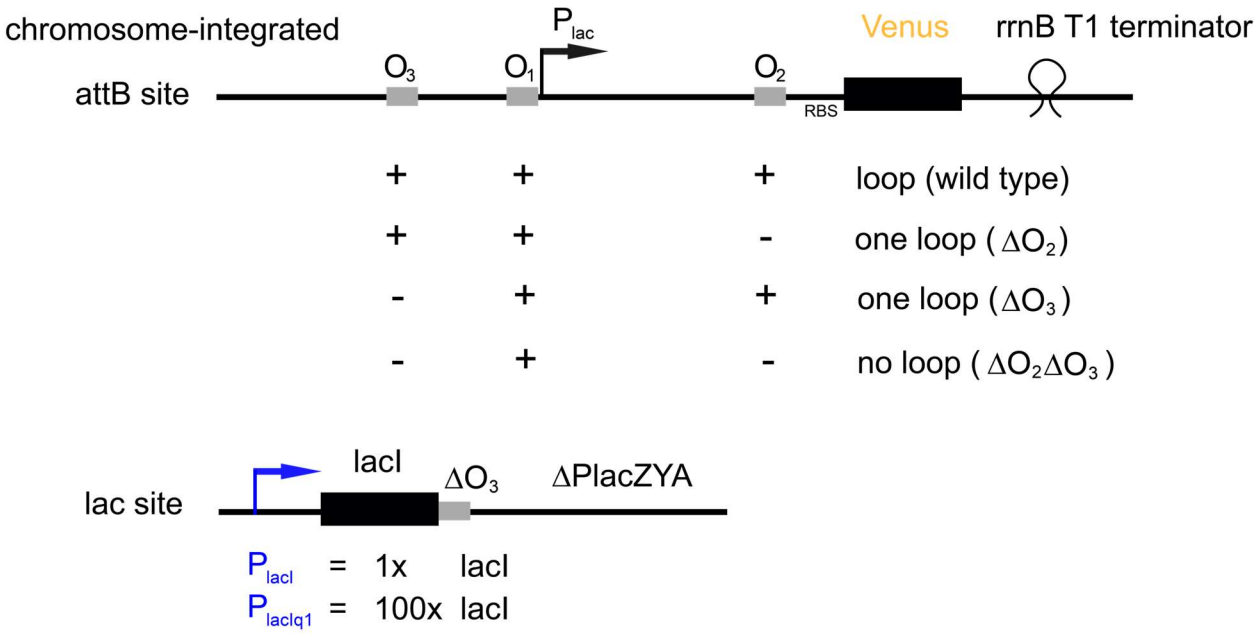

Fig. S1. Cloning design. All constructs are integrated into the chromosome rather than a plasmid to avoid cell-to-cell copy number variations. The native *lacI* site is decoupled from the DNA looping cassette site that was cloned into the *attB* site. At the *lac* site, O_3 partially overlaps with the 3' end of *lacI* that was kept intact after the O_3 deletion (Table S3), and no antibiotic marker is integrated.

In this study, we want to explore the effect of the number of functional operators in the DNA looping cassette and of the repressor concentration on repression strength of the *lac* operon. To decouple those two factors, we kept the *lacI gene* intact at the native site but deleted the looping cassette from the native site and cloned it at the *attB* site (Fig. S1). First, we constructed CC-50 ((Δ*lacI-lacA*)::P_{araB}-*ccdB*) from CC-41 using a linear fragment from P_{araB}-*ccdB* cassette to knock out all genes between *lacI* and *lacA* at the *lac* site. Secondly, we inserted *lacI* with different promoters back to the *lac* site but without DNA looping cassette nor any antibiotic marker: P_{lacI} for CC-51 (Δ(*lacO₃-lacA*)) that has wild type repressor expression, and P_{lacIq1} for CC-53 (ΔP_{lacI}::P_{lacIq1}, Δ(*lacO₃-lacA*)) that over-expresses LacI at an order of 100 times (5).

In a final step, CC-54, CC-58, CC-64 and CC-67 were constructed from CC-51 and CC-53 by cloning the DNA looping cassette to the *attB* site, and those strains consist of high and low LacI

repressor expression combined with either the presence or absence of the DNA looping cassette (Table S1). The looping cassette is under the control of the native *lac* promoter. For simplicity, we denote CC-54 as the *Loops* strain, CC-58 as the *100x/Loops* strain, CC-64 as the *No-loop* strain, and CC-67 as the *100x/No-loop* strain (Table S1), where the number before the slash refers to the order of the relative LacI concentration to the wild type.

We constructed CC-54 (Δ(*lacO₃-lacA*), *attB*::P_{lac}-*lacZ*(1-430)-SD-*mVenusNB*) from CC-51, and CC-58 (ΔP_{lacI}::P_{lacIq1}, Δ(*lacO₃-lacA*), *attB*::P_{lac}-*lacZ*(1-430)-SD-*mVenusNB*) from CC-53, with a linear fragment from pCC-12 (Table S2). pCC-12 (sc101, attB1-P_{lac}-*lacZ*(1-430)-SD-*mVenusNB*-attB2) has a fragment sharing homology with two sides of the *attB* site. It contains the native *lac* operon, with *lacZ* being truncated at ~ 50 base pair after O_2, following by the fast maturating fluorescence reporter VenusNB (6) that is under the control of the *lac* promoter and is followed by a kanamycin resistant gene (after the *rrnB* T1 terminator). The linear fragment amplified from pCC-12 is integrated into the *attB* site of CC-51 and that of CC-53, thus both CC-54 and CC-58 have the kanamycin selection marker.

We constructed CC-64 (Δ(*lacO₃-lacA*), *attB*::P_{lac}-Δ(*lacO₃*)-*lacZ*(1-430)-Δ(*lacO₂*)-SD-*mVenusNB*) from CC-51, and CC-67 (ΔP_{lacI}::P_{lacIq1}, Δ(*lacO₃-lacA*), *attB*::P_{lac}-Δ(*lacO₃*)-*lacZ*(1-430)-Δ(*lacO₂*)-SD-*mVenusNB*) from CC-53, with a linear fragment from pCC-16 (Table S2). pCC-16 (sc101, attB1-P_{lac}-Δ(*lacO₃*)-*lacZ*(1-430)-Δ(*lacO₂*)-SD-*mVenusNB*-attB2) also has a fragment sharing homology with *attB* site. pCC-16 is identical to pCC-12 except that O_2 and O_3 are knocked out to disable the DNA looping cassette, using sequences obtained from Refs. (7, 8) (Table S3). The linear fragment amplified from pCC-16 is integrated into the *attB* site of CC-51 and CC-53, thus both CC-64 and CC-67 also have the kanamycin selection marker.

In addition to those Loops and No-loop strains, we constructed four One-loop strains (O_1-O_2 One-loop, O_1-O_3 One-loop, 100x/O_1-O_2 One-loop, 100x/O_1-O_3 One-loop; see Table S1 and Fig. 3). The construction of those One-loop strains is similar to that for the Loops or No-loop strains. The parent strains are still CC-51 and CC-53, but the linear fragments integrated into the *attB* site are from pCC-14 or pCC-15 (Table S2). Those two plasmids are identical to pCC-12 except

that O_2 or O_3 is knocked out (Table S3). As a consequence, those One-loop strains also have the kanamycin selection marker.

In addition to those strains, we also have one strain, CC-45, with wild type repressor expression level and the DNA looping cassette at the *lac* site (Table S1). This strain (Δ*fliC,* (Δ*lacZ-lacA*)::*lacZ*(1-430)-SD-*mVenusNB*) also has a truncated *lacZ* and the fluorescence reporter VenusNB, but it does not have an antibiotics marker. For CC-45, the sequence between P_{lac} to *VenusNB* is identical to that between P_{lac} to *VenusNB* at the *attB* site in CC-54. The modifications at the *lac* site of CC-45 were done in a 'scarless' way.

Strain	Parent strain	Genotype	Notation used in the text
CC-41		*E. coli, str.* MG1655, Δ*fliC*	background
CC-45		(Δ*lacZ-lacA*)::*lacZ*(1-430)-SD-*mVenusNB*	lac::Loops
CC-50	CC-41	(Δ*lacI-lacA*)::P_{araB}-*ccdB*	
CC-51	CC-50	Δ(*lacO₃-lacA*)	
CC-53	CC-50	ΔP_{lacI}::P_{lacIq1}, Δ(*lacO₃-lacA*)	
CC-54	CC-51	*attB*::P_{lac}-*lacZ*(1-430)-SD-*mVenusNB*	Loops
CC-58	CC-53	*attB*::P_{lac}-*lacZ*(1-430)-SD-*mVenusNB*	100x/Loops
CC-62	CC-51	*attB*::P_{lac}-*lacZ*(1-430)-Δ(*lacO₂*)-SD-*mVenusNB*	O_1-O_3 One-loop
CC-63	CC-51	*attB*::P_{lac}-Δ(*lacO₃*)-*lacZ*(1-430)-SD-*mVenusNB*	O_1-O_2 One-loop
CC-64	CC-51	*attB*::P_{lac}-Δ(*lacO₃*)-*lacZ*(1-430)-Δ(*lacO₂*)-SD-*mVenusNB*	No-loop
CC-65	CC-53	*attB*::P_{lac}-*lacZ*(1-430)-Δ(*lacO₂*)-SD-*mVenusNB*	100x/O_1-O_3 One-loop
CC-66	CC-53	*attB*::P_{lac}-Δ(*lacO₃*)-*lacZ*(1-430)-SD-*mVenusNB*	100x/O_1-O_2 One-loop
CC-67	CC-53	*attB*::P_{lac}-Δ(*lacO₃*)-*lacZ*(1-430)-Δ(*lacO₂*)-SD-*mVenusNB*	100x/No-loop

Table S1. *E .coli* strains in this study. The source of all strains is from this work.

Plasmid	Genotype/Description	Antibiotics	Source
pCC-12	sc101, attB1-P_{lac}-*lacZ*(1-430)-SD-*mVenusNB*-attB2	Kan	This work
pCC-14	sc101, attB1-P_{lac}-*lacZ*(1-430)-Δ(*lacO₂*)-SD-*mVenusNB*-attB2	Kan	This work

pCC-15	sc101, attB1-P_{lac}-Δ(*lacO₃*)-*lacZ*(1-430)-SD-*mVenusNB*-attB2	Kan	This work
pCC-16	sc101, attB1-P_{lac}-Δ(*lacO₃*)-*lacZ*(1-430)-Δ(*lacO₂*)-SD-*mVenusNB*-attB2	Kan	This work

Table S2. Plasmids in this study.

Operator	Sequence
$lacO_2$	GGTTGTTACTCGCTCACATTT
Δ$lacO_2$	GGCTGCTATAGCTTGACGTTT
$lacO_3$	GGCAGTGAGCGCAACGCAATT
Δ$lacO_3$	GGCAGTGATGAAGCTTGTCAG

Table S3. Operator and their sequences. The sequences of Δ$lacO_2$ and Δ$lacO_3$ are obtained from Refs. (7, 8).

Site	Primer	Sequence
lac	mCC-1	AGCAAAACAGATCGAAGAAGGG
	mCC-9	GGTCAAAGAGGCATGATGCGAC
attB	mCC-46	AAGACCGCAGAGCAGAGAAC
	mCC-47	TGTTGTCACCTGCTACGACC
fliC	mCC-98	GTTGCCGTCAGTCTCAGTTAATCAGGTTAC
	mCC-99	ACCCGACTCCCAGCGATGAAATAC

Table S4. Primers used in this study.

We performed Sanger sequencing to verify these strains (Table S4). We used the primers mCC-1 and mCC-9 for checking the *lac* site, mCC-46 and mCC-47 for the *attB* site, and mCC-98 and mCC-99 for the *fliC* site, verify all news strains with sequencing. Sequencing primers are properly chosen to make sure that all regions related to experiments were examined.

Experimental setup

M9 media was prepared with M9 minimal salt (BD, Difco, catalog number: 248510), complemented with 0.1 mM $CaCl_2$ (MilliporeSigma, catalog number: EM1.02378.0500), 2mM $MgSO_4$ (Sigma-Aldrich, catalog number: M1880), 1μg/ml Thiamine (Sigma-Aldrich, catalog

number: T4625), 0.85g/L Pluronic F-108 (Sigma-Aldrich, catalog number: 542342), 0.50% casamino acids (BD, Bacto, catalog number: 223050), and 0.40% Glycerol (VWR, BDH, catalog number: BDH1172).

We used Zeiss Axiovert 200M microscope, with all the settings be identical as in (6). In preparation for a typical experiment, cells were grown in M9 media for overnight at 30°C (no antibiotics for the background strain, and 25 ug/ml kanamycin for Loops, One-loop and No-loop strains). At the day of the experiment, the cell culture was centrifuged and loaded into the inlets of microfluidics device by pipetting. Syringes with M9 media (no kanamycin) were connected to the inlets (BD, 60ml), and the outlets were connected to an empty beaker. The initial flow rate was 35 µl/min (~ 1 hour) for cleaning the inlets and outlets and the flow rate was adjusted to 7 µl/min during the experiment. All experiments were performed at 30°C. In this study, one mother machine microfluidics device (9, 10) has four independent quadrants, each has a series of growth channels allowing the observation of the old-pole mother cell and its progeny. This device allows the simultaneous observation of up to four different conditions, each for one quadrant. In each experiment, we included the background strain as a control, and we filled the other quadrants with Loops, One-loop or No-loop strains. Phase contrast and fluorescence images were obtained for each field of view every 5 minutes. After loading the microfluidic device with bacteria, the first five hours of data is discarded to ensure that the cells under observation are in the exponential phase, and subsequent duration of the recorded data is ≥ 40 hours. The number of lineages of a Loops, One-loop or No-loop strain included in the analysis is ≥ 50 (the 100x/O_1-O_3 One-loop strain only has 21 lineages), and the number of lineages of the background strain in the same experiment is ≥ 20.

The doubling time of an *E. coli* cell is only sub-optimal in 30ºC, which is the default temperature for our experiments. With this temperature, we can acquire (a) more data points within each division cycle, and (b) the fluorescent reporter has better properties such as smaller fluctuations of the maturation time as reported in Balleza *et al.* (6).

Image processing

We analyzed our microscopy images based on the software of *molyso*, which includes image registration, cell segmentation and lineage tracking (11). Here, we only tracked the old-pole mother cells, which always stayed at the end of growth channels. We modified the codes of *molyso* to enable manual corrections of the results of cell segmentation. Cell divisions were determined automatically based on cell length, complemented with manual check.

We calculated the total fluorescence per cell as the sum of the fluorescent intensities of all pixels belonging to a cell, subtracted with local background level. We also analyzed a cell without a fluorescent reporter and calculated its total auto-fluorescence.

The definition of repression

In our experiments, we monitor the spontaneous leakiness of the promoter, as a measure for the repression level of the *lac* promoter in the presence or absence of DNA looping. Our definition of repression differs from that of Müller-Hill et al. in Refs. (12, 13) that includes the fully induced promoter, however, they both yield the same qualitative pictures.

Characterizing background fluorescence signals

With the tight regulation of the DNA looping, most cells contain only few copies of the protein under the control of the *lac* promoter (14). As expected, from our experiments, the distributions of total fluorescence per cell of the Loops and No-loop strains show a low mean in such a way that they are not clearly separated from the distributions of the total auto-fluorescence per cell of the background strain (Fig. S2). This result indicates that, we need to explicitly consider the contribution of the auto-fluorescence to the total measured fluorescence signal in order to identify which part of the measured fluorescent signal is actually directly associated with bursts of VenusNB synthesis and not random fluctuations of auto-fluorescence.

The total auto-fluorescence per cell of the background strain is used for the inference of promoter activity for the Loops and No-loop strains. To determine if the auto-fluorescence signals from different experiments are consistent, we compared the distributions of the auto-fluorescence levels from two repeated experiments. Although small differences exist, they

overlap well with each other (Fig. S3). We also found that the total auto-fluorescence is proportional to the cell size (Fig. S3).

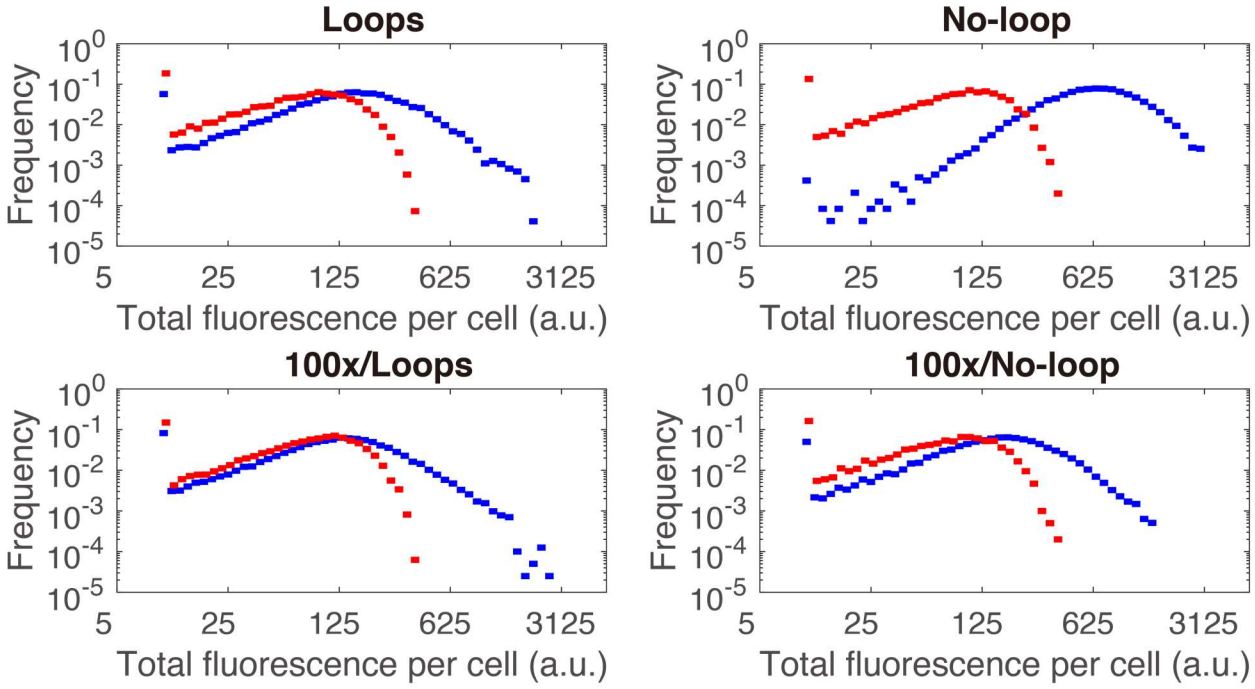

Fig. S2. Distributions of the total fluorescence per cell from the background strain and the four strains (Loops and No-loop) described in Table S1. The red dots represent the distribution of the total auto-fluorescence per cell from the background strain CC-41 (table S1), and the blue dots represent the distribution of total fluorescence per cell for each of the four strains (Loops and No-loop) expressing Venus as described in table S1. The distribution from the Loops strain is almost identical to that of 100x/No-loop and also similar to 100x/Loop, but is different from that of the No-loop strain.

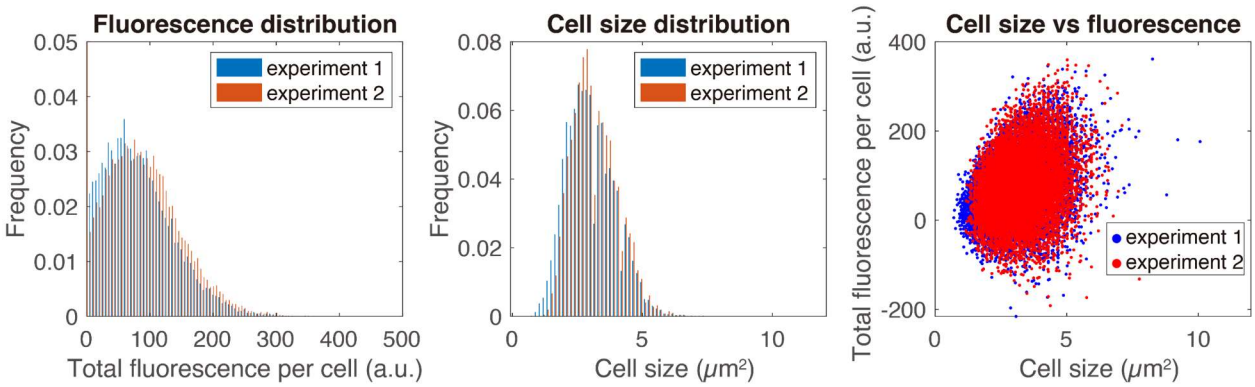

Fig. S3. Characterization of the auto-fluorescence signal from the background strain. (left panel) The distributions of the total auto-fluorescence per cell, and (middle panel) the cell size distributions of background strain at different timepoints in two independent experiments. (right panel) Scatter plot of cell size versus total fluorescence per cell.

Signal-to-noise ratio and the amplitude of pulses

To determine to which extent we can distinguish the "true" signal from the background, we calculated the signal-to-noise ratio

$$SNR = \frac{\mu_{ON} - \mu_{background}}{\sigma_{background}} \quad (1)$$

where μ_{ON} is the mean fluorescent signal when the promoter is inferred as ON (based on a probabilistic inference algorithm mentioned in the later part of the Supplementary Information), and $\mu_{background}$ and $\sigma_{background}$ are the mean and standard deviation, respectively, of the auto-fluorescence calculated from the background strain that is grown within the same microfluidic chip as the Loops or No-loop strains. A Savitzky-Golay filtering (15) is applied to the fluorescence signals of both background and experimental strains. We also estimated the standard deviation for signal-to-noise ratio by applying Eq. (1) to all fluorescent signals when the promoter is ON. The signal-to-noise ratio is 6.0 ± 5.2 for the Loops strain, 5.9 ± 5.2 for the 100x/Loops strain, 6.0 ± 4.7 for the 100x/No-loop strain, and 18.0 ± 10.5 for the No-loop strain. Interestingly, the signal-to-noise ratio for the first three strains are similar.

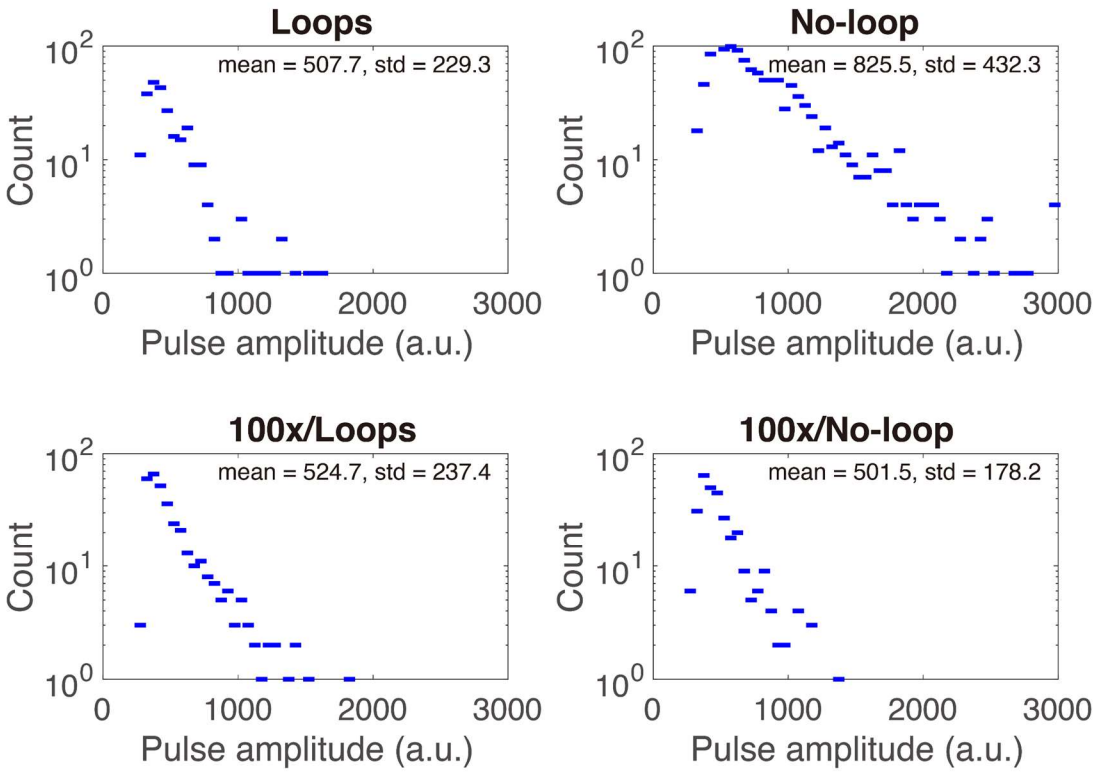

Fig. S4. Estimation of pulses amplitude. To determine to which extent the "true" signal is separated from the background (shown in Fig. S3), we directly estimated the distributions of the pulse amplitude from fluorescence time series. In line with the observations from Fig. S2, the distribution of the Loops strain is similar to those of the 100x/Loops and 100x/No-loop strains, but different from that of the No-loop strain.

For the same purpose, we also make a rough estimate of the amplitude of the pulses observed from the Loops and No-loop strains. The estimate of pulse amplitudes is using only fluorescence signal within a single cell cycle (thus a pulse lasting more than one cell cycle will not be counted in this estimation). If the maximum fluorescence signal within a cell cycle is larger than a threshold (three sigma above mean background signals), and the difference between the maximum and the average fluorescence signals within the same cell cycle is larger than a second threshold (three times of background signals standard deviation), the difference between the maximum and the minimum fluorescence signals is considered as a candidate for estimating the pulses. We find that the estimated amplitude of pulses from the loops strain is quite similar to

those of the 100x/Loops and 100x/No-loop strains (Fig. S4). On the other hand, the estimated amplitude of the No-loop strain is larger than that of the Loops strain, but their pulse peaks are still of the same order (~ several hundred a.u.; Fig. S4). However, the ratio between the mean fluorescence level of the No-loop strain and the Loops strain is 3.97. Together, these observations suggest that the expression difference between the Loops strain and the No-loop strain is not solely caused by the amplitude of expression, the frequency of expression of the pulses may also play a role, since frequent protein expression can lead to higher fluorescence signal if the fluorescent proteins being produced are not diluted out when the second pulse happened.

A probabilistic inference algorithm

We developed a probabilistic method to robustly infer the promoter activity in the regime of weak signals, considering not only the cell-size dependent auto-fluorescence but also fluorescence signals of adjacent timepoints. Overall, the detected distributions of OFF intervals and burst size agree with those of the input distributions. The codes are available at https://github.com/changsysbio/ProbabilisticInferenceForPromoterActivity .

Correlation between promoter activity and gene dosage during cell growth

According to the Cooper-Helmstetter model there should be a significant lag after division before the replication at the attB site. Consequently, the promoter activity should exhibit a flat region at the early phase of the cell cycle as observed in Ref. (16) and explained in our main text. However, the promoter is steadily active in Ref. (16) following a steady IPTG induction. By contrast, in our experiments, the promoter activity is sparse and highly stochastic with ON and OFF firing. And indeed, with our current analysis, we do not observe a flat region.

To reconcile our results with that of a steadily active promoter like in Ref. (16), we replotted our data by taking into account only the active ON regions and ignoring OFF regions in the firing time series to calculate the mean promoter activity (Fig. S5). Using this alternative analysis, identical to Ref. (16), we find that the promoter activity is flat in the early phase of the cell cycle as expected by the Cooper-Helmstetter model.

In Table S5, we also give the statistics of the doubling time for different strains as a reference.

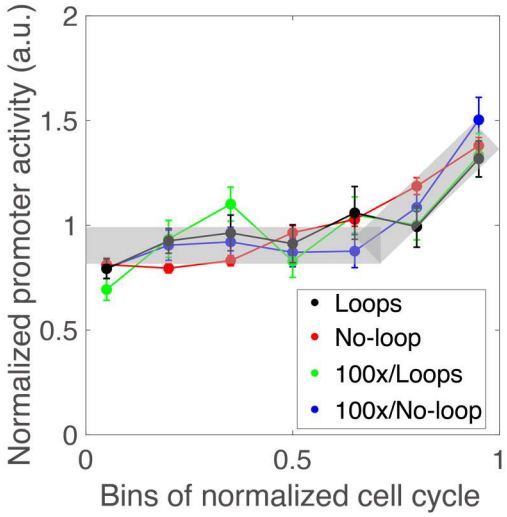

Fig. S5. Normalized non-zero promoter activity versus cell cycle progression. We used the same data set as in Fig. 3b, however the analysis is different because it only takes into account non-zero promoter activity from the time series. This analysis yields the same activity pattern as observed in Ref. (16) (grayed area).

Strains	Mean (min)	Standard deviation (min)	Standard error
Background	66	18	1
Loops	71	22	1
100x/Loops	64	16	0
No-Loop	64	16	0
100x/No-loop	69	21	1
O1-O3 One-loop	63	14	0
100x/O1-O3 One-loop	53	11	0
O1-O2 One-loop	62	15	0
100x/O1-O2 One-loop	55	13	0

Table S5. Statistics of the doubling time.

Robustness of the repression to temperature change

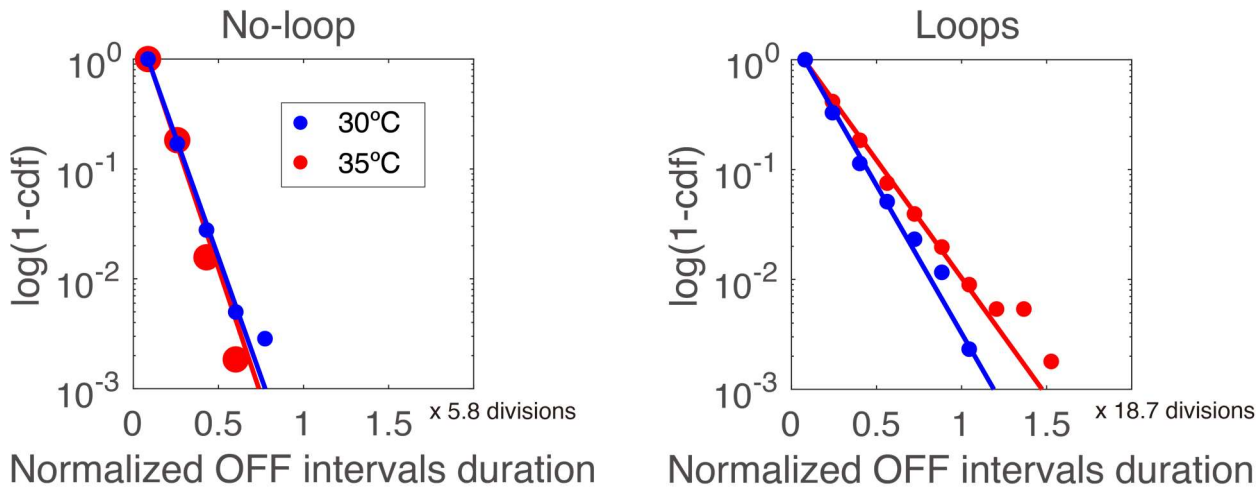

Fig. S6. Robustness of repression under different temperatures (30ºC and 35ºC), measured in units of division time. OFF intervals: the Loops strain (30ºC: 2.8 [SE] ± 0.1 divisions, 35ºC: 3.6 [SE] ± 0.2 divisions), and the No-loop strain (30ºC: 0.64 [SE] ± 0.01 divisions, 35ºC: 0.67 [SE] ± 0.01 divisions). Doubling time of the Loops strain: 30ºC, 71 [SE] ± 1 min; 35ºC: 43 [SE] ± 0 min. Doubling time of the No-loop strain: 30ºC, 64 [SE] ± 0 min; 35ºC: 43 [SE] ± 0 min.

The doubling time as well as the rates of intracellular processes are highly temperature dependent (17). To test whether repression mediated by DNA looping is sensitive to temperature, we compare the leakiness in the Loops and No-loop strains under 30ºC and 35ºC. We find that the OFF intervals duration measured in units of division time changed mildly between these two temperatures (Fig. S6). While the mean OFF intervals with DNA looping are much longer than the division time and shorter without DNA looping, however, they all exhibit exponential distributions (Fig. S6).

CRP and glucose experiments

In the native *lac* operon, it is likely in general that CRP associated with cAMP play a role in governing the lifetime of DNA looping. CAP/CRP binds between O_1 and O_3. Consequently, it can affect the stability of the O_1-O_3 loop. Indeed, it has been shown that it can stabilize loop formation between O_1 and O_3 up to -0.9 kcal/mol (-1.5 kT) (18, 19). However, CAP/CRP does not bind between O_1 and O_2, and therefore there is no obvious reason to assume that it would

affect the formation of the O_1-O_2 loop. Confirming this point, to our knowledge, there is no experimental or theoretical evidence that would suggest CAP/CRP potentially interacts with DNA looping between O_1 and O_2.

Fig. 3a shows that repression through the O_1-O_2 loop alone is robust across cell divisions, which is as robust as the native looping system. Therefore, the observed robustness is not a consequence of the interaction of CAP/CRP with looping. In addition, the repression controlled only O_1-O_3 loop is only slightly stronger than that with no loop and is much weaker than the native looping system (Fig. 3a), suggesting that the stabilization effect of CAP/CRP to O_1-O_3 may not be a major contributor to robustness under our conditions.

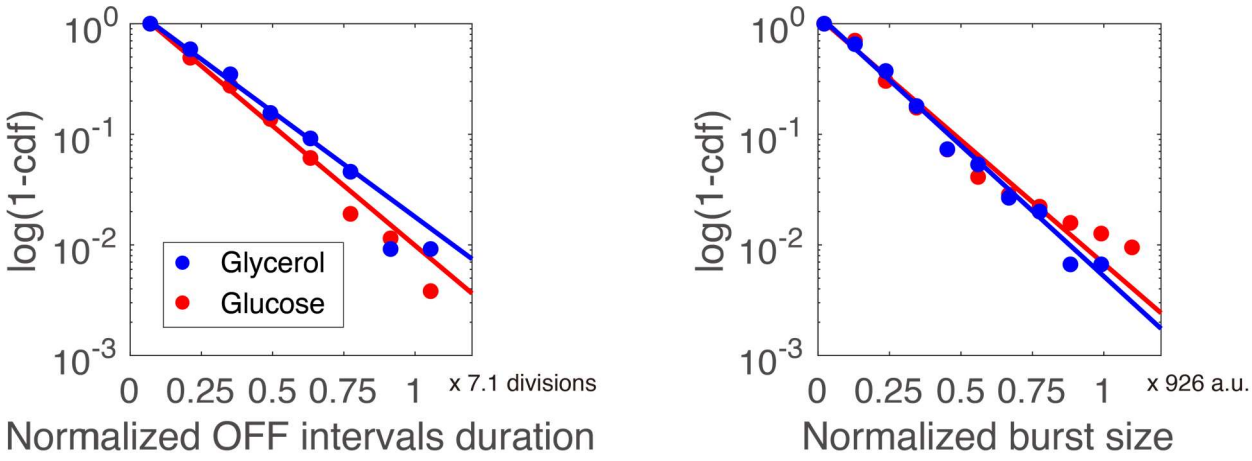

Fig. S7. Repression of the lac::Loops strain (Table S1) with two different carbon sources. (left panel) OFF intervals measured in units of division time, with a normalization factor of 7.1 divisions. Glycerol: 1.7 [SE] ± 0.0 divisions, with doubling time 93 [SE] ± 3 min; Glucose: 1.5 [SE] ± 0.0 divisions, with doubling time 75 [SE] ± 1.3 min. (right panel): Burst size, with a normalization factor of 926 a.u..

To estimate the impact of CAP/CRP on repression with DNA looping, we changed the cAMP levels by using a different carbon source. The intracellular levels of cAMP have been measured experimentally for glucose and glycerol in Ref. (20). The results show that cAMP levels in presence of glycerol are 4× greater than with glucose. At the transcriptional level, it has been

observed that the activity of the *lac* operon with glycerol is slightly higher than with glucose over a wide range of induction levels with IPTG (Fig. 1a of Ref. (21)).

Using the lac::Loops strain, we monitored the OFF intervals and the burst size under M9 media with 0.4% glucose versus with 0.4% glycerol, respectively (the duration of the recorded data is ~ 18 hr). We observed that repression remains robust with either carbon sources. With glucose, we observed that the durations of OFF intervals, as measured in units of division time, and burst size were similar to those with glycerol (Fig. S7). Those observations are consistent with previous cell-population transcriptional activity assays (Fig. 1a in (21)). Finally, we should note that these experiments cannot exclude that untested growth conditions in this paper could yield much lower relative level of cAMP to CAP/CRP, which in turn could affect robustness.

Modeling the effects of DNA looping in gene regulation

Description of the system. The *lac* repressor, with two DNA binding domains, can bind two operators simultaneously by looping the intervening DNA. We consider the simplest realistic model of the *lac* operon that incorporates DNA looping, including the main operator O_1 and an auxiliary operator, referred to as O_a. The repressor's binding to O_1 prevents transcription by the RNA polymerase irrespective of its binding to O_a, which does not prevent transcription.

The canonical description considers that there is a set of transcriptional states s and that mRNA, m, is produced at a rate g_s for each transcription state (22). We consider explicitly 5 transcriptional states, which are labeled as follows:

State	Description
1	O_1 and O_a free
2	O_1 free and O_a bound to the repressor
3	O_1 bound to the repressor and O_a free
4	O_1 and O_a bound to a repressor looping DNA
5	O_1 and O_a each bound to a repressor

We use a vectorial representation of the system in the state space. The transcription rates g_s are expressed as the components of the vector

$$\mathbf{g} \equiv (k_t \quad k_t \quad 0 \quad 0 \quad 0)^T,$$

which specifies transcription taking place at a rate k_t only when the main operator O_1 is free. Analogously, transitions between states result from the binding and unbinding of the repressor. The transition rates $k_{s,s'}$ from the state s to the state s' are specified through the elements of the matrix

$$\mathbf{k} \equiv \begin{pmatrix} 0 & n_R k_{\text{on}} & n_R k_{\text{on}} & 0 & 0 \\ k_{\text{off-Oa}} & 0 & 0 & k_{\text{loop}} & (n_R - 1)k_{\text{on}} \\ k_{\text{off-O1}} & 0 & 0 & k_{\text{loop}} & (n_R - 1)k_{\text{on}} \\ 0 & k_{\text{off-O1}} & k_{\text{off-Oa}} & 0 & 0 \\ 0 & k_{\text{off-O1}} & k_{\text{off-Oa}} & 0 & 0 \end{pmatrix},$$

where k_{on} is the association rate of the repressor for an operator; $k_{\text{off-O1}}$ and $k_{\text{off-Oa}}$ are the dissociation rates of the repressor from O_1 and O_a, respectively; k_{loop} is the rate of loop formation when the repressor is bound to one operator; and n_R is the number of repressors. This description, developed originally in Ref. (23), has been shown to accurately describe the *lac* operon under an exhaustive range of experimental conditions (24, 25), with 1.7-fold accuracy over a 10,000-fold variation of the expression level.

The time evolution of the probability P_s of the state s is given by

$$\frac{dP_s}{dt} = \sum_{s'} [\, k_{s',s} P_{s'} - k_{s,s'} P_s],$$

which takes into account the transitions between transcriptional states.

The steady-state expression of the probability P_s is obtained by solving $0 = \sum_{s'}[k_{s',s}P_{s'} - k_{s,s'}P_s]$, which follows straightforwardly from the preceding equation. The solution, using the statistical weights Z_s, is expressed in vector form as

$$\mathbf{P}^{\text{ss}} = \mathbf{Z}/\|\mathbf{Z}\|_1,$$

where $\|\boldsymbol{Z}\|_1$ is the partition function expressed using the one norm.

To obtain compact expressions, we express the dissociation and the looping rates in terms of the repressor-operator association constants, K_{O1} and K_{Oa}, and looping local concentration, n_L, as

$k_{\text{off-O1}} = k_{\text{on}}/K_{\text{O1}}$, $k_{\text{off-A}} = k_{\text{on}}/K_{\text{Oa}}$, and $k_{\text{loop}} = n_L k_{\text{on}}$. In terms of these parameters, the statistical weights are

$$\boldsymbol{Z} \equiv (1 \quad n_R K_{\text{Oa}} \quad n_R K_{\text{O1}} \quad n_R n_L K_{\text{Oa}} K_{\text{O1}} \quad n_R(n_R - 1) K_{\text{Oa}} K_{\text{O1}})^T.$$

The association constants and looping local concentration are related to the free energies of binding to O_1, ΔF_{O1}, and O_a, ΔF_{Oa}, and of looping, ΔF_L, as $K_{\text{O1}} = e^{-\Delta F_{\text{O1}}}$, $K_{\text{Oa}} = e^{-\Delta F_{\text{Oa}}}$, and $n_L = e^{-\Delta F_{\text{L}}}$, respectively, which use the thermal energy ($k_B T$) as energy units.

Parameter values. We use the number of molecules, abbreviated molec, as the units of substance; 1 molecule/cell, equivalent to 1.5 nM for an *E. coli* cell, as the units of concentration; and minutes, abbreviated min, as the units of time.

The association rate constant for the repressor tetramer binding to an operator at 30°C was set to $k_{\text{on}} = 0.28\ \text{molec}^{-1}\text{min}^{-1}$, consistently with the upper and lower bounds established in figure 2 of Ref. (8) for the repressor dimer at 37°C and 25°C, respectively.

The repressor-operator association constants and the looping concentration were obtained from the free energies inferred in Refs. (24, 25) as $K_{\text{O1}} = 2.76\ \text{molec}^{-1}$, $K_{\text{Oa}} = 0.32\ \text{molec}^{-1}$, and $n_L = 1080$ molec. These values lead to a single O_1 operator repression level, defined as the maximum transcription over the actual transcription and computed as $L = k_t/(\mathbf{P}^{\text{ss}} \cdot \mathbf{g})$, of $L = 27$, to a full system repression level of $L = 2291$, and to the main operator O_1 being 10 times stronger than the auxiliary operator O_2, consistently with the experimental observations of Ref. (12), with similar growing conditions as used here but with a slightly different temperature of 32°C.

These values of k_{on} and K_{O1} lead to $k_{\text{off-O1}} = 0.10\ \text{min}^{-1}$, corresponding to an average occupancy time of O_1 by the repressor of 9.7 min, consistently with the upper and lower bounds established at 37°C and 25°C, respectively, in figure 2 of Ref. (8).

The transcription rate was set to $k_t = 20\ \text{molec} \cdot \text{min}^{-1}$ as reported in Ref. (26).

Burst size and OFF interval definition. We define the OFF interval as the time between the first binding to O_1 after transcription and the subsequent transcriptional event. Therefore, the OFF interval can include multiple rounds of binding and unbinding to O_1. Correspondingly, the ON interval is defined as the time between the first transcriptional event after unbinding from O_1 and the subsequent binding to O_1. In this way, we can unambiguously define the burst size as the number of transcripts produced during the ON interval.

To obtain analytical results, we consider a reduced description. Explicitly, we aggregate the states without potential for transcription into the bound state, denoted by B, which includes the configurations with the repressor bound to O_1; namely, one repressor bound to O_1 and the auxiliary operator free, a repressor bound simultaneously to O_1 and the auxiliary operator by looping the intervening DNA, and one repressor bound to O_1 and another one bound to the auxiliary operator. Its probability is given by $P_B = P_3 + P_4 + P_5$. Analogously, we aggregate the states with potential for transcription into the free O_1 state, denoted by E, which includes the configurations with the repressor not bound to O_1; namely, O_1 free and the auxiliary operator occupied, and both operators free. Its probability is given by $P_E = P_1 + P_2$.

OFF interval statistics. To account for transcriptional events in the reduced description, we introduce a transcription start state, denoted by TS, so that a transcript is produced when the system reaches this state. The resulting transitions between the reduced states are described by

$$B \underset{k_b}{\overset{k_u}{\rightleftarrows}} E \overset{k_t}{\rightarrow} TS,$$

where k_b is the effective binding rate for the repressors to O_1, k_u gives the unbinding rate for a repressor from O_1, and k_t is the effective transcription rate.

When the system reaches the state E, there is a probability

$$\alpha = \frac{k_b}{k_b + k_t}$$

of returning to the bound state. The probability of l unbinding events before transcription occurs is

$$P_l = \alpha^{l-1}(1 - \alpha).$$

Considering that $k_u \ll k_b$, as implied by the physical parameters of the system, the timing at which l unbinding events happen is given by the composition of l exponential decays, described by the Erlang distribution,

$$w_{t|l} = \frac{(k_u t)^{l-1}}{(l-1)!} k_u e^{-k_u t},$$

which results in a distribution of waiting times between transcriptional events (OFF intervals) given by

$$w_t = \sum_l w_{t|l} P_l = \sum_l \frac{(\alpha k_u t)^{l-1}}{(l-1)!} k_u e^{-k_u t} (1-\alpha) = e^{-(1-\alpha)k_u t} k_u (1-\alpha)$$

The average duration of the OFF interval is

$$\tau_{\mathrm{OFF}} = \int_0^\infty t w_t dt = \frac{1}{k_u}\left(1 + \frac{k_b}{k_t}\right)$$

More generally, the distribution of OFF intervals can be computed as the survival probability distribution of the system being in the states B and E before a transcriptional event occurs (27, 28):

$$w_t = -\frac{d}{dt}(P_B + P_E).$$

The dynamics of the corresponding probabilities is given by

$$\frac{d}{dt}P_B = -k_{on} n_L P_1 + k_{on}(n_R + n_L)P_E - k_{\mathrm{off-O1}} P_B$$

$$\frac{d}{dt}P_E = k_{on} n_L P_1 - [k_{on}(n_R + n_L) + k_t]P_E + k_{\mathrm{off-O1}} P_B$$

which forms a closed set of equations when the contribution of P_1 is negligible compared to that of P_2. In that case, we obtain the explicit values $k_u = k_{\mathrm{off-O1}}$ and $k_b = k_{on}(n_R + n_L)$. Using the initial conditions $P_B(0) = 1$ and $P_E(0) = 0$, the resulting distribution of OFF intervals is given by

$$w_t = \frac{k_u k_t e^{-\frac{1}{2}(k_u+k_b+k_t)\left(1+\sqrt{1-\frac{4k_u k_t}{(k_u+k_b+k_t)^2}}\right)t}\left(e^{(k_u+k_b+k_t)\sqrt{1-\frac{4k_u k_t}{(k_u+k_b+k_t)^2}}\,t} - 1\right)}{\sqrt{(k_u+k_b+k_t)^2 - 4k_u k_t}}$$

Since $k_u \ll k_b + k_t$, by expanding the square root, we obtain

$$w_t \simeq \frac{k_u k_t}{k_b + k_t} e^{-\frac{k_u k_t}{k_b + k_t} t}$$

which coincides with the expression obtained previously.

Burst size statistics. To account for multiple transcriptional events in the reduced description, we consider that, after reaching the transcription start state TS and producing a transcript, the system returns instantaneously to the free O_1 state E. The resulting transitions between the reduced states are described by

$$B \underset{k_b}{\leftarrow} E \underset{\infty}{\overset{k_t}{\rightleftarrows}} TS.$$

The probability of making r E-TS transitions before O_1 is occupied is

$$P_r = (1-\alpha)^{r-1}\alpha$$

and the average number of transcripts per burst

$$\langle r \rangle = \sum_r r P_r = \frac{1}{\alpha} = 1 + \frac{k_t}{k_b}.$$

Stochastic simulations. We performed stochastic simulations of the Master equation that describes the dynamics of the system under steady conditions following the approach of Ref. (22). Explicitly, the time evolution of the joint probability $P(p, m, s)$ of the number p of proteins, the number m of mRNA molecules, and the system state s is governed by the Master equation

$$\begin{aligned}\frac{dP(p,m,s)}{dt} = &\sum_{s'} [\, k_{s',s} P(p,m,s') - k_{s,s'} P(p,m,s)] + g_s[P(p,m-1,s) - P(p,m,s)] \\ &+ \lambda_m[(m+1)P(p,m+1,s) - mP(p,m,s)] \\ &+ k_p m[P(p-1,m,s) - P(p,m,s)] + \lambda_p[(p+1)P(p+1,m,s) - pP(p,m,s)],\end{aligned}$$

which takes into account the transitions between transcriptional states, mRNA production, mRNA degradation, protein production, and protein dilution. Here, λ_m is the mRNA degradation rate, k_p is the translation rate, and λ_p is the dilution rate, which equals the growth rate. As discussed in the "Description of the system" section, $k_{s,s'}$ is the transition rate between the transcriptional states of the operon, and g_s is the transcription rate in the state s.

The ON and OFF intervals are computed at discrete times $t_i = i\Delta t$ equally spaced by Δt as $p(t_{i+1}) - p(t_i) > 0$ and $p(t_{i+1}) - p(t_i) \leq 0$, respectively. The burst size is computed as the protein produced during an ON interval.

Computational results. The table below shows the average OFF interval duration, τ_{OFF}, and burst size, $\langle r \rangle$, computed analytically and from stochastic simulations for systems with the WT number of repressors ($n_R = 10$), with 100 times the WT number of repressors ($n_R = 1000$), with DNA looping, and without DNA looping ($K_{\mathrm{Oa}} = 0$ and $n_L = 0$):

Looping	Repressors	τ_{OFF} (min) analytical	τ_{OFF} (min) simulation	$\langle r \rangle$ analytical	$\langle r \rangle$ simulation
Yes	WT	162	160 (154)	1.1	1.5 (1.4)
Yes	100×WT	299	310 (305)	1.0	1.1 (1.1)
No	WT	11.1	22.5 (13.3)	7.9	23.3 (13.0)
No	100×WT	149	159 (153)	1.1	1.2 (1.1)

The parameter values for the analytical computations are described in the "Parameter values" section. Additional parameters in the simulations include $\lambda_m = 0.33\ \mathrm{min}^{-1}$, $\lambda_p = 0.0059\ \mathrm{min}^{-1}$, and $k_p = 26.7\ \mathrm{min}^{-1}$. Averages from simulations were computed over 1.9×10^6 min runs after discarding the initial 3.9×10^5 min. Simulation results were computed by sampling the time series every 5 min (results without parentheses) or 1 min (results within parentheses). The average burst size from the stochastic simulations in the previous table has been normalized by the average protein produced by a single transcript.

Simulations with cell division

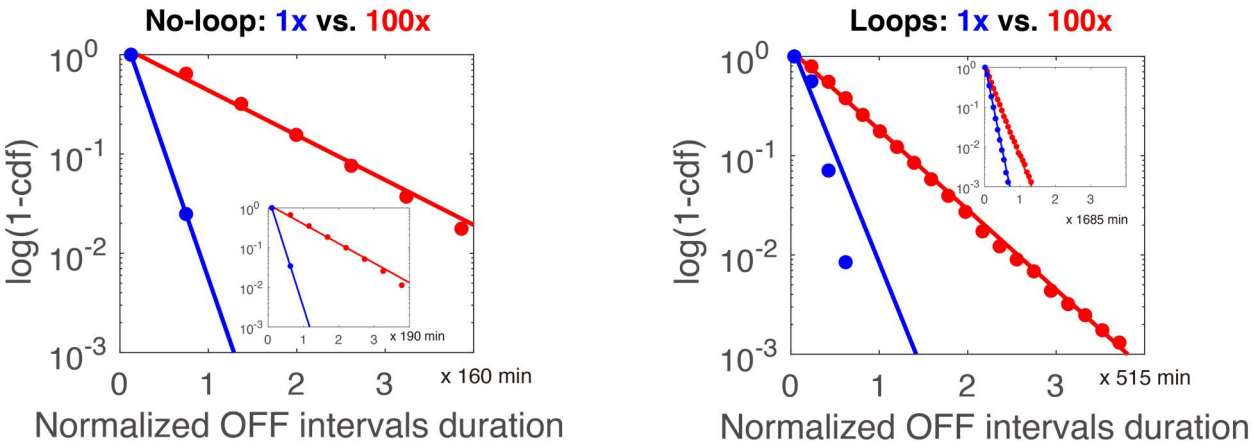

Fig. S8. Statistics of the OFF intervals in the simulations with cell division. We calculated the cumulative distributions of OFF intervals for simulations with cell division by forcing the repressor to unbind from the operators periodically. The insets give the simulations without cell division, which are identical to those in the insets of Fig. 2c.

To assess the effects of cell division on the repression of the *lac* promoter, we performed stochastic simulations that differ from those in the insets of Fig. 2c by forced periodic, simultaneous unbinding events of the repressor from the operators (Fig. S8). In the Loops strain, the statistics of the OFF intervals in the presence of cell division deviates from that of an exponential distribution, and the mean interval duration in the presence of cell division is half shorter than in its absence (Fig. S8, right panel and inset). Furthermore, with DNA looping, the duration of OFF intervals in the presence of cell division are sensitive to repressor concentration changes and decreases by four-fold (Fig. S8, right panel). On the other hand, the effect of cell division on the No-loop strains is very mild (Fig. S8, left panel and inset).

To mimic the effect of the replication fork progression, we made the simplifying assumption in the insets of Fig. 2c that both repressors unbind simultaneously. Alternatively, we also simulated uncoordinated unbinding of the repressors caused by the replication fork crossing the operators at different times. We estimated a delay between the two unbinding times of the order of ~0.3 sec. based on Refs. (29, 30) and a size of the replication fork 1200bp (as measured by the Okazaki fragment length (31)). This more detailed assumption yields almost identical OFF

intervals distribution to that associated with simultaneous unbinding (Fig. S9). For the sake of simplicity, we presented in the main text the simultaneous unbinding results.

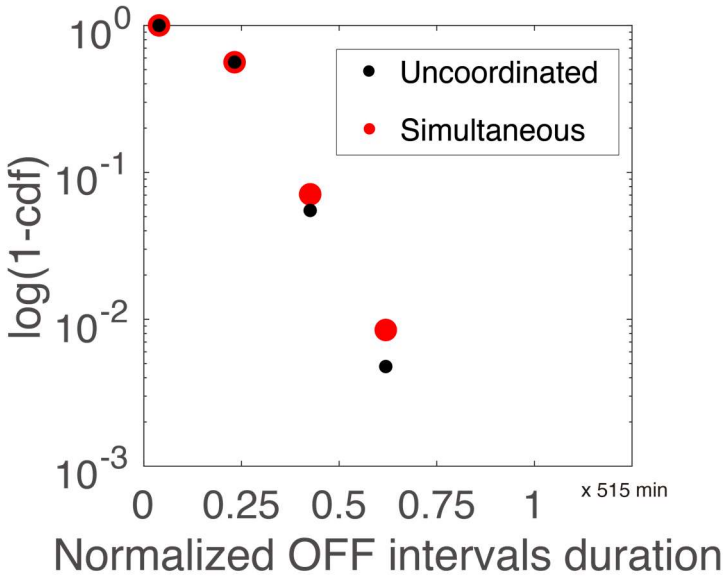

Fig. S9 Simulations that include either uncoordinated unbinding events of LacI due to the replication fork progression or simplified simultaneous unbinding events of LacI from O_1 and O_2.

References

1. N. G. Copeland, N. A. Jenkins, D. L. Court, Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* **2**, 769-779 (2001).
2. S. K. Sharan, L. C. Thomason, S. G. Kuznetsov, D. L. Court, Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* **4**, 206-223 (2009).
3. S. Datta, N. Costantino, D. L. Court, A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109-115 (2006).
4. J. M. Kim, M. Garcia-Alcala, E. Balleza, P. Cluzel, Stochastic transcriptional pulses orchestrate flagellar biosynthesis in Escherichia coli. *Sci Adv* **6**, eaax0947 (2020).
5. C. B. Glascock, M. J. Weickert, Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in Escherichia coli. *Gene* **223**, 221-231 (1998).
6. E. Balleza, J. M. Kim, P. Cluzel, Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods* **15**, 47-51 (2018).
7. P. Hammar *et al.*, The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595-1598 (2012).
8. P. Hammar *et al.*, Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. *Nat Genet* **46**, 405-408 (2014).
9. P. Wang *et al.*, Robust growth of Escherichia coli. *Curr Biol* **20**, 1099-1103 (2010).
10. J. R. Moffitt, J. B. Lee, P. Cluzel, The single-cell chemostat: an agarose-based, microfluidic device for high-throughput, single-cell studies of bacteria and bacterial communities. *Lab Chip* **12**, 1487-1494 (2012).
11. C. C. Sachs *et al.*, Image-Based Single Cell Profiling: High-Throughput Processing of Mother Machine Experiments. *PLoS One* **11**, e0163453 (2016).
12. S. Oehler, E. R. Eismann, H. Kramer, B. Muller-Hill, The three operators of the lac operon cooperate in repression. *EMBO J* **9**, 973-979 (1990).
13. S. Oehler, M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, B. Muller-Hill, Quality and position of the three lac operators of E. coli define efficiency of repression. *EMBO J* **13**, 3348-3355 (1994).
14. L. Cai, N. Friedman, X. S. Xie, Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358-362 (2006).
15. A. Savitzky, M. J. Golay, Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* **36**, 1627-1639 (1964).
16. N. Walker, P. Nghe, S. J. Tans, Generation and filtering of gene expression noise by the bacterial cell cycle. *BMC Biol* **14**, 11 (2016).
17. S. Klumpp, Z. Zhang, T. Hwa, Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366-1375 (2009).
18. J. M. Hudson, M. G. Fried, Co-operative interactions between the catabolite gene activator protein and the lac repressor at the lactose promoter. *Journal of molecular biology* **214**, 381-396 (1990).
19. L. Saiz, J. M. Vilar, Ab initio thermodynamic modeling of distal multisite transcription regulation. *Nucleic Acids Res* **36**, 726-731 (2008).

20. T. Inada, K. Kimata, H. Aiba, Mechanism responsible for glucose-lactose diauxie in Escherichia coli: challenge to the cAMP model. *Genes Cells* **1**, 293-301 (1996).
21. T. Kuhlman, Z. Zhang, M. H. Saier, Jr., T. Hwa, Combinatorial transcriptional control of the lactose operon of Escherichia coli. *Proc Natl Acad Sci U S A* **104**, 6043-6048 (2007).
22. J. M. G. Vilar, L. Saiz, Systems biophysics of gene expression. *Biophys J* **104**, 2574-2585 (2013).
23. J. M. G. Vilar, S. Leibler, DNA looping and physical constraints on transcription regulation. *J Mol Biol* **331**, 981-989 (2003).
24. J. M. G. Vilar, L. Saiz, Reliable Prediction of Complex Phenotypes from a Modular Design in Free Energy Space: An Extensive Exploration of the lac Operon. *ACS Synth Biol* **2**, 576-586 (2013).
25. L. Saiz, J. M. G. Vilar, Ab initio thermodynamic modeling of distal multisite transcription regulation. *Nucleic Acids Res* **36**, 726-731 (2008).
26. D. Kennell, H. Riezman, Transcription and translation initiation frequencies of the Escherichia coli lac operon. *J Mol Biol* **114**, 1-21 (1977).
27. N. G. van Kampen, *Stochastic processes in physics and chemistry*, North-Holland personal library (Elsevier, Amsterdam ; London, ed. 3rd, 2007).
28. N. F. Polizzi, M. J. Therien, D. N. Beratan, Mean First-Passage Times in Biology. *Isr J Chem* **56**, 816-824 (2016).
29. T. M. Pham *et al.*, A single‐molecule approach to DNA replication in E scherichia coli cells demonstrated that DNA polymerase III is a major determinant of fork speed. *Molecular microbiology* **90**, 584-596 (2013).
30. M. M. Elshenawy *et al.*, Replisome speed determines the efficiency of the Tus-Ter replication termination barrier. *Nature* **525**, 394-398 (2015).
31. L. Balakrishnan, R. A. Bambara, Okazaki fragment metabolism. *Cold Spring Harb Perspect Biol* **5** (2013).