Electronic Supplementary Material (ESI) for Molecular BioSystems

Supplementary material for "Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of *Escherichia coli*"

Jason Lloyd-Price, Maria Lehtivaara, Meenakshisundaram Kandhavelu, Sharif Chowdhury, Anantha-Barathi Muthukrishnan, Olli Yli-Harja, and Andre S. Ribeiro

Table of Contents

1. Introduction

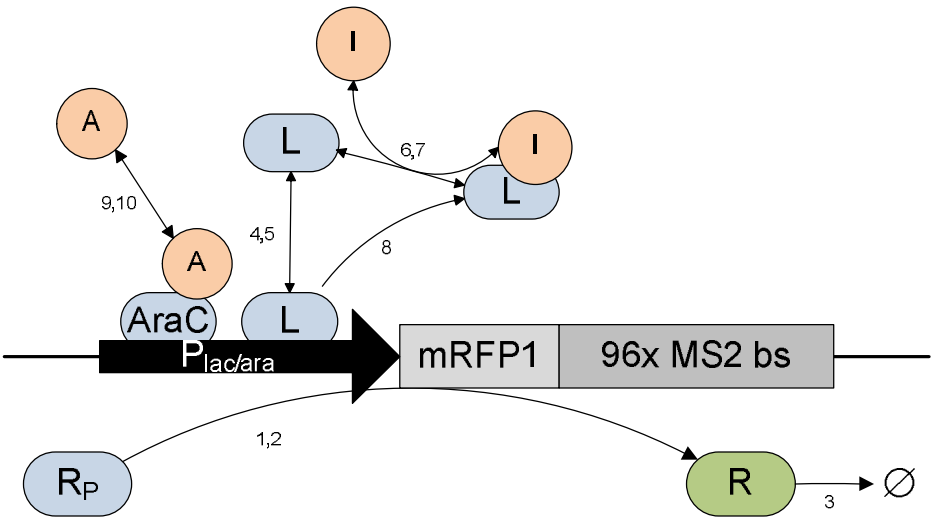
This document describes the delayed stochastic model of the dynamics of the $P_{lac/ara}$ promoter (Golding and Cox, 2004) as well as the tuning procedure to match the mean expression levels measured experimentally. Our model follows the delayed stochastic modeling strategy proposed in (Ribeiro et al, 2006), and is implemented in SGNSim (Ribeiro and Lloyd-Price, 2007). An initial model of this genetic system at the single cell level is first presented where the relevant chemical components are represented explicitly (section 2). From that complete model, we then introduce a reduced version of the model by

approximating some components in order to facilitate to tuning and prediction, without affecting the relevant dynamics of the system (section 3). The complete set of reactions of the reduced model is presented in section 3.4, while the values used for the parameters are presented in section 4.5.

Afterwards, we present results concerning the spatial distribution of free MS2d-GFP molecules in live *E. coli* cells (section 5), and details on the method of quantification of mRNA target for MS2d-GFP in life cells (section 6). Finally, we present an example image of cells taken by epifluorescence microscopy (section 7) and supporting information regarding the degree of synchronization of cell division following heat shock (section 8).

2. Explicit Model

The chemical components and interactions are depicted in Supplementary Fig. 1. The production of RNA molecules (and thus fluorescent spots in the cells) is accomplished when an RNA polymerase binds to the promoter region and transcribes the DNA into RNA. Transcription is regulated by a repressor protein (LacI) and inducers (IPTG and Arabinose). IPTG can bind to LacI, causing a conformational change in LacI which results in the protein falling off the promoter, allowing transcription to occur (Lutz and Bujard, 1997). Another protein (AraC) can also bind to the promoter. This protein does not modify the transcription initiation rate on its own, but when bound to Arabinose, it increases the affinity of the RNA polymerases to the promoter region, thereby promoting transcription. GFP molecules are not explicitly represented since they always exist in sufficient amounts so as not to be a limiting factor of RNA detection.



Supplementary Fig. 1: Components and reactions in the model of the $P_{lac/ara}$ promoter and RNA production. Numbers indicate the corresponding reactions. Molecule labels are defined in section 2.1.

2.1. Notation

Hereafter, we use *P* to denote the promoter, R_P to denote RNA polymerase, *R* to denote RNA, *L* to denote LacI, *I* to denote IPTG, and *A* to denote Arabinose in reactions and equations.

The model contains several reversible bimolecular reactions. In general, rates of unbinding are denoted by k_{uXY}, where X and Y denote the interacting species. Dissociation constants (K_{dXY}) are denoted similarly. For example, K_{dLI} is the dissociation constant between LacI and IPTG.

The promoter can be in several 'states', depending on which substances are bound to it. In the following reactions, the promoter with LacI bound to it is denoted P_L, whereas $P_{L;^-}$ denotes the promoter with no LacI bound. This notation also applies to Arabinose, where P_A and $P_{A;^-}$ denote the presence or absence of an Arabinose molecule bound to an AraC protein which is in turn bound to the promoter. For simplicity, we assume that an AraC is always bound to the promoter, given its small dissociation constant (on the order of 10^{-8} M (Timmes et al, 2004)). Similarly, $P_{L;^-A}$ denotes a promoter with no LacI bound, but with Arabinose bound to it.

Following standard chemical notation, the number of molecules of a chemical species present in a cell is denoted $[X]$. For example, $[L]$ denotes the current number of LacI proteins in the cell. Additionally, we use $[X]_0$ to denote the number of a given molecule in the cell at time 0 (the start of the experiment).

2.2. Transcription

Transcription is initiated when an RNAP binds to the promoter region, forms the open complex, and begins elongating the RNA. This process is modeled in reaction (1) (Ribeiro et al, 2006).

$$P_{\bar{L}\bar{A}} + R_P \xrightarrow{k_t} P_{\bar{L}\bar{A}}(\tau_P) + R_P(\tau_R) + R(\tau_R) \qquad (1)$$

Here, τ_P and τ_R denote time delays used to model the time it takes for this highly complex, multi-step reaction to occur (Ribeiro et al, 2006). This notation denotes that, for example, the RNA (R) is fully transcribed and visible in the cell τ_R seconds after the transcription reaction began.

When Arabinose binds to the AraC molecule bound to the promoter, it induces transcription by actively recruiting RNA polymerases. This is modeled by a reaction which differs from reaction (2) in the value of its rate constant ($k_{tA} > k_t$).

$$P_{\bar{L}A} + R_P \xrightarrow{k_{tA}} P_{\bar{L}A}(\tau_P) + R_P(\tau_R) + R(\tau_R) \qquad (2)$$

Normally, the RNA molecules are assumed to degrade via a first-order chemical reaction. However, the MS2-coated RNA molecules have been shown to have a considerably longer lifetime than normal RNA molecules, and cell division was shown to be the largest term in the (Golding et al, 2005).

2.3. Decay

The degradation of transcripts is modeled as a first-order process in reaction (3) (Ribeiro et al, 2006):

$$R \xrightarrow{k_d} \varnothing \qquad (3)$$

2.4. Interactions of the promoter with LacI and IPTG

LacI's binding/unbinding to/from the promoter is modeled by reactions (4) and (5). When LacI is bound to the promoter, transcription cannot occur.

$$P_{\bar{L}} + L \xrightarrow[K_{dLP}]{k_{uLP}} P_L \quad (4)$$

$$P_L \xrightarrow{k_{uLP}} P_{\bar{L}} + L \quad (5)$$

IPTG can bind to LacI (reactions (6) and (7)). The complex *IL* represents LacI with a different conformation that has much weaker affinity to the promoter than LacI. In reality, the stability of the bond between this complex and the promoter is much weaker. For simplicity, this is modeled by an immediate dissociation from the promoter (reaction (8)).

$$I + L \xrightarrow[K_{dLI}]{k_{uLI}} IL \quad (6)$$

$$IL \xrightarrow{k_{uLI}} I + L \quad (7)$$

$$P_L + I \xrightarrow[K_{dLI}]{k_{uLI}} P_{\bar{L}} + IL \quad (8)$$

2.5. Interactions of the promoter with AraC and Arabinose

Since an AraC is assumed to be bound at all times to the promoter, we only model the binding and unbinding of the Arabinose to AraC. This event changes the rate of transcription initiation. The binding and unbinding of Arabinose to AraC is modeled in reactions (9) and (10), respectively.

$$P_{\bar{A}} + A \xrightarrow[K_{dAP}]{k_{uAP}} P_A \quad (9)$$

$$P_A \xrightarrow{k_{uAP}} P_{\bar{A}} + A \quad (10)$$

3. Reduced Model

The explicit model contains several parameters that are currently difficult to measure *in vivo*, and some reactions which, under normal conditions, do not significantly affect the dynamics of gene expression. To reduce the complexity of the model, we implement some approximations and justify why they are appropriate and do not compromise the realism of the simulation. Generally, these approximations consist of removing a reaction species which is in sufficient abundance to be considered constant.

3.1. RNA Polymerase

Under normal conditions, the amount of R_P available for transcription events is approximately constant in an *E. coli* at all times (McClure, 1983). For that reason, instead of representing R_P explicitly, the stochastic rates of reactions (1) and (2) can be multiplied by 20, the known quantity of free RNA polymerases per gene in *E. coli*.

3.2. Arabinose

Similar to R_P, when Arabinose is present in the cell, it is present sufficient amount so as to be assumed as constant. With 6.67 mM (1%) of Arabinose, and given the mean volume of an *E. coli* (10^{-15} L, from http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi), we estimate that the number of molecules of Arabinose in the cell is on the order of 10^{10}. Reactions (9) and (10) can then be simplified to:

$$P_{\bar{A}} \xrightarrow{\frac{[A]k_{uAP}}{K_{dAP}}} P_A \tag{11}$$

$$P_A \xrightarrow{k_{uAP}} P_{\bar{A}} \tag{12}$$

This simplification has the additional important benefit that the Arabinose concentration and K_{dAP} no longer need to be expressed in units of molecules per cell. Instead, both can be expressed as a ratio of 6.67 mM, decreasing the number of parameters required in the model (the cell volume is removed).

3.3. LacI/IPTG

Just as Arabinose and R_P are present in the system in abundance, so are LacI and IPTG. These species can be removed with the same benefits as removing Arabinose, by the equilibrium point of reactions (6) and (7). The equilibrium values can be then used in simplified versions of reactions (4), (5) and (8). From reactions (6) and (7), this equilibrium point is reached when:

$$[I][L] = [IL]K_{dLI} \tag{13}$$

Assuming that are no IPTG·LacI complexes initially present in the system, the amount of the complex can be written in terms of the initial concentrations of LacI (denoted by $[L]_0$). Similarly, the amount of IPTG can also be written as a function of LacI.

$$[IL] = [L]_0 - [L] \tag{14}$$

$$[I] = [L] + [I]_0 - [L]_0 \tag{15}$$

The solution for $[L]$ is:

$$[L] = \frac{-b \pm \sqrt{b^2 + 4K_{dLI}[L]_0}}{2}, b = [I]_0 - [L]_0 + K_{dLI} \tag{16}$$

These calculations remove the need for reactions (4) and (5), and the equilibrium concentrations of $[L]$ and $[I]$ can be inserted into the reaction rates of reactions (6), (7) and (8).

This approximation yields two advantages. First, as with Arabinose, it is no longer necessary to translate IPTG and LacI concentrations and K_{dLP} into molecules per cell. Second, we have sped up the simulation considerably by removing two high-frequency reactions, since the runtime of the SSA largely depends on the propensity of the reactions.

The validity of this approximation depends on the actual rate of k_{uLI} (and therefore the time it takes to reach equilibrium), and the amounts of LacI and IPTG. In vitro studies have measured k_{uLI} to be 0.2 s^{-1} (Dunaway et al, 1980), implying that the system should be sufficiently close to equilibrium within one minute. The amount of LacI proteins in these cells has been measured to be on the order of 5000

(Lanzer and Bujard, 1988). Based on the average volume of an *E. coli* (10^{-15} L, from http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi), we estimate that the amount of molecules of IPTG in the cell when induced by 1mM of IPTG is on the order of 10^9. All the conditions for the applicability of this approximation are therefore met by at least one order of magnitude.

3.4. The Final Model

The final delayed stochastic model consists of reactions (17) to (24):

$$P_{\overline{LA}} \xrightarrow{[R_p]k_t} P_{\overline{LA}}(\tau_P) + R(\tau_R) \tag{17}$$

$$P_{\overline{LA}} \xrightarrow{[R_p]k_{tA}} P_{\overline{LA}}(\tau_P) + R(\tau_R) \tag{18}$$

$$R \xrightarrow{k_d} \varnothing \tag{19}$$

$$P_{\overline{L}} \xrightarrow{\frac{[L]k_{uLP}}{K_{dLP}}} P_L \tag{20}$$

$$P_L \xrightarrow{k_{uLP}} P_{\overline{L}} \tag{21}$$

$$P_L \xrightarrow{\frac{[I]k_{uLI}}{K_{dLI}}} P_{\overline{L}} \tag{22}$$

$$P_{\overline{A}} \xrightarrow{\frac{[A]k_{uAP}}{K_{dAP}}} P_A \tag{23}$$

$$P_A \xrightarrow{k_{uAP}} P_{\overline{A}} \tag{24}$$

4. Constants and Tuning

4.1. Promoter Occupancy

Reaction pairs (20)-(21) and (23)-(24) determine the occupancy state of the promoter. It is useful in this case to precisely define the quantity 'Promoter Occupancy' as the expected fraction of time that the promoter is bound by LacI, or Arabinose (by the AraC/Arabinose complex). This fraction can be calculated from the propensities of the reactions as:

$$O(X, K_d) = \frac{X}{X + K_d} \tag{25}$$

In (25), X denotes the concentration of the binding molecule, and K_d is the dissociation constant. For example, $O([L]_0, K_{dLP})$ is the fraction of time that the lac repressor is bound to the promoter in the absence of IPTG.

4.2. Accounting for degradation and the promoter open complex formation duration (τ_P)

To calculate the values of the transcription rate constants (k_t and k_{tA}) necessary to obtain a given mean level of RNA after a transient time t, we first estimate the mean field behavior of a system

producing RNA with rate k_p, which can degrade with rate k_d, without accounting for the effects of τ_P as a limiting factor of transcription initiation. The differential equation describing such mean field behavior is:

$$\frac{d[R]}{dt} = k_p - [R]k_d \tag{26}$$

Solving for k_p yields:

$$k_p([R]) = \frac{[R]k_d}{1-e^{-tk_d}} \tag{27}$$

Finally, the values of the transcription rate constants (k_t and k_{tA}) are not exactly equal to k_p due to the effects of the promoter open complex formation. This effect can be accounted for by discounting the expected time that the promoter spends in this state, giving the mean rate at which the transcription reaction must occur to reach a mean RNA count of [R] after t seconds:

$$k([R]) = \frac{k_p([R])}{1-\tau_p k_p([R])} \tag{28}$$

We can now use (28) to calculate the necessary transcription rate to reach a given mean amount of RNA molecules after a transient.

4.3. Tuning k_t, k_{tA}, K_{dLP}

We now tune some of the free parameters (not yet experimentally measured) of the model to match the experimental results. The parameters k_t and k_{tA} determine the maximum possible RNA production rate, while K_{dLP} determines how strongly LacI represses the system.

First, we deduce the production rate of RNA, according to reactions (17) and (18). The fraction of time that the promoter is, on average, not repressed by LacI (and thus is free for these reactions) is given by (1-O([L],K_{dLP})). Similarly, the fraction of time that the promoter is in state P_A is given by O([A],K_{dAP})). The mean production rate of RNA is a mixture of the propensities of the two reactions that lead to RNA production, weighted by the fraction of times each is expected to occur:

$$k([R]) = [R_P](1-O([L],K_{dLP}))[k_t(1-O([A],K_{dAP})) + k_{tA}O([A],K_{dAP})] \tag{29}$$

We can now write formulas for the mean production rates of three cases: full repression of the promoter ($k_1 = k([R]_1)$, where $[R]_1$ was measured to be 0.532 RNAs), activation by Arabinose alone ($k_A = k([R]_A)$, where $[R]_A$ was measured to be 0.612 RNAs), and full activation ($k_M = k([R]_M)$, where $[R]_M$ was measured to be 3.36 RNAs).

$$k_1 = [R_P]k_t(1-O([L],K_{dLP})) \tag{30}$$

$$k_A = [R_P](1-O([L]_0,K_{dLP}))[k_t(1-O(1,K_{dAP})) + k_{tA}O(1,K_{dAP})] \tag{31}$$

$$k_M = [R_P][k_t(1-O(1,K_{dAP})) + k_{tA}O(1,K_{dAP})] \tag{32}$$

Solving the system of equations, we get:

$$k_t = \frac{k_M k_1}{[R_P]k_A} \tag{33}$$

$$K_{dLP} = \frac{[R_P]k_t[L]_0}{[R_P]k_t - k_1} - [L]_0 \tag{34}$$

$$k_{tA} = \frac{k_M - k_t\left(1 - O\left(1, K_{dAP}\right)\right)}{[R_P]O\left(1, K_{dAP}\right)} \tag{35}$$

4.4. Cell Division

The simulation of multiple cells, subject to cell division, is modeled by the CellLine simulator (Ribeiro et al, 2007), which can model cell division and impose desired distributions of partitioning of RNA molecules between the daughter cells at cell division. However, all calculations thus far have assumed that there is no cell division. To obtain the correct mean production rate with asynchronous division, the value of k_d in section 4.2 is increased by adding $\frac{\ln 2}{g}$, where g is the generation time.

4.5. Values of Constants and Parameters

Constant	Value	Source
$[R_P]$	20 molecules	McClure (1983)
t	3600 – $E(\tau_R)$ = 3465.9 s	Ourselves
τ_P	32[*] s	Lutz and Bujard (1997)
τ_R	τ_P + Γ(Gene Length, Elongation Rate^{-1}) s	
Gene Length	mRFP1 Length + 96 BS Length = 4287 bp	
mRFP1 Length	654 bp	Zhang et al (2002)
96 BS Length	3633 bp	Golding and Cox (2004)
Elongation Rate	42 bp·s^{-1}	Gotta et al (1991)
$[L]_0$	5000 molecules = 8.3 × 10^{-3} × 1x $[I]$	Lanzer and Bujard (1988)
k_{uLP}	0.04 s^{-1}	Dunaway et al (1980)
K_{dLI}	0.1[†] × 1x $[I]$	Lutz and Bujard (1997)
k_{uLI}	0.2 s^{-1}	Dunaway et al (1980)
K_{dAP}	0.1[†] × 1x $[A]$	Lutz and Bujard (1997)
k_{uAP}	1.5 s^{-1}	Miller et al (1983)

Supplementary Table 1: Constants.

Parameter	Value	Source/Section
k_d	0 s^{-1}	Golding et al (2005)
k_t	4.191 × 10^{-5} $[R_P]^{-1}$s^{-1}	Section 4.3

[*] Since the transcription events are rare (a few per cell lifetime), we do not model the promoter delay as a distribution, but rather as a constant value. We observed no significant difference when τ_P followed a distribution with realistic variance (Lutz and Bujard, 1997).

[†] Approximated from the induction curve from Lutz and Bujard (1997)

k_{tA}	4.886×10^{-5} $[R_P]^{-1}s^{-1}$	Section 4.3
K_{dLP}	$1.879 \times 10^{-3} \times$ 1x $[I]$	Section 4.3

Supplementary Table 2: Model tuning in Fig. 5.

Parameter	**Value**	**Source/Section**
k_d	1/600 s^{-1}	Bernstein (2002)
g	1800 s	Section 4.4
k_t	3.727×10^{-4} $[R_P]^{-1}s^{-1}$	Section 4.3
k_{tA}	4.370×10^{-4} $[R_P]^{-1}s^{-1}$	Section 4.3
K_{dLP}	$1.492 \times 10^{-3} \times$ 1x $[I]$	Section 4.3

Supplementary Table 3: Model tuning in Table 2 and subsequent figures.

5. Uniformity of the MS2d-GFP distribution

The MS2d-GFP tagging proteins are expressed from a strong promoter ($P_{LtetO-1}$) on a high-copy number plasmid (PROTET-K133). Within the duration of our measurements we observed that there was always enough MS2d-GFP in the cells to properly detect all target RNA molecules. This can be assessed by measuring the uniformity of the background fluorescence. This assessment further shows that if partitioning errors in the MS2d-GFP exist, they have negligible effects on the detection of target RNA in daughter cells.

The simplest way to make this assessment is to quantify the uniformity of the fluorescence background of cells with no target RNA molecules. We first establish a measure of uniformity based on the local spatial entropy of the fluorescence in the image each cell and then show that it is able to detect clumping and gradients in model cells with clumps and gradients. The measure is then applied to cells expressing MS2d-GFP without the target RNA to determine if the MS2d-GFP clumps and/or tends to be localized in any particular region of the cell.

5.1. Clumping and spatial distribution of MS2d-GFP in the cells

To determine if the MS2d-GFP molecules form clumps and/or tend to be co-localized in any particular region of the cell, we compute the local spatial entropy of the intensity of the pixels composing a bacterium, which allows us to quantify the degree of randomness of a set of variables (Shannon 1948). Here, we aim to show that all neighborhoods of pixels within the cell have similar distributions of fluorescence intensities among the pixels within each neighborhood, that is, that they have no detectable gradients or clumps of MS2d-GFP. The entropy H_k in a neighborhood of k pixels is defined as:

$$H_k = \sum_{\mathbf{y} \in \Omega^k} p(\mathbf{y}) \log(p(\mathbf{y})) \tag{36}$$

where $\mathbf{y}$ is the vector of pixel intensities in the neighborhood, Ω is the domain of the elements of $\mathbf{y}$, and $p(\mathbf{y})$ is its probability measure.

Entropy H_1 informs us of how diverse the intensities of the pixels composing the cell are, but not whether there are correlations between the intensities of neighboring pixels. If MS2d-GFP molecules clump or tend to be preferentially located in any particular region of the cell, these correlations ought to exist. If such correlations do not exist, then the entropy of the joint distribution of the intensities of

neighboring pixels should equal kH_1 (since the pixel intensities are independent), otherwise, it is smaller than this value. The minimum possible (totally correlated pixel intensities) H_k equals the entropy of a single pixel H_1. A simple way to quantify effects of possible spatial correlations or gradients in pixel intensities is then the ratio between H_k and H_1. To have a normalized measure of correlation, we define J_k, ranging from k^{-1} to 1, as:

$$J_k = \frac{H_k}{kH_1} \tag{37}$$

We measure the two-pixel neighborhood entropy H_2 from both vertical and horizontal pairs of adjacent pixels. For each cell segmented from the images from the microscope, we first subtract the mean pixel intensity from each pixel, and scale the resulting intensities by the variance of the distribution of pixel intensities. The scaled pixel intensities were then binned into bins of size 0.2 per unit variance. The aforementioned scaling is required for the entropies of different cells to be comparable, due to the effects of the binning.

5.2. Generating spatial patterns and clumpiness in model cells

Model cells with various degrees of clumpiness are generated from cell shapes taken from real cells with the following algorithm. Let the *x* axis correspond to the major axis of the cells, and I(x,y) be the gray level of the pixel at (x,y). The algorithm proceeds as follows:

1. Set all I(x,y) = 0
2. Repeat N times:
    a. Select x_c, y_c uniformly from the pixels in the real cell.
    b. Set $I(x_c, y_c) = I(x_c, y_c) + 1$
3. Convolve I with a Gaussian kernel with standard deviation σ

The degree of clumpiness of model cells is then determined by the choice of parameters N and σ. Larger N and/or larger σ produce less clumpiness. We use two sets of values of (N, σ), namely, (25, 1) and (100, 3), as these produce different, both detectable, degrees of clumpiness.

We also model cells with gradients, where the gradients aim to mimic what would be observed if the MS2d-GFP molecules were preferentially localized near, e.g., the cell poles or approximately along the cell border. Gradients are generated as follows:

$$I(x, y) = c(x - x_0)^p + (y - y_0)^p \tag{38}$$

where x_0 and y_0 are the coordinates of the cell center and, *p* determines the order of the gradient. To attain a linear gradient we set *p* to 1. For a quadratic gradient, we would set *p* to 2. By changing *c*, the eccentricity of the gradient can be varied.

5.3. Null model cells with no spatial correlations between pixel intensities

Null model cells without spatial correlations between pixel intensities can be generated by, for each pixel, generating an intensity value, drawn from the distribution of pixel intensities of a real cell.

However, a procedure is necessary prior to this, since real cells have two external sources of spatial correlations in the intensities of neighbor pixels, regardless of the existence or not of clumping or accumulation of MS2d-GFP molecules at any region of the cell. One source is the point spread function of the microscope. Since its effects cannot be removed, we expect slightly higher local correlations in real cells than in null model cells. The other source is the rod shape of the cells, whose effect can be accounted for in null model cells for proper comparison.

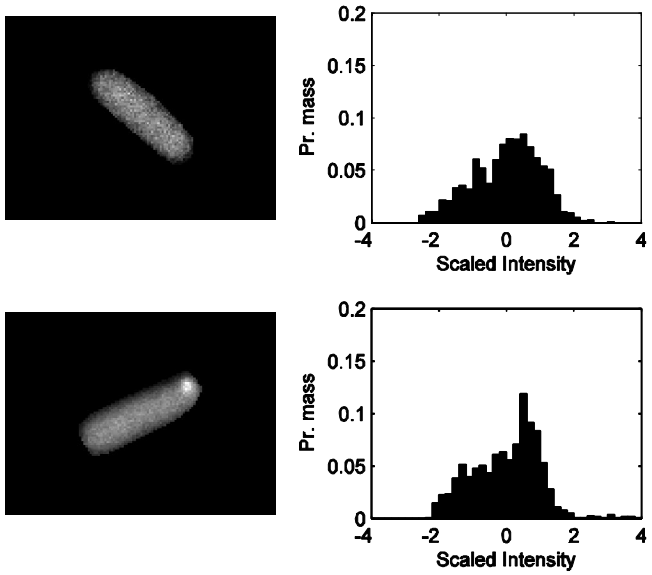
To generate null model cells from real ones that account for the shape of a cell, first, we remove the effect of the rod-shape from the pixels of a real cell by dividing by a scaling factor (described below). From the distribution of resulting pixel intensities, we generate the intensity of each pixel of the model cell. Next, we reintroduce the effect of the rod shape into the pixel intensities of the model cell by multiplying each pixel by the scaling factor. This allows null model cells to be generated that lack spatial correlations in pixel intensities except due their shape. The scaling factor to account for the rod shape of the cells for the pixel at (x,y) is:

$$s(x,y) = \sqrt{\frac{d_{\min}(x,y)}{\max(d_{\min})}} \tag{39}$$

where $d_{min}(x,y)$ is the Euclidean distance to the nearest pixel that is not in the cell.

5.4. Clumping of MS2d-GFP

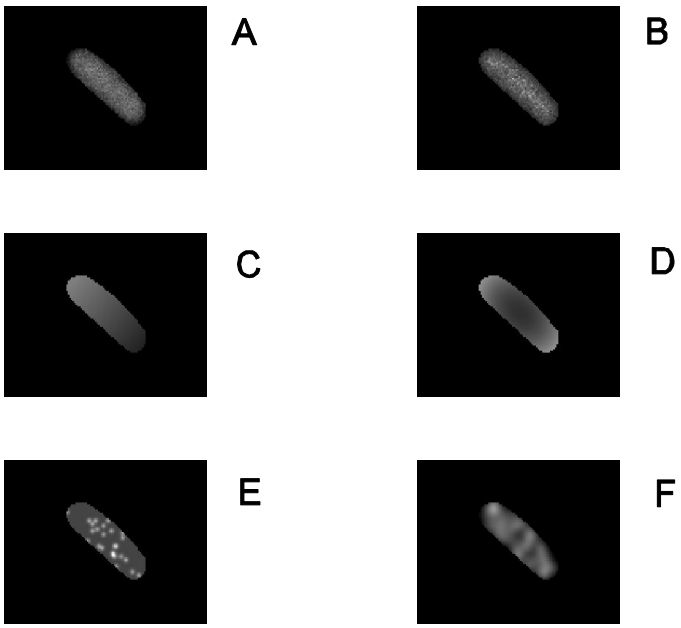
We now study whether there is a tendency of MS2d-GFP molecules to accumulate at any particular region of the cell or to clump in vivo. We start by performing a test on the method of detection of spatial correlations from images of cells taken by confocal microscopy. If the measure of J_k of a cell is accurate enough to detect spatial correlations due to clumping of MS2d-GFP, its value should differ measurably between a cell with no target RNA and a cell with a target RNA, to which ~60 MS2d-GFP molecules are bound at any moment (Golding and Cox, 2004). Supplementary Fig. 2 shows the image of a cell with no visible RNA-MS2d-GFP spot (top left). Also shown is its distribution of scaled pixel intensities (top right). The image of a cell with one RNA-MS2d-GFP spot is also shown (bottom left) along with its distribution of scaled pixel intensities (bottom right).



Supplementary Fig. 2: A cell with no visible RNA-MS2d-GFP spot (top left) and its scaled pixel intensities (top right). Also shown is a cell with one RNA-MS2d-GFP spot (bottom left) and its distribution of scaled pixel intensities (bottom right).

Comparing the distributions in Supplementary Fig. 2, it is visible that the RNA-MS2d-GFP spot creates a small peak in the highest intensities in comparison to the cell with no spot. More spots would further increase the height of this peak and thus the difference between the two distributions. The values of J_2 are 0.91 for the cell with no spot and 0.84 for the cell with one spot. As shown below, and given that this measure varies from 0.5 to 1, this difference is significant, allowing the detection of MS2d-GFP clumps or gradients, if these exist (regardless of their origin).

In Supplementary Fig. 3 we show the image of a real cell (5A) and of a null model cell generated from this real cell (5B). Also shown are model cells, one with a linear gradient along the major axis (5C), and another with an eccentric quadratic gradient that results in stronger pixel intensities near the cell poles (5D). Two model cells with different degrees of artificial clumpiness and their distributions of pixel intensities are also shown (5E and 5F).



Supplementary Fig. 3. Images of a real cell (A) and of a null model cell (B) generated from the real cell. Also shown are model cells, one with a linear gradient along the major axis (C) and another with a quadratic gradient that creates stronger pixel intensities near the poles (D). Two model cells with different degrees of artificial clumpiness are also shown (E and F).

Each model cell is used as a null model to test for the existence of a type of pattern in the spatial distribution of MS2d-GFP molecules in real cells. The cells generated by randomly choosing the pixel intensities from the original distribution of pixel intensities are used to determine, by comparison, if there are local correlations between the intensities of neighbor pixels that are not detectable by eye. If no such local pixel intensity correlations exist, cells and model cells with random pixels intensities ought to have identical values of J_k.

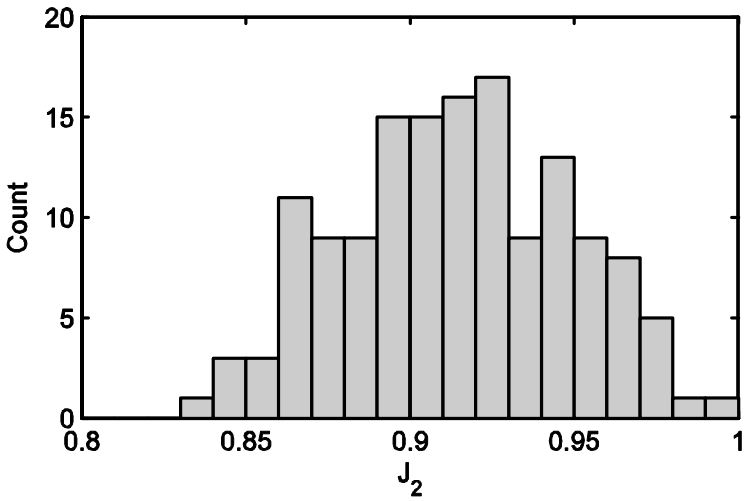
The model cells with gradients are used as a null model for possible accumulation of MS2d-GFP molecules at any particular location of the cell. If such preferential locations were to exist, they would result in gradients of pixel intensities in the real cells and cause J_k to be lower than in the cells with randomized pixel locations.

Finally, the model cells with artificial clumps are used as a null model for possible accumulation of MS2d-GFP molecules at certain locations in the cell. If these exist, they would cause J_k to be lower than in the cells with randomized pixel locations. The values of J_2 of the cells A to F depicted in Supplementary Fig. 3 are shown in Supplementary Table 4. For this particular cell, it is possible to conclude that there are no spatial correlations between neighboring pixel intensities, as its value of J_2 is identical to that of the model cell with randomized pixel locations. Since J_2 is much higher than the J_2 of

the model cells with artificial gradients and clumps, we can also conclude that this cell has no gradients or clumps.

Cell	H_2	H_1	J_2
Real Cell, no spot	5.44	2.98	0.91
Real Cell, one RNA-MS2d-GFP spot	4.88	2.92	0.84
Null model cell	5.44	2.97	0.92
Linear gradient	3.58	2.92	0.61
Quadratic gradient	3.71	2.79	0.66
Small artificial spots	2.76	1.78	0.77
Large artificial spots	4.74	2.98	0.79

Supplementary Table 4: Values, for the cell and the null-model cells depicted in Supplementary Fig. 3, of their entropy in a two-pixel neighborhood (H_2), their entropy of individual pixels (H_1) and the value of J_2, a measure of spatial correlations, attained from the ratio between H_2 and H_1.



Supplementary Fig. 4: Distribution of J_2 for cells induced only with aTc (data is from 145 cells).

Supplementary Fig. 4 shows the distribution of J_2 values for 145 cells induced with only aTc for one hour. The mean value of J_2 of each of these cells is 0.914. We generated 10 models cells, from each of these 145 cells, with pixel intensities randomly drawn from the distribution of pixel intensities of the real cell. The mean value of J_2 of the 1450 model cells is 0.93, identical to that of the real cells. This demonstrates that there is no indication of accumulation of MS2d-GFP molecules at any particular region of the cells, or formation of clumps, one hour after induction by aTc.

Further note that the value of J2 for the cell with a target RNA (spot), shown in Table 2 is in fact smaller than more than 95% of the values in the distribution shown in Supplementary Fig. 4.

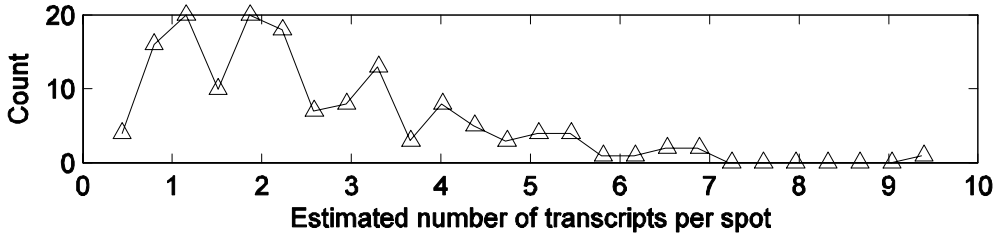
A final test can be made to further verify these conclusions. If MS2d-GFP molecules do not form clumps or accumulate at any region of the cells, then the value of J_2 of cells measured over a long period of time ought to be constant (aside from small stochastic fluctuations). We imaged cells at t = 326 s, t = 1791 s, and t = 3591 s after placed under the microscope, and measured J_2 (Table 2). The results show that this quantity does not change significantly over time in any cell, further verifying that MS2d-GFP molecules neither tend to accumulate at any particular region of the cell, nor aggregate.

Cell index	J_2 at t = 326 s	J_2 at t = 1791 s	J_2 at t = 3591 s
1	0.89	0.88	0.89
2	0.87	0.87	0.87
3	0.88	0.88	0.87
4	0.89	0.88	0.89
5	0.89	0.88	0.88
6	0.87	0.86	0.87
7	0.89	0.89	0.89

Supplementary Table 5: Values of J_2 of cells at t = 326, t = 1791, and t = 3591 s.

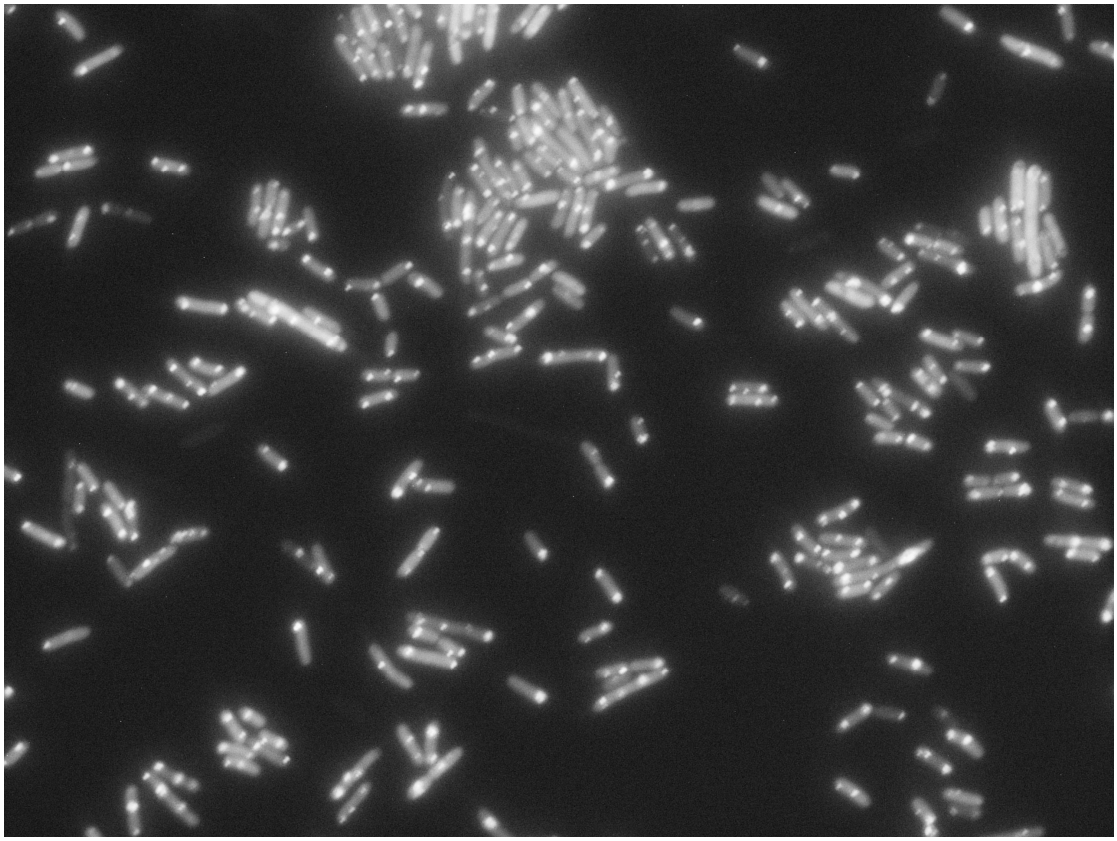
6. Quantification of mRNA in cells

The RNA quantification method used here was proposed in (Golding et al, 2005). The number of RNA molecules in each spot is quantified by assuming that the first peak in the distribution of intensities of many RNA spots from cells on the same slide corresponds to individual RNA molecules. The intensities are then normalized by the intensity of this peak to obtain the number of RNA molecules in each spot. This is possible due to the discrete nature of the peaks and the approximately uniformity of the distance between consecutive peaks. An example of such a distribution of intensities is shown in Supplementary Fig. 5.



Supplementary Fig 5: Example distribution of spot intensities obtained from a single slide, normalized by the mean intensity of the first peak in the distribution which corresponds to a single tagged RNA molecule.

7. Example image of cells expressing MS2d-GFP



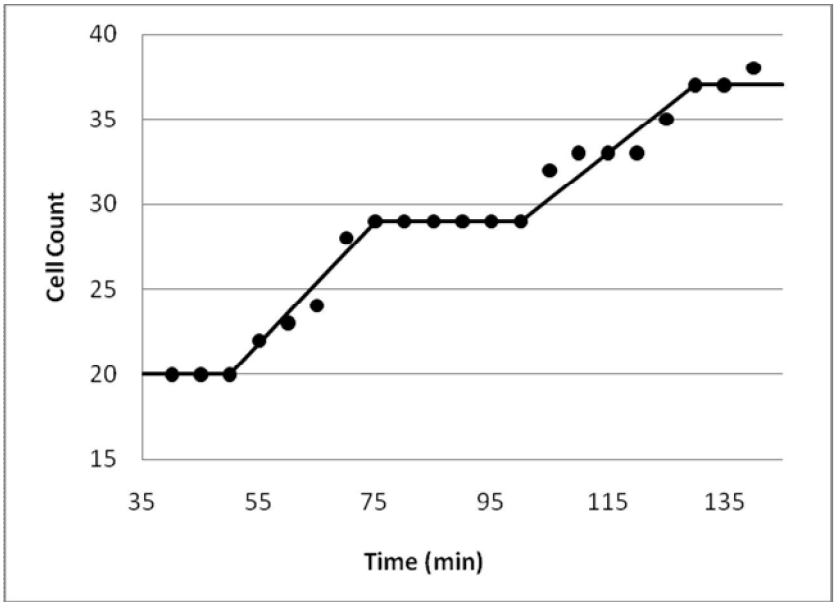
Supplementary Fig 6: Cells expressing MS2d-GFP and RNA target. Bright spots in each cell correspond to RNA molecules tagged with ~60-100 MS2d-GFP molecules. Cellular background is also fluorescent due to the freely diffusing MS2d-GFP molecules.

8. Assessing the degree of synchrony of cells following a heat shock

To determine the degree of synchrony in division of cells subject to heat shock, we redid the experiment as described in Materials and Methods, except that 30 minutes after induction by IPTG, 8 μL of culture was placed between a 1% LB agarose gel and a microscope cover slip. Starting from 40 minutes after induction, images of a set of cells were taken by DIC every 5 minutes for the following 100 minutes. Cells were held at 37^{o}C while under the microscope.

The number of cells visible at each point in time is shown in Supplementary Fig. 7. Two approximately synchronous divisions were observed (between 55 minutes and 75 minutes and between 105 minutes and 130 minutes), indicating that the heat shock successfully synchronized the divisions with the same efficiency as reported in (Lomnitzer and Ron, 1972). The divisions appear slightly later than the jumps in CV of RNA numbers observed in the synchronous experiment (see main document). This is expected because the imaging procedure for each time point of the synchronized experiment took approximately 10 minutes (this includes obtaining the cells from the liquid culture, placing them under the microscope, etc…). From the figure, it is observable that the fraction of cells that do divide under the

microscopy at each generation is in line with this type of experiments (Lomnitzer and Ron, 1972)(Laskin and Lechevalier, 1984).



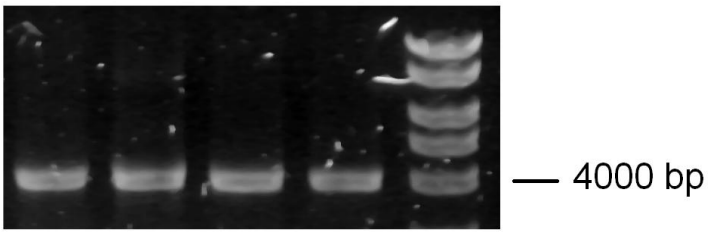
Supplementary Fig 7: Number of cells observed under the microscope over time. Dots are data points, the line shows the trend. Two approximately synchronous divisions can be seen.

9. PCR Analysis of mRFP1-96bs

To characterize the heterogeneity of the plasmids in the cells and to determine whether recombination errors cause the loss of binding sites or some region of the BAC clone (due to the tandem repeats of the binding sites), we have isolated several colonies from the transformant and purified the BAC clone after several cell divisions (overnight cell culturing at 37˚C at 250 rpm). These cells carried the BAC clone with the mRFP1-96bs target gene. To amplify the target gene, the plasmid was isolated and purified using a plasmid purification kit (Fermentas).

The target gene was amplified with Forward primer 5’ GACGTCTGTGTGGAATTGTGAGCGG 3’ and Reverse primer 5’ ACGCGTTCGAAGCTTCGGACGCTA 3’ (Thermo Scientific) from the purified BAC clone. Standard PCR protocol was used to amplify the target which was then run using 1% agarose gel electrophoresis, shown in Supplementary Fig. 8. The target gene is clearly visible at ≈ 4kb in each independent colony, in agreement with the expected length of the original target gene reported in (Golding and Cox, 2004).

From Supplementary Fig 8, it is possible to state that there is no significant variance within each sample since the width of each band is equal to the width of the bands of the ladder. Also, there are no significant differences between bands from different colonies. This indicates that our quantification of RNA numbers per cell is not affected by heterogeneity in the number of MS2-GFP binding sites of the target RNA in different cells, given the observed homogeneity of the results from the PCR analysis of the target.



Supplementary Fig 8: 1% agarose gel electrophoresis of the amplified target gene (mRFP1-96bs) from four different colonies (lanes 1-4), and a 10kb ladder (lane 5).

As a side note, we do not rule out the possibility that recombination events in the 96 binding site region may have occurred and lowered the number of binding sites of the original construct. What can be stated from the results here reported is that if this has occurred, it is a rare event, as it did not introduce diversity in the length of the target RNA of cells of the various colonies used in this study.

References

Bernstein, J, Khodursky A, Lin P, Lin-Chao S, and Cohen S (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA* **99**: 9697–9702.

Dunaway M et al (1980) Kinetic Studies of Inducer Binding to *lac* Repressor-Operator Complex. *Journal of Biological Chemistry* **255**: 10115-10119.

Golding I, and Cox EC (2004) RNA dynamics in live *Escherichia coli* cells. *Proc Natl Acad Sci USA* **101**: 11310–11315.

Golding I, Paulsson J, Zawilski SM, and Cox EC (2005) Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* **123**: 1025-1036.

Gotta SL, Miller OL, and French SL (1991) rRNA transcription rate in *Escherichia coli*. *Journal of Bacteriology* **173**: 6647-6649.

Lanzer M, Bujard H (1988) Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci USA* **85**: 8973-8977.

Laskin AI and Lechevalier HA (1984) Handbook of Microbiology: Growth and Metabolism, Vol. 6. CRC Press.

Lomnitzer R and Ron E (1972) Synchronization of Cell Division in Escherichia coli by Elevated Temperatures: a Reinterpretation. *Journal of Bacteriology* **109**: 1316-1318.

Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Research* **25**: 1203-1210.

Miller DM, Olson JS, Pfluhrath JW, Quiocho FA (1983) Rates of Ligand Binding to Periplasmic Proteins Involved in Bacterial Transport and Chemotaxis. *Journal of Biological Chemistry* **258**: 13665-13672.

McClure WR (1983) A biochemical analysis of the effect of RNA polymerase concentration on the in vivo control of RNA chain initiation frequency. In *Biochemistry of Metabolic Processes*, Lennon DLF, Stratman FW, and Zahlten RN (ed) pp. 207-217. New York: Elsevier Science Publishing Co. Inc.

Ribeiro AS, Zhu R, Kauffman SA (2006) General modeling strategy for gene regulatory networks with stochastic dynamics. *Journal of Computational Biology* **13**: 1630–1639.

Ribeiro AS, Lloyd-Price J (2007) SGN Sim, a Stochastic Genetic Networks Simulator. *Bioinformatics* **23**: 777-779.

Ribeiro AS, Charlebois DA, and Lloyd-Price J (2007) *CellLine*, a stochastic cell lineage simulator. *Bioinformatics* **23**: 3409-3411.

Timmes A, Rodgers M, and Schleif R (2004) Biochemical and Physiological Properties of the DNA Binding Domain of AraC Protein. *Journal of Molecular Biology* **340**: 731-738.

Zhang J, Campbell RE, Ting AY, and Tsien RY (2002) Creating new fluorescent probes for cell biology. *Nature Reviews Molecular Cell Biology* **3**: 906-918.