Supplementary Material for "Kinetics of the cellular intake of a gene expression inducer at high concentrations"

Huy Tran[1], Samuel M.D. Oliveira[1], Nadia Goncalves[1] and Andre S. Ribeiro[1,*]

[1]Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology.
TC316, Korkeakoulunkatu 10, 33720 Tampere, Finland
*andre.ribeiro@tut.fi

I. Measurements and data extraction

Media and growth condition

Cells were grown overnight at 30°C with aeration and shaking in Luria-Bertani (LB) medium, supplemented with the necessary antibiotics. Cells were then diluted in fresh M63 medium. When reaching an optical density of OD600≈0.3–0.5, cells were pre-incubated for 45 min with 100 ng/ml anhydrotetracycline (aTc) to produce enough matured MS2d-GFP proteins to detect RNAs at the start of the microscopy measurements. During the microscopy measurements, cells were kept in M63 media, so as to extend cells' division time, which increases the chances for each cell present at the start of the measurements to produce at least one target RNA before it divides. The contents of (i) LB and (ii) M63 media are:

(i) 10g/L of Tryptone (Sigma Aldrich, USA), 5g/L of yeast extract (LabM, UK) and 10g/L of NaCl (LabM, UK);

(ii) 2mM MgSO4.7H2O (Sigma-Aldrich, USA), 7.6mM (NH4)2SO4 (Sigma Life Science, USA), 30μM FeSO4.7H2O (Sigma Life Science, USA), 1mM EDTA (Sigma Life Science, USA), 60mM KH2PO4 (Sigma Life Science, USA) pH 6.8 with Glycerol 0.5% (Sigma Life Science, USA) and Casaminoacids 0.1% (Fluka Analytical, USA).

Microscopy

After pre-incubation with aTc, cells are placed on a microscope slide with 3% agarose gel to restrict movements. A peristaltic pump is used to provide cells with a constant flow of fresh, pre-warmed M63 media and of IPTG at specified concentrations throughout the measurement period. With the pump initialized at a speed of 0.3 mL/min, the collection of time lapse images by confocal microscopy is initiated as soon as the flow reaches the microscope slide (detected visually).

Microscopy time series were 3 hours long, with cells being imaged once per minute. The data from the first ~5 minutes following induction is not recorded (although time is) as the gel slide slightly shifts due to the initialization of flow of fresh media by the pump, hampering a proper cell tracking.

During the microscopy measurements, the cells' fluorescent background was found to be stable, which indicates that the ability of target RNA counting of the MS2d-GFP system does not change during the course of measurements. Also, from previous studies[1–4], the amounts of fluorescence in the cell background observed suffice to accurately report the appearance of new target RNA molecules in the cells.

Image and data analysis

Image analysis was performed as in [1]. We use a semi-automated cell segmentation strategy [5] as in [6]. Afterwards, fluorescent spots in each cell at each time moment are detected automatically (Figure S1) as in [7], by estimating the cell background intensity distribution using its median and median absolute deviation, and then performing thresholding with a given confidence level assuming that this distribution is Gaussian. Finally, we extracted the

moment when the first RNA appears in each cell and the time intervals between consecutive RNA production events are extracted from the time series of total spot intensities.

We fit a monotonically increasing piecewise-constant function to the corrected total spot intensity in a cell over time using least squares and infer on the moments of appearance of novel target RNAs as in [2,3,8]. The number of terms for the fit was selected by an F-test with a p-value of 0.01. Each discontinuity, i.e. jump, corresponds to the production of one target RNA[3]. An example of the results of applying these methods is shown in Figure S1. Validation of this method is provided in section VII of this document.

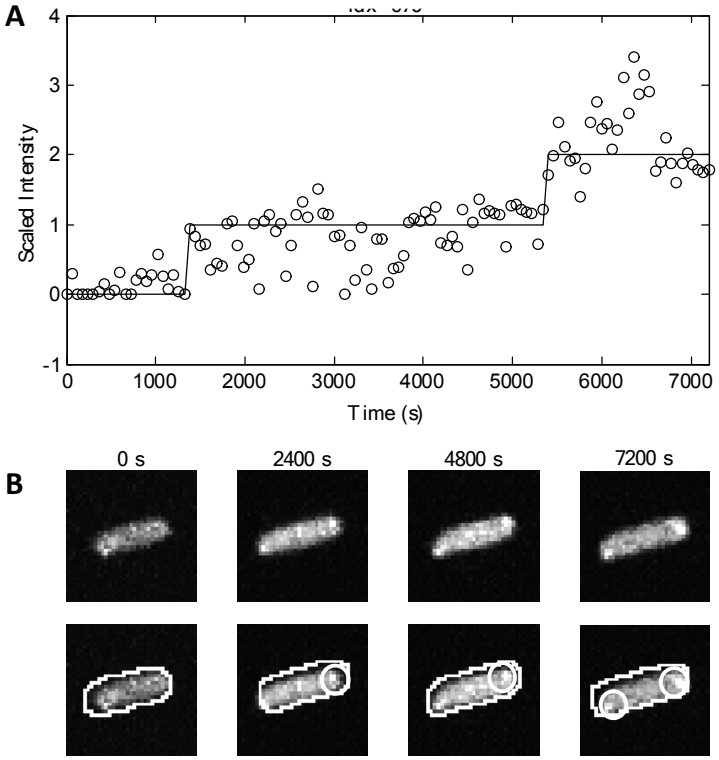


Figure S1. Tagged RNAs in *E. coli* cells. (A) Unprocessed frames and segmented cells and RNA spots. The moments when images were taken are shown for each frame. (B) Examples of time series of scaled spot intensity levels from one cell (circles) and the corresponding estimated RNA numbers (solid lines).

II. Collection and analysis of censored data

The problem of right censored data is well described in [9,10], where each individual in the population has a limited life time drawn from a random variable *Y*. We measure from each individual of the population the time for a certain event *X* to occur. We assume that the time for this event to occur, without the effect of limited life time, is a random variable *T*. Given that *X* has no effect on the 'health' of the individuals under observation, *T* and *Y* are independent of one another.

Collection of censored data

For the i^{th} individual in the population, we draw from the bivariate variable $<T, Y>$ a pair $<t_i, y_i>$, where y_i is the life time of that individual and t_i is the time for event *X* to occur. We define δ_i and z_i as follow:

$$\delta_i = [t_i < y_i] \text{ and } z_i = min(t_i, y_i) \quad (1)$$

where δ_i is the type of sample and z_i is the value of the sample of the i^{th} individual. If the event occurs before the death of the individual, we obtain an actual sample (δ_i=1), else we obtain a censored sample (δ_i=0).

Measurements of the time for the first RNA to appear in each cell, t_0, are obtained from cells present at the start of the microscopy sessions. For measurements of Δt, the intervals between consecutive RNA productions in each cell, the individuals under the observation are any cells that produce one or more RNAs during the last hour of the measurements. For both measurements of Δt and of t_0, the event X to observe is the appearance of the next novel RNA molecule in the cell. Cell 'death' is due to division or the end of the measurement time.

Likelihood function of censored data

To find the likelihood function of the parameter set θ characterizing the model of T, we calculate the possibility to obtain the outcome $<\delta_{1..n}, z_{1..n}>$ from n individuals in the population with this model: $\Lambda\left(\langle\delta_{1..n},z_{1..n}\rangle \,\middle|\, \theta\right)$. With each parameter set θ, the model of T is defined by the probability distribution function $P_{T|\theta}(t|\theta)$ and the cumulative distribution function $F_{T|\theta}(t|\theta)$.

The life time Y of individuals in the population has the probability distribution function $P_Y(y)$ and the cumulative distribution $F_Y(y)$. These distribution functions can be obtained directly by measuring the life time of the individuals in the population.

The likelihood function of the parameter set θ of T's model with the outcome $<\delta_{1..n}, z_{1..n}>$ is given by[9]:

$$\Lambda\left(\langle\delta_{1..n},z_{1..n}\rangle \,\middle|\, \theta\right) = \prod_{i=1}^{n}\left[P_{T|\theta}(z_i|\theta)\left(1-F_Y(z_i)\right)\right]^{\delta_i}\left[P_Y(z_i)\left(1-F_{T|\theta}(z_i|\theta)\right)\right]^{1-\delta_i} \tag{2}$$

Here, $P_{T|\theta}(z_i|\theta)\left(1-F_Y(z_i)\right)$ is the probability of obtaining an actual sample with the value z_i ($<\delta_i$=1, $z_i>$), and $P_Y(z_i)\left(1-F_{T|\theta}(z_i|\theta)\right)$ is the probability of obtaining a censored sample with the value z_i ($<\delta_i$=0, $z_i>$).

While probing for the value of θ that maximizes the likelihood function, the functions $P_Y(y)$ and $F_Y(y)$, which are independent of T, remain constant. Therefore, the objective function to maximize can be simplified to:

$$Obj\left(\theta \,\middle|\, \langle\delta_{1..n},z_{1..n}\rangle\right) = \prod_{i=1}^{n}\left[P_{T|\theta}(z_i|\theta)\right]^{\delta_i}\left[1-F_{T|\theta}(z_i|\theta)\right]^{1-\delta_i} \tag{3}$$

Model distribution of *T* subject to censoring

With the inferred parameter set θ, the probability distribution of T is given as $P_{T|\theta}(t|\theta)$.

The life time of an individual cell in the measurement depends on various factors, such as the division moment and the duration of the measurements. Here, the distribution of the life time Y is obtained directly from the observations of cell life times during the microscopy measurements, rather than being modeled. The inferred distribution of actual samples T' with the distribution of life time Y known is:

$$P_{T'|\theta}(t \mid \theta) = P_{T|\theta}(t \mid \theta) \times P(Y > t) = P_{T|\theta}(t \mid \theta)(1 - F_Y(t)) \tag{4}$$

By comparing $P_{T'|\theta}(t \mid \theta)$ with the empirical distribution of the actual samples (δ_i=1, z_i) using Pearson's chi-squared test, we can calculate the goodness of fit of θ's estimation.

III. Solving the deterministic model of inducer dynamics

Model of inducer dynamics

The model of inducer dynamics is described (as in equations (7) and (8) in the manuscript) as follows:

$$\delta I_m / \delta t = k_o - I_m \times k_i \tag{5}$$

$$\delta I / \delta t = I_m \times k_i - I \times d_I \tag{6}$$

We first find the solution for the inducer level in the periplasmic space (I_p):

$$\frac{\delta I_m}{k_o - I_m k_i} = \delta t \tag{7}$$

By integrating both sides of the equation, we obtain:

$$-\frac{\ln\left(k_o - I_m k_i\right)}{k_i} = t + C_1 \tag{8}$$

$$\leftrightarrow k_o - I_m k_i = C_1 . e^{-k_i t} \tag{9}$$

$$\leftrightarrow I_m = \frac{k_o - C_1 e^{-k_i t}}{k_i} \tag{10}$$

At $t=0$, $I_m(0)=0$, thus $C_1=k_o$. The solution for I_m is:

$$I_m(t) = \frac{k_o(1 - e^{-k_i t})}{k_i} \tag{11}$$

The differential equation for $I(t)$ becomes a first order linear differential equation:

$$\frac{\delta I}{\delta t} + I.d_I = k_o\left(1 - e^{-k_i t}\right) \tag{12}$$

The general solution for this equation is:

$$I(t) = \frac{\int u(t) k_o\left(1 - e^{-k_i t}\right) dt + C_2}{u(t)} \tag{13}$$

in which $u(t) = e^{\int d_I dt} = e^{d_I t}$. C_2 is a constant determining the initial condition $I(0)$. Thus:

$$I(t) = \frac{\int e^{d_I t} k_o\left(1 - e^{-k_i t}\right) dt + C_2}{e^{d_I t}}$$

$$= \frac{k_o(\int e^{d_I t} dt - \int e^{\left(d_I - k_i\right)t} dt) + C_2}{e^{d_I t}} = \frac{k_o}{d_I} - \frac{k_o e^{-k_i t}}{d_I - k_i} + \frac{C_2}{e^{d_I t}} \tag{14}$$

$$= \frac{k_o(d_I - k_i) - d_I k_o e^{-k_i t} + C_2 d_I (d_I - k_i) e^{-d_I t}}{d_I (d_I - k_i)}$$
$$= \frac{k_o\left(d_I e^{-k_i t} + C_2 d_I (k_i - d_I) e^{-d_I t} + k_i - d_I\right)}{d_I (k_i - d_I)}$$

At $t=0$, $I(0)=0$, $C_2 d_I (k_i - d_I) = -k_i$.

The final solution for the intracellular inducer quantity over time is therefore:

$$I(t) = \frac{k_o\left(d_I e^{-k_i t} - k_i e^{-d_I t} + k_i - d_I\right)}{d_I (k_i - d_I)} \quad (15)$$

IV. Model distribution of t_0

From the models of inducer intake and of transcription, we use the Chemical Master Equation (CME) [11] to calculate the first moment of open complex formation in each cell, which is followed, shortly after, by the release of a transcript [12,13]. For this, we assume that, upon this release, the promoter is unable to transcribe any subsequent RNA. Given this approximation, the master equation for the promoter in each of its three possible states is given by:

$$\delta P(\mathrm{Pr},t)/\delta t = -k_c f_R(t) \times P(\mathrm{Pr},t) \quad (16)$$

$$\delta P(\mathrm{Pr}_c,t)/\delta t = k_c f_R(t) \times P(\mathrm{Pr},t) - k_o \times P(\mathrm{Pr}_c,t) \quad (17)$$

$$\delta P(\mathrm{Pr}_o,t)/\delta t = k_o \times P(\mathrm{Pr}_c,t) \quad (18)$$

$P(Pr,t)$, $P(Pr_c,t)$ and $P(Pr_o,t)$ are the probabilities that the promoter is in its primary state, in closed complex state and in open complex state, respectively, at time t. Due to the high amount of repressors in the cells [14], we ignore the leakiness of the target gene (from our measurements, we observed that, on average, it takes more than 1 hour for ~10% of the cells to produce one spurious RNA, when not induced). Given this, we set the probability of the promoter to be in its primary state, P(Pr,0), to 1 and to be in the other two states ($P(Pr_c,0)$ and $P(Pr_o,0)$) to 0.

V. Dilution rate of regulatory molecules at various induction levels

The dilution rate of regulatory molecules (d_I) is calculated from the expansion rate of the cells' volume. As *E. coli* grows mostly by elongating through its major axis length, while leaving its minor axis length unchanged, the relative increase in cell's volume can be approximated by the increase in the cell's major axis length.

Cell growth in liquid media

To test for the effect of IPTG induction on the cells growth rate at 37 ˚C, we first measured cell growth in liquid media. Cells were grown overnight at 30 ˚C with aeration and shaking in LB media, supplemented with the appropriate antibiotics, before being diluted in fresh LB medium until an OD600 ≈ 0.1 and pre-incubated for 2 hours without inducers. In the remaining hours, cells were either left to grow normally or grown in the presence of IPTG at the concentration of 0.25mM and 1mM. The optical density (OD) curves at 0mM, 0.25mM and 1mM IPTG concentrations were sampled every 30 minutes for 5 hours (Figure S2).

From Figure S2, during the first 4 hours of the measurements there is little difference between the normalized OD curves, indicating that, in the range of concentrations tested, IPTG does not have any notable effect on cell growth.

Cell growth on agarose gel

Next, we obtained the cell growth rate during the microscopy measurements, where cells are kept on agarose gel as described in the Methods section of the main manuscript. As only a few cell cycles were observed in M63 media during 2 hour-long measurements, we estimated the cell growth rate from the elongation rate of all cells' major axis rather than the cells' doubling time.

From the time lapse confocal images, cells were segmented and the length of the major axis was extracted at each frame. For each cell, we fitted a linear function to the logarithm of the major axis length over time and obtained the slope coefficient d_I', equivalent to the cell's elongation rate. The doubling time T_d' of each cell is inferred from d_I' as follows:

$$T_d' = \frac{ln(2)}{d_I'} \tag{19}$$

The distributions of T_d' at different induction levels spans over a wide range of durations, suggesting a noisy dilution rate when cells are on the 3% agarose gel. The distributions share a mode of around ~8400 seconds. To eliminate any effects of noise in the dilution rate of regulatory molecules, for the analysis of t_0, we selected 'normal' cells with a doubling time T_d'~8400 s, using a margin for selection of 15% of the mode's value. Finally, from the value of T_d', the dilution rate d_I of the selected cells is found to be:

$$d_I = \frac{ln(2)}{T_d'} = 8.25 \times 10^{-5}\ (s^{-1}) \tag{20}$$

Since cells grew exponentially during the measurements at a rate of d_I ~8.25×10^{-5} s^{-1} (doubling time of ~140 minutes) in all conditions, it is reasonable to assume that the cells were unaffected by the inducer in the range of concentrations tested (in this regard see, e.g. [15]).

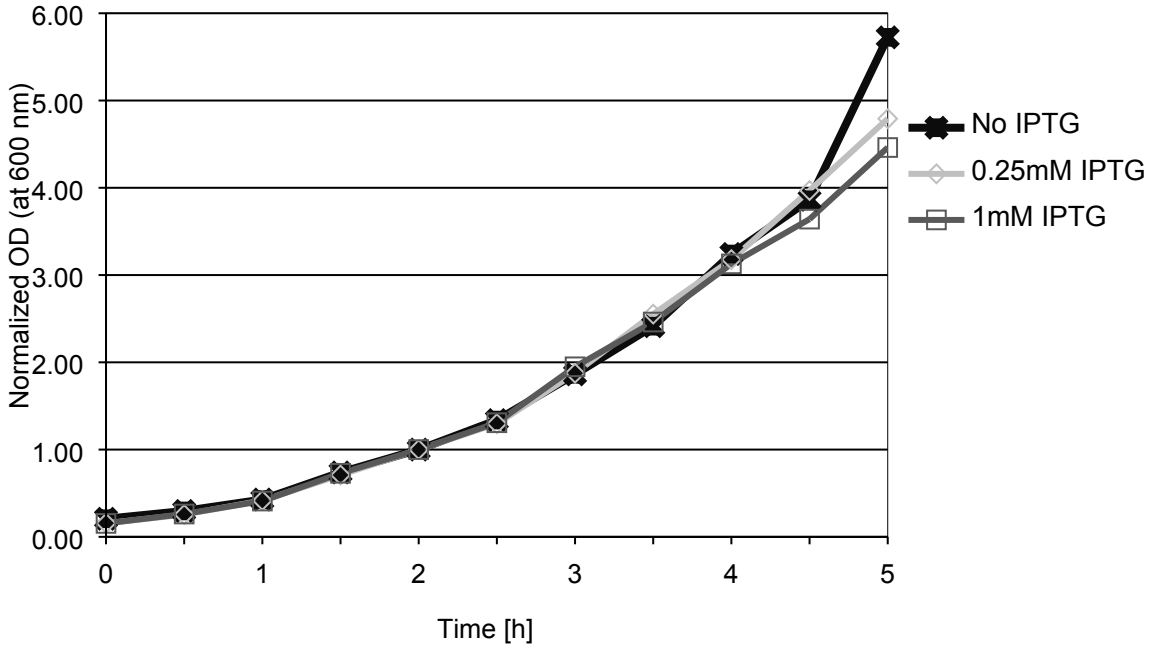
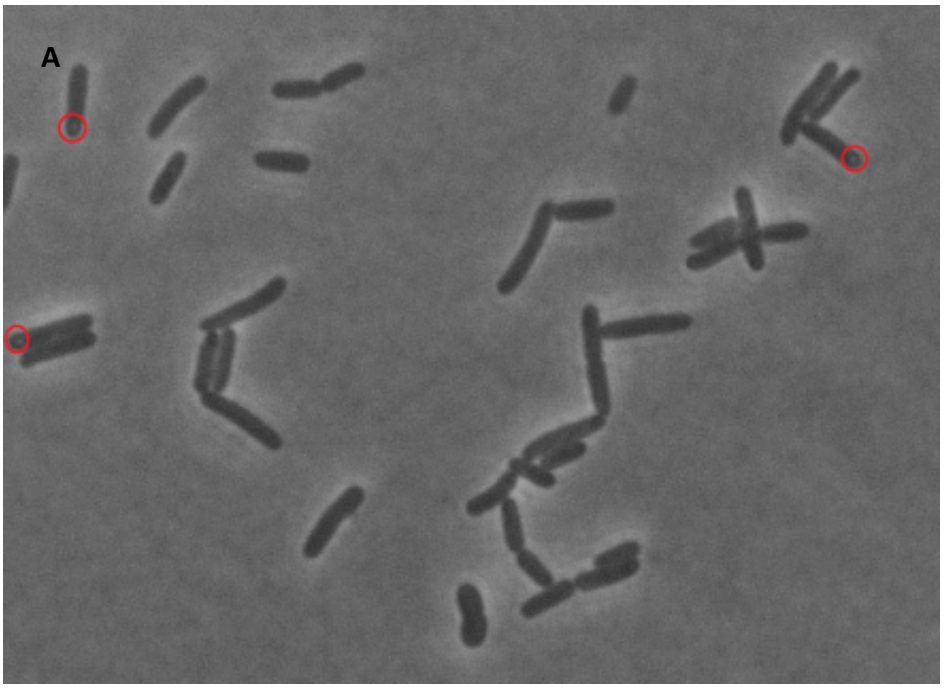
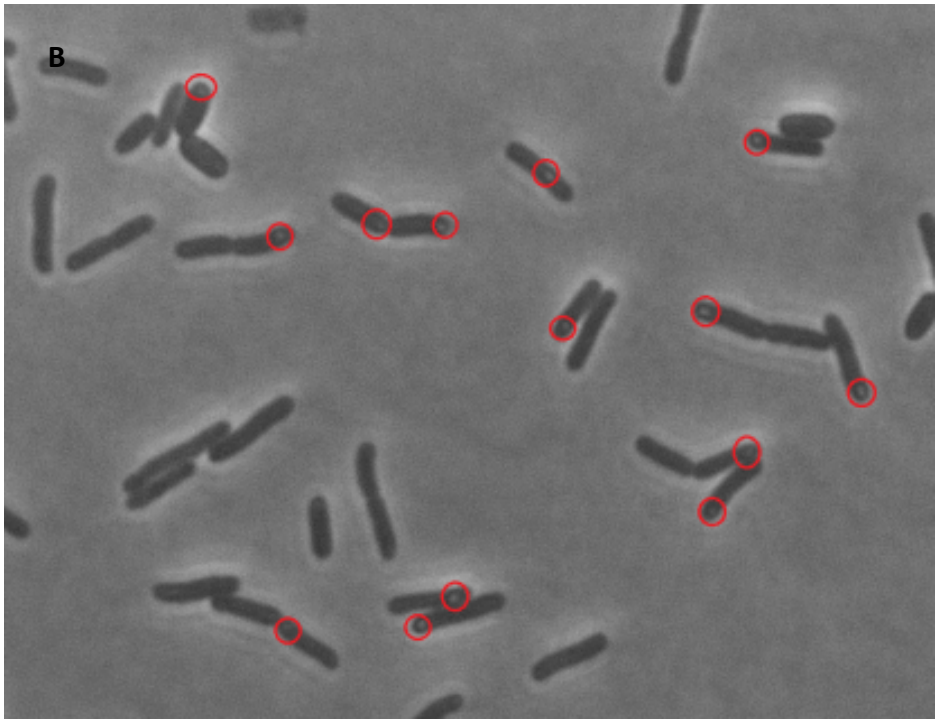


Figure S2. Normalized optical density (OD) curve at 0.25mM (diamond) and 1mM (square) IPTG and without IPTG (cross). Inducers are added at the end of the second hour, where the normalized OD's values equal 1.

VI. Formation of inclusion bodies at high inducer concentrations

We use phase contrast microscopy to examine the fraction of cells with inclusion bodies as a function of IPTG concentration in the media. Cells were grown overnight at 30 ˚C with aeration and shaking in LB media, supplemented with the appropriate antibiotics, before being diluted in fresh LB medium until an OD600 ≈ 0.1 and pre-incubated for 2 hours without inducers. In the remaining hours, cells were incubated in the presence of aTc at 100ng/L and IPTG at 1mM, 2mM and 4mM before being placed under the microscope. From the phase contrast images, we manually detected the presence of inclusion bodies (shown as a bright spot) in each cell. Example images of cells with marked inclusion bodies are shown in Figure S3.


A


B

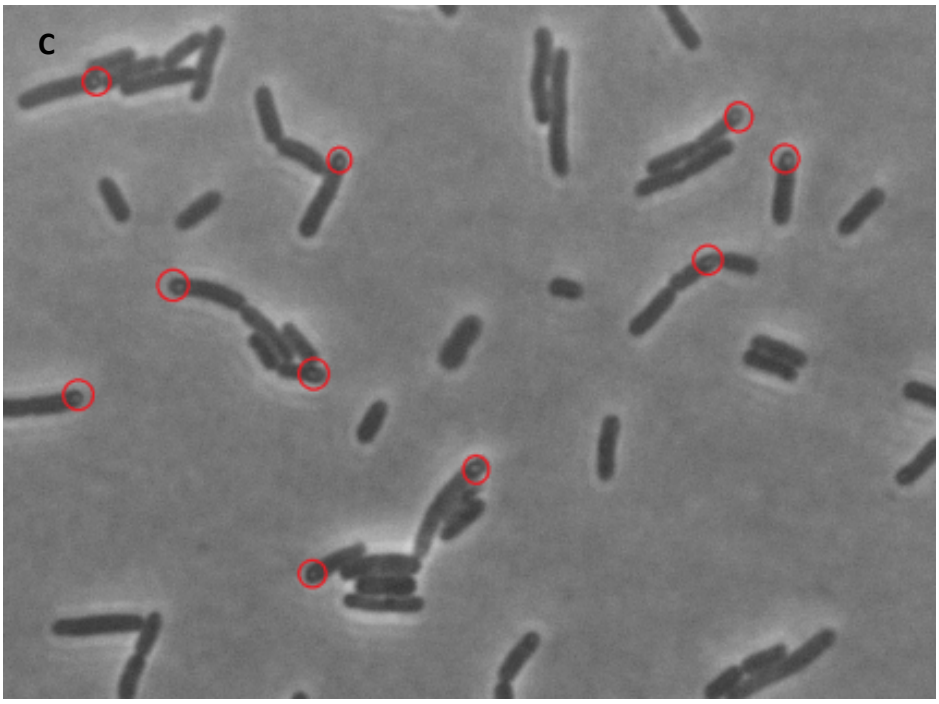


Figure S3. Phase contrast images with cells with marked inclusion bodies (appears as a bright spots), when induced with IPTG concentrations of (Top) 1mM, (Middle) 2mM and (Bottom) 4mM.

VII Temporal fluorescence intensity of MS2d-GFP tagged RNA molecules

The technique of detecting new RNA molecules in individual cells as these appear in time lapse microscopy images using the MS2d-GFP RNA-tagging system (ref. 6 in main manuscript) consists of fitting the total corrected RNA spot intensity with a step-increasing function (see section I of this document).

For this method to be valid, it is necessary that new RNA molecules appear nearly fully-tagged when first detected, so as to cause a significant "jump" in the total spots fluorescence intensity of the cell[7]. This is possible if the speed of elongation at the target gene and MS2d-GFP binding is not much longer than the interval between consecutive images, which in our measurements is 1 minute long.

Also, it is necessary that an MS2d-GFP tagged RNA, once tagged, does not degrade significantly (neither abruptly nor gradually) during the measurement period (so as to allow using a step increasing function). Note that, nevertheless, the method can tolerate infrequent "blinking" of the tagged RNAs, due to moving out of focus transiently, without loss of information[7].

To validate the two assumptions, we observed the fluorescence intensity of individual, RNA spots over time (1 min^{-1}). As newly produced RNA spots could appear and compensate for the loss of intensity (abrupt or gradual) of the existing spots (resulting in the underestimation of the spots' degradation rate), we conducted the observation on a non-induced target gene. Namely, following the protocol described in the main manuscript (except for the induction of expression of the RNA target for MS2d-GFP), we observed sufficient cells during a period of 3 hours so that at least 40 RNA spots appearances could be detected (during that period of time, less than 1 in 10 cells produced an RNA spot). Note that, by

inspection, we never observed the appearance of two new fluorescent spots in a cell at the same time moment and no cell ever contained 2 spots.

To test the first assumption, from the time-lapse images, we obtained the fluorescence intensity of 40 individual tagged RNAs for 30 minutes, since first detected. From these, we found that there is no significant RNA fluorescence increase after its detection. That is, new RNA molecules are nearly fully-tagged when first detected, as expected from the frequency of image acquisition (1 min^{-1}) and the expected speed of transcription elongation and MS2d-GFP binding (tens of seconds[16,17]). This is visible in Figure S4, where the mean spot fluorescence over time is shown. Note how, following the detection of the spots at moment 0 (synchronized for easier visualization), their mean fluorescence over time does not increase further in subsequent time moments.

To test the second assumption, we fitted the intensity of each RNA spot over time with a decaying exponential function and inferred the degradation rate of the spot intensity. We obtained a mean decaying rate of ~8.1×10^{-5} s^{-1}, corresponding to a mean half-life of ~144 mins, which is much longer than our observation window for Δt (60 mins). As such, we conclude that, during the measurement period, the fluorescence of tagged RNAs does not decrease significantly over time (gradually or abruptly), in agreement with previous reports using the same RNA detection system[2,3,16,17].

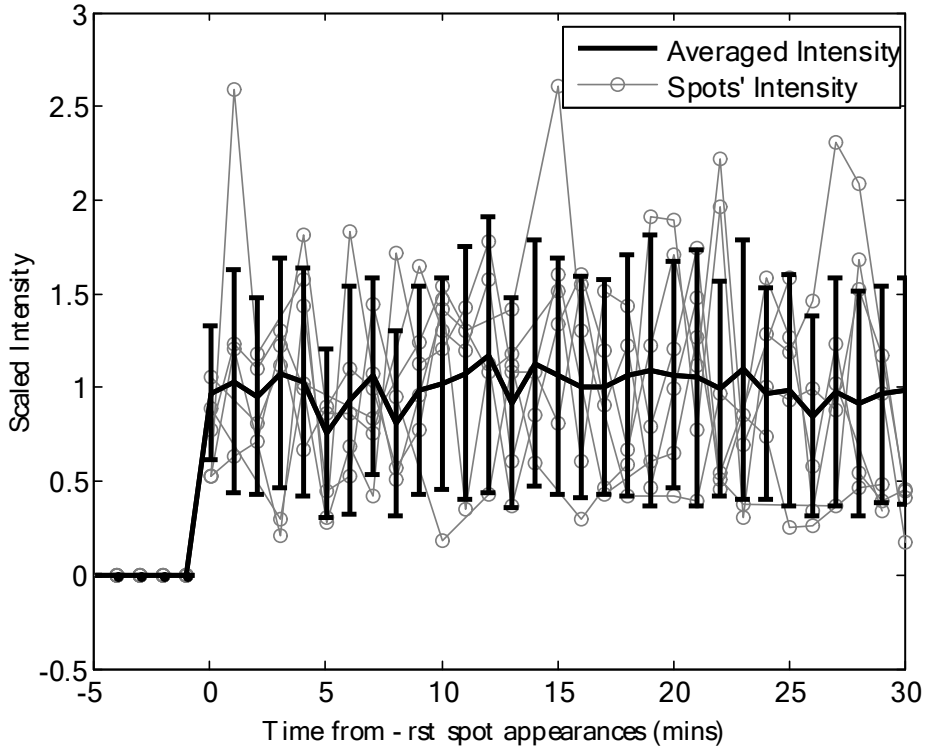


Figure S4. Fluorescence of tagged RNAs in *E. coli* cells over time. Each of the five thin lines shows the fluorescence of a single tagged RNA molecule (randomly selected from the data) since first detected, for a period of 30 minutes. The solid black line shows the mean fluorescence intensity of individual tagged RNA molecules (40 molecules tracked), along with the sample standard deviation (vertical bars).

The above results are in agreement with previous studies. Regarding the dynamics of RNA production, the present results agree with previous data on the rate of transcription elongation in *E. coli*. Namely, at 37ºC, this rate is expected to be between ~60 and ~90 base pairs (bp) per second[18–20]. Given that the target gene is ~3200 bp long[16], the RNA polymerase should produce a complete transcript in ~35 to ~50 s, which is faster than our imaging interval (60 s).

Meanwhile, regarding the lack of degradation of tagged RNAs, our results are expected given previous studies on the coat protein of bacteriophage MS2[16,21,22], which showed that most of the MS2 binding sites are constantly occupied by (at least 70) MS2d-GFP proteins, which results in the 'immortalization' of the target RNA due to isolation from RNA-degrading enzymes [16,17].

REFERENCES

1 J. Mäkelä, M. Kandhavelu, S. M. D. Oliveira, J. G. Chandraseelan, J. Lloyd-Price, J. Peltonen, O. Yli-Harja and A. S. Ribeiro, *Nucleic Acids Res.*, 2013, **41**, 6544–52.

2 A. B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja and A. S. Ribeiro, *Nucleic Acids Res.*, 2012, **40**, 8472–83.

3 M. Kandhavelu, J. Lloyd-Price, A. Gupta, A. B. Muthukrishnan, O. Yli-Harja and A. S. Ribeiro, *FEBS Lett.*, 2012, **586**, 3870–5.

4 M. Kandhavelu, H. Mannerström, A. Gupta, A. Häkkinen, J. Lloyd-Price, O. Yli-Harja and A. S. Ribeiro, *BMC Syst. Biol.*, 2011, **5**, 149.

5 S. Chowdhury, M. Kandhavelu, O. Yli-Harja and A. S. Ribeiro, *J. Microsc.*, 2012, **245**, 265–75.

6 A. Häkkinen, A.-B. Muthukrishnan, A. Mora, J. M. Fonseca and A. S. Ribeiro, *Bioinformatics*, 2013, **29**, 1708–9.

7 M. Kandhavelu, A. Häkkinen, O. Yli-Harja and A. S. Ribeiro, *Phys. Biol.*, 2012, **9**, 026004.

8 H. Mannerstrom, O. Yli-Harja and A. S. Ribeiro, *Eurasip J. Bioinforma. Syst. Biol.*, 2011, **2011**, 11–15.

9 D. R. Cox, *J. R. Stat. Soc. Ser. B*, 1972, **34**, 187–220.

10 H. Koul, V. Susarla and J. Van Ryzin, *Ann. Stat.*, 1981, **9**, 1276–1288.

11 D. T. Gillespie, *J. Stat. Phys.*, 1977, **16**, 311–318.

12 L. M. Hsu, *Biochim. Biophys. Acta*, 2002, **1577**, 191–207.

13 K. M. Herbert, A. La Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick and S. M. Block, *Cell*, 2006, **125**, 1083–94.

14 R. Lutz and H. Bujard, *Nucleic Acids Res.*, 1997, **25**, 1203–10.

15 S. Cooper, *Theor. Biol. Med. Model.*, 2006, **3**, 10.

16 I. Golding and E. C. Cox, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11310–5.

17 I. Golding, J. Paulsson, S. M. Zawilski and E. C. Cox, *Cell*, 2005, **123**, 1025–36.

18 U. Vogel and K. a J. F. Jensen, 1994.

19 P. P. Dennis, M. Ehrenberg, D. Fange and H. Bremer, *J. Bacteriol.*, 2009, **191**, 3740–3746.

20 J. Ryals, R. Little and H. Bremer, 1982, **151**, 879–887.

21 S. J. Talbot, S. Goodman, S. R. E. Bates, C. W. G. Fishwick and P. G. Stockley, *Nucleic Acids Res.*, 1990, **18**, 3521–3528.

22 D. Fusco, N. Accornero, B. Lavoie, S. M. Shenoy, J. M. Blanchard, R. H. Singer and E. Bertrand, *Curr. Biol.*, 2003, **13**, 161–167.