Measurement of gene regulation in individual cells reveals rapid switching between promoter states

Leonardo A. Sepúlveda,1,2 Heng Xu,1,2 Jing Zhang,1,2 Mengyu Wang,1,2,3 Ido Golding2,3,4

In vivo mapping of transcription-factor binding to the transcriptional output of the regulated gene is hindered by probabilistic promoter occupancy, the presence of multiple gene copies, and cell-to-cell variability. We demonstrate how to overcome these obstacles in the lysogenic maintenance promoter of bacteriophage lambda, P_{RM}. We simultaneously measured the concentration of the lambda repressor CI and the number of messenger RNAs (mRNAs) from P_{RM} in individual Escherichia coli cells, and used a theoretical model to identify the stochastic activity corresponding to different CI binding configurations. We found that switching between promoter configurations is faster than mRNA lifetime and that individual gene copies within the same cell act independently. The simultaneous quantification of transcription factor and promoter activity, followed by stochastic theoretical analysis, provides a tool that can be applied to other genetic circuits.

Sequence-specific transcription factors drive the diversity of cell phenotypes in development and homeostasis (1). For each target gene, alternative transcription-factor binding configurations (by different transcription factors or by multiple copies of the same one) result in varied transcriptional outputs, in turn leading to alternative cell fates and behaviors (2, 3). Elucidating the relations between transcription-factor configurations [which can number in the hundreds (4–6)] and the resulting transcriptional activity remains a challenge. Application of traditional genetic and biochemical approaches usually requires a genetically modified system or assays of purified components in vitro (7). Ideally, however, one would like to map transcription-factor configuration to promoter activity inside the cell, with minimal perturbation to the endogenous system.

Multiple factors hinder such direct measurement. First, individual cells vary in both transcription-factor concentration and the resulting transcriptional activity (8, 9); averaging over many cells thus filters out details of the regulatory relation. Second, even within the single cell, more than one copy of the regulated gene is typically present, with each copy individually regulated (10). Finally, even at the level of a single gene copy, multiple binding configurations are possible at a given transcription-factor concentration (11, 12). The relative probabilities of these different configurations and the rate of switching between them will define the stochastic activity of the regulated promoter (13).

We simultaneously measured, in individual cells, the concentration of a transcription factor and the number of mRNAs produced from the regulated gene. We also measured how the gene copy number changes through the cell cycle. We then analyzed the full single-cell data using a theoretical model, which allowed us to identify the contributions of different transcription-factor binding configurations to the stochastic activity of the promoter.

Specifically, we examined the lysogenic maintenance promoter of phage lambda, P_{RM}. The regulation of this promoter by its own gene product, the lambda repressor (CI), is a paradigm for how alternative binding configurations drive transcriptional activity and the resulting cell fate—stable lysogenic or lytic induction resulting in cell death (7). The number of possible CI configurations is very large (>100 (4, 5)). Briefly, as CI concentration increases, CI dimers gradually occupy three proximal (O_{l1-3}) and three distal (O_16-18) operator sites, leading first to activation, then repression, of P_{RM} (Fig. 1A). Cooperative CI binding, and looping of DNA between the O_{l3} and O_{l4} sites, play important roles in shaping the P_{RM}(CI) regulatory curve (14).

In a lysogen (a bacterium carrying a prophage), CI concentration is believed to be such that P_{RM} fluctuates between the activated and repressed states (15) (Fig. 1A), and this has been suggested to stabilize the lysogenic state against random fluctuations in CI levels (16). However, the nature of the lysogenic “mixed state” (activated/repressed) is unknown: Are the promoter fluctuations slow...
respectively? Alternatively, are promoter fluctuations fast, such that all cells exhibit an intermediate, well-defined level of \( P_{\text{RM}} \) expression (Fig. 1B)?

**Fig. 1. Schematic of \( P_{\text{RM}} \) regulation by CI.** (A) As the concentration of CI increases, the probabilities of different binding configurations of CI dimers at the \( O_L \) and \( O_H \) operators change (color shading), resulting in varying \( P_{\text{RM}} \) activity (gray curve). Three configurations, expected to be the most probable, are depicted. In lysogenic cells, \( P_{\text{RM}} \) has comparable probabilities of being in the activated and repressed promoter states (gray shading). (B) The rate of switching between activated and repressed states drives the stochastic activity of \( P_{\text{RM}} \) in lysogenic cells. Two alternative hypotheses are illustrated: If switching is slow relative to the mRNA lifetime (left), two subpopulations of cells will exist, with low and high mRNA levels. If switching is fast (right), the mRNA distribution in the population will be unimodal.

**Fig. 2. Measuring the number of CI molecules in individual cells.** (A) CI proteins were labeled with antibodies to CI and fluorescently labeled secondary antibodies (left). Under the microscope, lysogenic *E. coli* cells exhibited a strong CI signal (right) whereas nonlysogens showed a weak background signal (center). (B) Method 1 for measuring the number of CI proteins per cell. The typical fluorescence of a single CI dimer was obtained from the spot intensity distribution in lysogenic cells (green, \( N = 23,631 \) spots), distinguishable from that of the negative sample (black, \( N = 1764 \) spots). (C) Method 2 for measuring the number of CI proteins per cell. The variance versus the mean of pixel intensity in individual cells (gray, \( N = 324 \) spots) was fitted to a linear function (green). The slope of this line was used to estimate the fluorescence intensity of a single CI dimer. (D) The estimated number of CI molecules in a lysogen, obtained with the two single-cell methods (green, mean ± SEM from six experiments, 327 to 704 cells each). Also shown is the value reported in the literature (gray, mean ± SD from three studies (19–21)). (E) The distribution of CI copy number in lysogenic cells (green; \( N = 560 \) cells). The data are described well by a gamma distribution (black).

To measure CI concentration in individual cells, we used antibody labeling (immunofluorescence). Lysogenic cells (see table S1) exhibited a strong CI signal, whereas nonlysogenic (uninfected) cells showed only a weak background signal (Fig. 2A and fig. S1). To verify that the antibody signal reliably represents CI levels, we expressed a CI–yellow fluorescent protein (YFP) fusion protein (16) in nonlysogenic cells and compared the YFP fluorescence to the signal exhibited by the antibody to CI in each cell. The two signals were linear with each other (fig. S2A), and single-molecule imaging revealed that most YFP molecules were colocalized with an antibody to CI, as expected (fig. S2B).

To convert the antibody signal to CI concentration in each cell, we needed to know the fluorescence value corresponding to a single antibody–labeled CI molecule [a CI dimer, which is the dominant species in the cell (27)]. To obtain this calibration constant, we used two methods (18) (Fig. 2, B and C): In the first method, we used automated image analysis to identify individual fluorescent particles (spots, fig. S3). These spots displayed a well-defined intensity value, distinct from the corresponding signal found in negative samples (Fig. 2B). We identified the positive-sample spot intensity as corresponding to individual CI dimers (fig. S4A) (each one decorated by ~20 fluorescent dyes, due to the stoichiometry of antibody labeling; fig. S5) and used it to convert cell fluorescence to CI concentration. In the second method, we used the fact that the Poisson statistics of random protein positions within the cell lead to a linear relation between the
fluorescence in situ hybridization (smFISH) (individual lysogenic cells, we used single-molecule to label and count proteins each, during the 30-min cell cycle (table 1220 values also agreed with those reported in the results (Fig. 2D and fig. S4B). These measured fluorescence mean and the pixel-to-pixel variance of a single labeled fluorescence corresponding to a single labeled copy of cI with a (more accurate) value that we obtained when accounting for the presence of two to four gene copies per cell (Fig. 3B), this value is consistent with the burst frequency estimated from the CI protein histogram (Fig. 2E).

To measure the regulatory relation between CI concentration and P Ram activity, we used a reporter system in which the autoregulatory feedback from CI to P Ram existing in the lysogen is broken: CI is expressed from an inducible promoter, whereas P Ram transcribes the lacZ gene rather than cI (Fig. 4A). To simultaneously measure CI concentration and P Ram activity in the same cell, we combined immunofluorescence (using antibody to CI) with smFISH (using lacZ probes) (18) (Fig. 4B and fig. S10) and measured the corresponding protein and mRNA numbers as described above. Performing this measurement over a range of CI levels, then plotting lacZ mRNA numbers versus CI concentration from many individual cells, produced highly scattered data (Fig. 4C), as expected from the stochasticity of the regulation and transcription processes (9). Averaging within finite windows of CI concentration revealed the mean regulatory relation between CI and P Ram known as the gene regulation function (16) (Fig. 4C and fig. S11). The shape of the regulation function agreed with that from previous reports, with P Ram activity first increasing, then decreasing, with CI concentration (4, 14, 28). However, our measurement provides the absolute numbers for both the input (CI concentration) and output (mRNA numbers), rather than relative expression levels (4, 5, 14, 28). The absolute values are crucial for the subsequent steps in our analysis of P Ram regulation.

As the first step in this analysis, we wrote down a theoretical model in which the probabilities of different CI binding configurations are given by their thermodynamic weights (15) (fig. S12A). This thermodynamic model successfully reproduced the regulation function (Fig. 4C and fig. S13). In performing this procedure, most free-energy values used in the model were identical to those reported (15) (table S5). The model also provided the probabilities of observing different promoter activity states—basal, activated [with the DNA between O1 and O2 either looped or unlooped (15)], and repressed—as a function of CI concentration (Fig. 4D). The overlap between the different states underlines the challenge in identifying the transcriptional signature of a single promoter state: For example, the probability of P Ram being in the activated state does not surpass ~50%.

To reveal the activity of individual promoter states, we introduced a stochastic version of the theoretical model (Fig. 4E and fig. S12). In the model, the CI binding configurations are grouped based on the expected promoter activity: basal, activated unlooped, activated looped, and repressed.

Analyzing the single-gene mRNA distribution (Fig. 3D) revealed that a single copy of P Ram produces a burst of cI mRNA every ~6 min on average (table S4). This burst frequency estimated from the CI protein histogram (Fig. 2E). Comparing the protein and mRNA data also allowed us to directly calculate the number of CI proteins produced from each cI mRNA, ~6 on average (table S4). This value is in good agreement with a previous theoretical calculation (23).

To measure the regulatory relation between CI concentration and P Ram activity, we used a reporter system in which the autoregulatory feedback from CI to P Ram existing in the lysogen is broken: CI is expressed from an inducible promoter, whereas P Ram transcribes the lacZ gene rather than cI (Fig. 4A). To simultaneously measure CI concentration and P Ram activity in the same cell, we combined immunofluorescence (using antibody to CI) with smFISH (using lacZ probes) (18) (Fig. 4B and fig. S10) and measured the corresponding protein and mRNA numbers as described above. Performing this measurement over a range of CI levels, then plotting lacZ mRNA numbers versus CI concentration from many individual cells, produced highly scattered data (Fig. 4C), as expected from the stochasticity of the regulation and transcription processes (9). Averaging within finite windows of CI concentration revealed the mean regulatory relation between CI and P Ram known as the gene regulation function (16) (Fig. 4C and fig. S11). The shape of the regulation function agreed with that from previous reports, with P Ram activity first increasing, then decreasing, with CI concentration (4, 14, 28). However, our measurement provides the absolute numbers for both the input (CI concentration) and output (mRNA numbers), rather than relative expression levels (4, 5, 14, 28). The absolute values are crucial for the subsequent steps in our analysis of P Ram regulation.

As the first step in this analysis, we wrote down a theoretical model in which the probabilities of different CI binding configurations are given by their thermodynamic weights (15) (fig. S12A). This thermodynamic model successfully reproduced the regulation function (Fig. 4C and fig. S13). In performing this procedure, most free-energy values used in the model were identical to those reported (15) (table S5). The model also provided the probabilities of observing different CI binding configurations—basal, activated [with the DNA between O1 and O2 either looped or unlooped (15)], and repressed—as a function of CI concentration (Fig. 4D). The overlap between the different states underlines the challenge in identifying the transcriptional signature of a single promoter state: For example, the probability of P Ram being in the activated state does not surpass ~50%.

To reveal the activity of individual promoter states, we introduced a stochastic version of the theoretical model (Fig. 4E and fig. S12). In the model, the CI binding configurations are grouped based on the expected promoter activity: basal, activated unlooped, activated looped, and repressed.
(15). Each promoter activity state is described by stochastic bursty kinetics of mRNA production (29). P\textsubscript{Rm} stochastically switches between its four activity states. The switching rates are initially unknown, but the thermodynamic model above provides us with the equilibrium constant (ratio between switching left and right) for each pair of states, at a given CI concentration. For each set of parameters, the stochastic model can be solved to yield the expected mRNA copy-number distribution for the population of multicopy cells.

We used the stochastic model to analyze the full P\textsubscript{Rm}(CI) single-cell data set (Fig. 4C). Applying maximum-likelihood estimation, we found good agreement between the experimental and theoretical mRNA distributions over the full range of CI concentrations (Fig. 4F, fig. S14, and movie S1). The fitting procedure allowed us to extract the mRNA statistics corresponding to the different activity states of P\textsubscript{Rm} (fig. S15). The calculated distributions were in good agreement with those obtained with genetic controls: cells expressing no CI (basal), and cells overexpressing CI in wild-type P\textsubscript{Rm} (repressed) and in a mutant lacking the O\textsubscript{L} operator (activated unlooped) (14) (fig. S15B and table S6). The stochastic kinetics of each promoter state exhibited a similar relation between expression level and burst size to that measured in other E. coli promoters (29) (fig. S15C).

Even though the measured mRNA distribution at each CI concentration represents a mixture of multiple promoter states, each of the histograms is unimodal and can be described by a simple kinetic model with a single burst size and frequency (Fig. 4F and fig. S16). The parameter that determines the shape of the “mixed state” mRNA distribution is the rate of switching between promoter states (Fig. 1B). Previous in vitro studies of O\textsubscript{L}-O\textsubscript{R} looping suggested that the switching between looped and unlooped promoter configurations is fast (~seconds) (30), but similar studies of looping in the cell left the question open (31). Our stochastic model predicts that if promoter switching is very slow relative to mRNA lifetime [here ~2 min (29)], the observed mRNA distribution will be the weighted sum of the underlying single-promoter-state distributions. Our experimental data strongly disagreed with this prediction (Fig. 4G). By contrast, if switching is fast, the observed distribution will be given by a (weighted) convolution of the underlying single-promoter-state distributions, and if the underlying states can each be described by simple bursty kinetics, the new mixed state can be as well. This is indeed what we observed (Fig. 4, F and G, and fig. S16). Thus, P\textsubscript{Rm} switches rapidly between different promoter states, resulting in a stochastic signature that (at a given CI concentration) is indistinguishable from that of a single promoter state, but with renormalized kinetic parameters. Our finding of rapid switching explains why, in the lysogen, we did not detect distinct “active” and “repressed” populations in either the protein populations in either the protein (Fig. 2E) or mRNA (Fig. 3A) histograms, but instead both data sets indicated a single, well-defined promoter activity.

Precise single-cell measurements, accompanied by theoretical analysis, can reveal new features even in well-studied model systems. When combined with genetic and synthetic-biology approaches (23), this strategy may allow prediction of the stochastic characteristics of promoter activity,
a prediction that remains a challenge to our understanding of gene regulation (9, 32).

REFERENCES AND NOTES


ACKNOWLEDGMENTS

We are grateful to the following people for generous advice and for providing reagents: I. Dodd, M. Elowitz, L. Finzi, H. Garcia, T. Gregor, T. Kuhlman, L. McLane, R. Phillips, A. Raj, E. Rothenberg, A. Sanchez, K. Shearwin, R. Singer, S. Skinner, L-H. So, A. Sokac, L. Zeng, and C. Zong. Work in the Golding lab is supported by grants from NIH (R01 GM062837), NSF (PHY 1347498, PHY 1430124 and PHY 1427654), The Welch Foundation (Q-1759), and The John S. Dunn Foundation (Collaborative Research Award). H.X. is supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface. We gratefully acknowledge the computing resources provided by the Computational and Integrative Biomedical Research Center of Baylor College of Medicine.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/351/6278/1218/suppl/DC1

Materials and Methods

Figs. S1 to S7

Tables S1 to S6

Movie S1

Supplementary Captions for Fig. 1

References (33–47)

20 July 2015; accepted 25 January 2016
Measurement of gene regulation in individual cells reveals rapid switching between promoter states
Leonardo A. Sepúlveda, Heng Xu, Jing Zhang, Mengyu Wang and Ido Golding

*Science* **351** (6278), 1218-1222.
DOI: 10.1126/science.aad0635

**Stochastic properties of phage promoter**

Full understanding of regulated gene expression requires characterization of stochastic variation in the activity of individual promoters. To avoid cell-to-cell variability and variation between the activity of specific gene copies, Sepúlveda *et al.* investigated the behavior of the lysogeny maintenance promoter of phage lambda in individual *Escherichia coli* cells. They measured the concentration of transcription factor and the actual number of mRNAs produced, and used mathematical modeling to discern the stochastic activity of the regulated promoter. The promoter underwent switching between configurations that occurred more rapidly than the lifetime of mRNA molecules produced, and individual copies of the same gene functioned independently in the same cell. Such studies can reveal new aspects of systems that have been well studied by more conventional techniques.

*Science*, this issue p. 1218