Filtering input fluctuations in intensity and in time underlies stochastic transcriptional pulses without feedback

Alberto Stefano Sassi[a], Mayra Garcia-Alcala[b,c,d], Mark J. Kim[b], Philippe Cluzel[b,1], and Yuhai Tu[a,1,2]

[a]Department of AI Science, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598; [b]Department of Molecular and Cellular Biology, John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; [c]Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Cuernavaca 62210, México; and [d]Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Coyoacán 04510, Mexico City, Mexico

Edited by Herbert Levine, Northeastern University, Boston, MA, and approved September 14, 2020 (received for review May 27, 2020)

Stochastic pulsatile dynamics have been observed in an increasing number of biological circuits with known mechanism involving feedback control and bistability. Surprisingly, recent single-cell experiments in *Escherichia coli* flagellar synthesis showed that flagellar genes are activated in stochastic pulses without the means of feedback. However, the mechanism for pulse generation in these feedbackless circuits has remained unclear. Here, by developing a system-level stochastic model constrained by a large set of single-cell *E. coli* flagellar synthesis data from different strains and mutants, we identify the general underlying design principles for generating stochastic transcriptional pulses without feedback. Our study shows that an inhibitor (YdiV) of the transcription factor (FlhDC) creates a monotonic ultrasensitive switch that serves as a digital filter to eliminate small-amplitude FlhDC fluctuations. Furthermore, we find that the high-frequency (fast) fluctuations of FlhDC are filtered out by integration over a timescale longer than the timescale of the input fluctuations. Together, our results reveal a filter-and-integrate design for generating stochastic pulses without feedback. This filter-and-integrate mechanism enables a general strategy for cells to avoid premature activation of the expensive downstream gene expression by filtering input fluctuations in both intensity and time so that the system only responds to input signals that are both strong and persistent.

transcriptional pulses | gene regulation | stochastic modeling | filtering | ultrasensitivity

The gene regulatory network for controlling the bacterial flagellar synthesis and assembly has long served as a canonical example of transcriptional cascade (1–8), in which a particular gene product (protein) can serve as the transcription factor (TF) for other genes, which encode proteins that serve as TFs for other downstream genes, and so on. The bacterial flagellar promoters are organized in three classes, each of which underlies specific stages of the complex flagellar assembly and chemosensory signaling processes (9–11). The class 1 promoter controls the transcription of *flhDC*, the operon encoding for the proteins FlhD and FlhC, that assemble into the heterohexamer $FlhD_4C_2$ (refer to as FlhDC hereafter), which is the master regulator of flagellar synthesis. The master regulator serves as the TF for class 2 promoters and initiates the transcription of seven operons necessary for the assembly of the basal body and the hook of the flagellum (12, 13). A specific class 2 protein, FliA, activates class 3 promoters to express class 3 genes coding proteins for the flagellar filament and the chemotaxis signaling pathway (14). The bacterial flagellar synthesis represents a remarkable process in which over 50 different genes are expressed in a coherent fashion over several generations to build a large functional protein complex. Over the past decades, experimental studies have yielded many valuable insights about the transcriptional network that underlies flagellar synthesis in *Escherichia coli* and *Salmonella* (9, 10, 15–18). In particular, population measurements suggested that the transcriptional cascade follows a deterministic temporal order where genes in subsequent classes are activated sequentially and, once turned on, promoters remain active continuously during exponential growth (10).

Recently, Kim et al. (19) measured gene expression dynamics of single *E. coli* cells over many generations by using a microfluidic device called the "mother machine" (20, 21). Surprisingly, the new single-cell data showed that flagellar promoters are stochastically activated in intermittent pulses, that is, bursts of strong activity amid a quiet background. Due to their stochastic nature, transcriptional pulses can only be detected by single-cell measurements, and they were missed in previous cell population study (10).

Indeed, rapid developments in quantitative single-cell measurements have revealed that quasi-periodic or stochastic pulsating dynamics is a common dynamic activity pattern in biological circuits in a wide range of organisms, for example, TF Msn2 in yeast (22, 23), TF p53 (24) and NF-κB (25) in mammalian cells, bacterial TF σ_b in the bacterium *Bacillus* (26), and stochastic cell fate switching in microbes (27). Typically, the underlying mechanism involves bistability caused by a combination of positive and negative feedback loops (28); a recent theoretical study showed

Significance

Stochastic pulsatile dynamics are ubiquitous in organisms ranging from *Escherichia coli* to mammalian cells. The molecular mechanism for these stochastic pulses has long been thought to depend on feedback control that leads to bistability in the underlying system. Here, by quantitative modeling constrained by recent single-cell data in *E. coli* flagellar synthesis, we find that an inhibitor of the transcription factor creates a monotonic ultrasensitive switch that serves as a digital filter to eliminate small-amplitude input fluctuations. Additionally, the high-frequency input fluctuations are filtered by integration over time. Together, our results reveal that the filter-and-integrate design can generate stochastic pulses without feedback. This filter-and-integrate strategy allows cells to turn on gene expression only for persistently strong input signals.

Author contributions: A.S.S., P.C., and Y.T. designed research; A.S.S., M.G.-A., M.J.K., P.C., and Y.T. performed research; A.S.S. and Y.T. developed the model and performed the simulations; A.S.S., M.G.-A., M.J.K., P.C., and Y.T. analyzed data; and A.S.S., M.G.-A., P.C., and Y.T. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]P.C. and Y.T. contributed equally in supervising the project.

[2]To whom correspondence may be addressed. Email: yuhai@us.ibm.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2010849117/-/DCSupplemental.

First published October 12, 2020.

that stochastic pulses may also be caused by a negative feedback loop alone (29). Suffice to say, most known mechanisms for pulses involve some form of feedback control. However, in Kim et al. (19), pulses were present even when the promoter of FlhDC was deleted and replaced by a synthetic/constitutive promoter that does not respond to flagellar endogenous regulator and thus rules out the involvement of any feedback control. Additionally, in a different system that controls cell fate, stochastic pulses were also successfully generated by transferring, in *E. coli*, a reconstituted gene circuit without feedback (30).

Prompted by these recent observations, we ask the question, what are the underlying design principles that govern the stochastic transcriptional pulses without feedback? To answer this question, we use the large single-cell expression time series dataset for different strains and mutants from Kim et al. (19) to quantitatively and consistently constrain a system-level stochastic model. Using this approach, we aim to identify the minimum design principles necessary to generate stochastic transcriptional pulses without feedback in general, and specifically during *E. coli* flagellar synthesis.

Results

We first present a system-level (coarse-grained) stochastic model for the regulation of class 2 genes by FlhDC. We then use the model to analyze and explain the steady-state distributions of class 2 activity and its dynamics observed in single-cell experiments (19). Next, we propose possible molecular mechanisms underlying the main findings from our analysis and describe a filter-and-integrate mechanism for controlling noisy FlhDC signal in *E. coli*. Finally, we use our model to make testable predictions for responses to realistic time-varying signals.

A System-Level Stochastic Model for Class 1 and Class 2 Gene Expression Dynamics. In the experiments by Kim et al. (19), they used strains whose expression of class 1 genes was controlled by promoters of different strengths (see *SI Appendix, Supplementary Information A* for details of the experimental data). For each strain, long time series (~ 70 generations) of class 1 and class 2 promoter activity were simultaneously measured within the same cells using fluorescent proteins as proxy. They found that distributions of class 2 promoter activity depend on the class 1 promoter strength as well as the presence or absence of the inhibitor YdiV (31–33) [also known as RflP (34)].

Here, we use the class 1 reporter fluorescence signal as a proxy to represent the FlhDC level C_1 as the input to the class 2 gene transcription process. Following Kim et al. (19), we define the class 2 gene expression activity (output), A_2, as the time derivative of the fluorescence time series of the class 2 reporter gene, which is assumed to be proportional to the degree of activation of class 2 promoters (see *SI Appendix, Supplementary Information A* for details of computing and denoising A_2 from experimental data obtained in the "mother machine"). We propose a system-level stochastic model to describe dynamics of these two coarse-grained observables, A_2 and C_1, without considering molecular details. In particular, the A_2 dynamics in a single cell is described by a stochastic differential equation (SDE) that has three basic terms: decay (degradation), activation (production), and noise,

$$\frac{dA_2}{dt} = -\frac{A_2 - f_s(C_1)}{\tau_2} + \theta_s = -\frac{A_2}{\tau_2} + \frac{f_s(C_1)}{\tau_2} + \theta_s, \qquad [1]$$

where $s = (+, -)$ represents cells with and without YdiV, respectively, and τ_2^{-1} is the decay rate with τ_2 as the characteristic timescale associated with class 2 expression. For convenience, we write activation rate of the class 2 promoter as $f_s(C_1)/\tau_2$ so that $\langle A_2 \rangle = \langle f_s(C_1) \rangle$. We call $f_s(C_1)$ the response function, as it characterizes how class 2 activity responds to class 1 protein level C_1. The white noise θ_s has zero mean $\langle \theta_s(t) \rangle = 0$ and correlation $\langle \theta_s(t)\theta_s(t') \rangle = 2\Delta_s(\langle C_1 \rangle)\delta(t - t')$ with noise strength given by $\Delta_s(\langle C_1 \rangle)$. Here, we assume that the noise strength $\Delta_s(\langle C_1 \rangle)$ depends only on the average C_1 concentration $\langle C_1 \rangle$, which is valid due to the timescale difference for A_2 and C_1, as will become clear later in this work.

The activity of class 2 genes A_2 depends on the class 1 protein level C_1, whose dynamics can be modeled by a similar SDE,

$$\frac{dC_1}{dt} = -\frac{(C_1 - \mu)}{\tau_1} + C_1\eta = -\frac{C_1}{\tau_1} + \frac{\mu}{\tau_1} + C_1\eta, \qquad [2]$$

where τ_1 is the decay time for class 1 protein. We define μ/τ_1 as the C_1 production rate, so that the steady-state average is $\langle C_1 \rangle = \mu$ with μ depending on the promoter strength of a specific strain. The white noise η has zero mean $\langle \eta(t) \rangle = 0$ and $\langle \eta(t)\eta(t') \rangle = 2\Delta_1\delta(t - t')$, with the noise strength Δ_1 as a constant for all promoter strengths. The noise η is multiplied by C_1 in Eq. **2**. This choice of a linear multiplicative noise is due to the fact that it provides a better fit to experimental data in comparison with other forms of noise such as additive and square root multiplicative noise (see *SI Appendix, Supplementary Information B* and Fig. S2 for a detailed comparison). Mechanistically, the multiplicative noise may originate from random partitioning of protein molecules during cell division and fluctuations in protein degradation rate.

An important assumption of our proposed model is that the observed class 2 pulses do not require transcriptional feedback on YdiV. Indeed, this assumption is fully supported by mother machine single-cell data published in Kim et al. (ref. 19, figure S(11)), which shows that, when the expression of YdiV is controlled by a synthetic constitutive promoter, the pulses are still present. Overall, our "minimal" system-level (coarse-grained) model consists of two SDEs (Eqs. **1** and **2**) with a few physiologically meaningful parameters and functions: two timescales τ_1 and τ_2, a constant multiplicative noise strength Δ_1, a strain-dependent noise strength $\Delta_s(\mu)$, and two response functions $f_s(C_1)$. In the following, by fitting all of the single-cell data for different strains (19) with our model quantitatively, we determine the parameters and response functions, which help us identify the underlying mechanism for transcriptional pulses.

YdiV Creates an Ultrasensitive Switch. We first use our model to study all of the steady-state class 2 activity distributions from experiments (see *Methods* for details on simulations of our model). In Fig. 1*A*, experimentally observed distributions of the normalized class 2 activity $A_2/\langle A_2 \rangle$ for cells with and without YdiV for three class 1 promoter strengths—low ($P1$, red), medium ($P4$, green), and high ($P7$, blue)—are shown (see *SI Appendix, Supplementary Information A* for details of the experimental dataset). The most drastic difference between cells with and without YdiV is for strain $P4$ that has a medium promoter strength similar to that of the wild-type cells. The $A_2/\langle A_2 \rangle$ distribution for the $P4$ strain in the presence of YdiV (green line in Fig. 1 *A*, *Left*) is strongly asymmetric with a peak that is significantly shifted with respect to the mean, whereas the distributions for lower and higher promoter strengths are more symmetric. This asymmetry is largely absent in strains without YdiV (Fig. 1 *A*, *Right*). For simplicity, we fit the experimental data by using Hill functions for the response functions f_s and tuning the shape of the Hill functions to minimize the difference between the distributions of C_1 and A_2 from the model and those from experiments for different promoter strengths (see *SI Appendix, Supplementary Information C* for details). Our model can explain all of the steady-state A_2 distributions from different strains together as shown by the direct comparison between the experimentally measured distributions (Fig. 1*A*) and those obtained from our model (Fig. 1*B*).

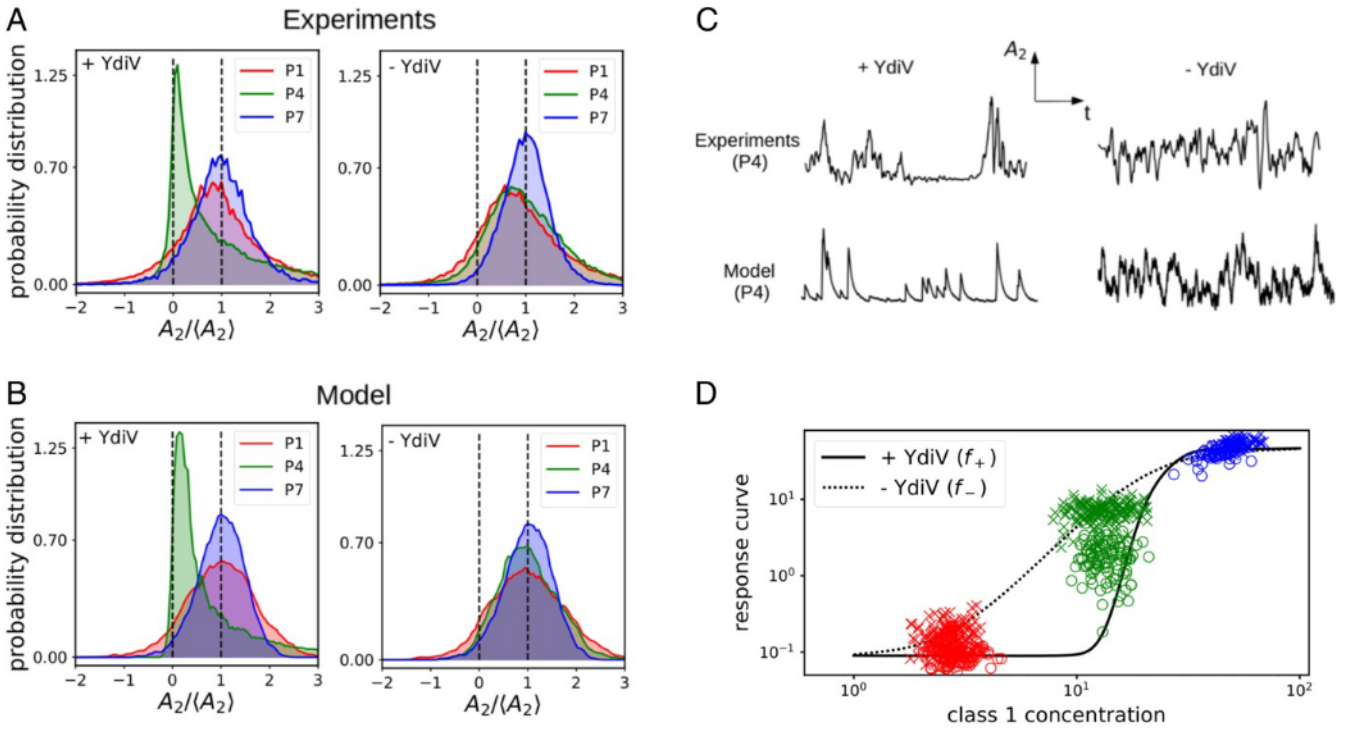

Fig. 1. Comparison between experimental data and model results. (*A*) Distributions of the normalized class 2 activity $A_2/\langle A_2\rangle$ in the presence (*Left*) and absence (*Right*) of YdiV for small (*P1*, red), medium (*P4*, green), and large (*P7*, blue) class 1 promoter strengths. Data are taken from ref. 19. We use the black dashed lines to indicate the positions of its mean value and zero for A_2. (*B*) The corresponding distributions obtained from our model. (*C*) Comparison of typical time series of $C_1(t)$ and $A_2(t)$ from experiments and model for the *P4* strain with and without YdiV. (*D*) Response functions obtained by fitting our model to experimental data (19) with (f_+, solid line) and without (f_-, dashed line) YdiV. The circles (+YdiV) and crosses (−YdiV) indicate the mean value from different mother cells in experiments. Note that f_s does not pass through the center of the data points, due to its high nonlinearity ($\langle f_s(C_1)\rangle \neq f_s(\langle C_1\rangle)$). The color code for each promoter is the same as in *A* and *B*.

The most significant difference in the dynamics of class 2 activity $A_2(t)$ between cells with and without YdiV also occurs in strain $P4$ with a medium class 1 promoter strength similar to that of wild-type cells. As shown in Fig. 1*C*, from the experiments, the A_2 dynamics contains well-defined pulses in the presence of YdiV, while pulses are not well separated and the presence of higher-frequency noise appears to be unfiltered in the absence of YdiV. These different activity dynamics are reproduced in our model.

By fitting our model to the single-cell data (see *SI Appendix, Supplementary Information C* for details of the fitting and values of all of the fitted model parameters), we obtain the response functions, $f_+(C_1)$ and $f_-(C_1)$, which cannot be directly determined from experimental data because $\langle A_2\rangle = \langle f_s(C_1)\rangle \neq f_s(\langle C_1\rangle)$, due to their nonlinearity. As shown in Fig. 1*D* (solid line), the response function $f_+(C_1)$ in the presence of YdiV behaves like a monotonic ultrasensitive switch, that is, it has a much larger Hill coefficient ($H_+ \approx 12$) than the Hill coefficient ($H_- \approx 3$) for $f_-(C_1)$ (see *SI Appendix, Supplementary Information C* for details). Due to this ultrasensitive (steep) response function, when the mean FlhDC concentration is close to the threshold (e.g., for strain $P4$), the distribution of $A_2/\langle A_2\rangle$ has the largest asymmetry, and there is a significant shift in the peak of the distribution (Fig. 1*A*, *Left* and *B*, *Left*). In the absence of YdiV, the response function $f_-(C_1)$ is found to be less steep (dotted line in Fig. 1*D*), and the probability distributions are more symmetric (Fig. 1 *A*, *Right* and *B*, *Right*). Since we only used experimental data from three strains ($P1, P4, P7$) to fit (calibrate) our model to determine the response functions and other parameters (such as τ_1 and τ_2), we can test the consistency of our model by applying it to other strains that are not used in the fitting. In particular, we ran simulations of our model for two other strains, $P2$ and $P5$, with different promoter strengths but with all other parameters and the response functions fixed. As shown in *SI Appendix*, Fig. S5, the A_2 distributions for $P2$ and $P5$ obtained from our model agree well with experiments, and the average values of C_1 are also consistent with the experiments (*SI Appendix*, Fig. S4).

The agreements between our model and the experiments not only validate our model, but, more importantly, they reveal that the role of YdiV is to create a ultrasensitive (steep) response of the class 2 activity to the class 1 protein (FlhDC) concentration.

Memory Effect and Integration over Time. Next, we use our model to analyze the stochastic gene expression dynamics in the flagellum system. As shown in Eq. **1**, there are two noise sources for the class 2 activity A_2: the intrinsic noise θ_s in the gene expression process and the extrinsic noise due to fluctuations of the FlhDC concentration C_1. From our model, the mean class 2 promoter activity averaged over the intrinsic noise can be written in the following integral form:

$$\langle A_2(t)\rangle = \int_{-\infty}^{t} G(t-t') f_s(C_1(t'))\, dt', \qquad [3]$$

where $G(t-t') = \tau_2^{-1}\exp[-(t-t')/\tau_2]$ is the integral kernel (Green's function) that decays with the timescale τ_2. Eq. **3** means that $A_2(t)$ depends on the values of $C_1(t')$ in a time window $t-\tau_2 \le t' \le t$; that is, there is a memory effect. Given that $C_1(t)$ varies with a timescale τ_1, the strength of this memory effect depends on the ratio of the two timescales τ_2/τ_1. For $\tau_2/\tau_1 \ll 1$, the kernel $G(t-t') \approx \delta(t-t')$ and $\langle A_2(t)\rangle \approx f_s(C_1(t))$; that is, the class 2 activity depends on the instantaneous level of class 1

protein $C_1(t)$ with negligible memory effect. However, if $\tau_2/\tau_1 > 1$, the memory effect becomes significant. In the limit $\tau_2/\tau_1 \gg 1$, $\langle A_2(t)\rangle \approx \langle f_s(C_1)\rangle_{C_1} \equiv \int f_s(C_1)P(C_1)dC_1$ becomes a constant independent of the instantaneous FlhDC level $C_1(t)$ ($P(C_1)$ is the steady-state C_1 distribution function).

To demonstrate the memory effect, we divide the FlhDC concentration in equally spaced bins (in log-scale) and then compute the class 2 activity and FlhDC concentration averaged within each bin to form the so-called "bin-averaged response curve" (BARC) for each strain with a different class 1 promoter strength (see *Methods* for details on BARC). As shown in Fig. 2*A*, when $\tau_2 > \tau_1$, the BARCs for different strains can be shifted from each other, consistent with experiments. This means that the class 2 activity is not determined just by the instantaneous FlhDC concentration; it depends also on its average over previous times, due to the memory effect. Indeed, as shown in Fig. 2*B*, in the absence of memory effect when $\tau_2 \ll \tau_1$, all BARCs collapse.

When $\tau_2 > \tau_1$, A_2 depends on an integral of the transformed input signal $f_+(C_1)$ over a time window τ_2 as shown in Eq. **3**. This integration effectively averages the input signal in time and eliminates high-frequency fluctuations as demonstrated in Fig. 2 *C* and *D*, where power spectra for the two cases with and without integration are shown. The power spectrum from experimental data is consistent with our model, with $\tau_2 = 3.5\tau_1$ where the power of the spectrum of $A_2(t)$ is concentrated in the low-frequency regime with a similar half-maximum-power frequency $f_{1/2} \approx 0.002/\text{min}$ as shown in Fig. 2*C*. For $\tau_2 \ll \tau_1$ (Fig. 2*D*), the power spectrum is much broader, with significant high-frequency fluctuations and $f_{1/2} \approx 0.008/\text{min}$ that is much higher than that from experiments. Quantitatively, by comparing BARCs and autocorrelation functions from the model and those from experiments, we estimate the timescale ratio to be in the range $2.1 \leq \tau_2/\tau_1 \leq 3.6$ (see *SI Appendix, Supplementary Information F* for details).

A YdiV-Mediated Sequestration Mechanism for Filtering. So far, we have used our model to explain the steady-state distributions and dynamics of the noisy single-cell data. In this section, we propose possible molecular mechanisms underlying our findings, with a focus on the role of YdiV in regulating FlhDC signal.

We first briefly discuss possible molecular origins of separation of timescales. The shorter timescale τ_1 is most likely related to the degradation time of the FlhC, and FlhD monomers and the longer timescale τ_2 are related to the decay time of the FlhDC-dependent class 2 promoter activation (see Fig. 3*A* for an illustration). Since FlhD and FlhC can only serve as functional TF when they form heterohexamer FlhDC, τ_2 is affected by the binding affinity of FlhDC to class 2 promoters as well as stability of the FlhDC complex. The heterohexamer is more stable than the monomers (35), which is consistent with $\tau_2 > \tau_1$ as observed in experiments and our model. It would be interesting to test the effects of changing τ_2 experimentally (see *Summary and Discussion* for more details).

We now study the role of YdiV in creating an ultrasensitive switch-like response of class 2 gene expression as shown in Fig. 1*D*. YdiV is known to have two main effects on FlhDC. First, by occupying the binding sites of FlhDC, YdiV prevents the binding of FlhDC to free class 2 promoters (32). Second, YdiV mediates the interaction between FlhDC and the degradation complex ClpXP (31, 32). Here, we only consider the effect of competitive binding on the response function; similar effect may be achieved by enhancing the degradation of FlhDC, which is not considered in this study. Intuitively, due to competitive binding, YdiV can introduce a threshold for FlhDC by sequestrating FlhDC to prevent it from binding to and activating the class 2 promoters.

To show the qualitative effects of this YdiV-mediated sequestration mechanism, we developed a simple model based on competitive binding of FlhDC by YdiV and the class 2 promoter as illustrated in Fig. 3*A* (see *Methods* for details of the sequestration model). In Fig. 3*B*, the fraction of promoters bound to FlhDC, which serves as a measure of the class 2 gene expression activity, is shown as a function of the scaled FlhDC concentration C_1/P_t, with P_t as the total promoter concentration, for different values of total YdiV concentration. We can see that higher YdiV concentrations lead to delayed but steeper response curves, which agrees qualitatively with the response curves

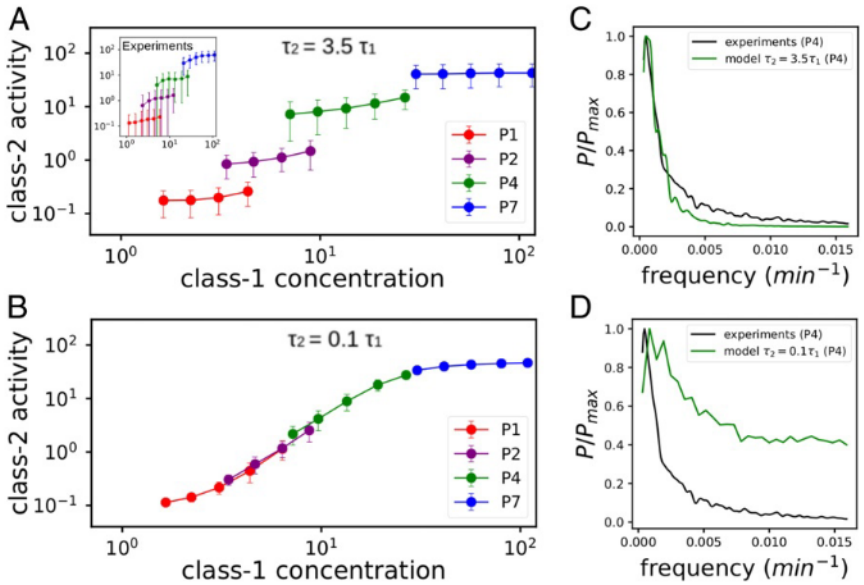

Fig. 2. The bin-averaged response curves and the ratio of two timescales τ_2/τ_1. (*A*) Model results for $\tau_2 = 3.5\tau_1$. *Inset* shows the same behavior from experimental data. (*B*) Model results for $\tau_2 = 0.1\tau_1$. We used −YdiV strains here; see *SI Appendix*, Fig. S7 for similar results for +YdiV strains. (*C* and *D*) The normalized power spectra (green lines) from our model for (*C*) $\tau_2 = 3.5\tau_1$ and (*D*) $\tau_2 = 0.1\tau_1$ for the *P4* strain. The normalized power spectrum (black line) from experiments (*P4*) is also shown for comparison. We used $\tau_1 = 45$ min within the range of C_1 correlation time estimated from experiments.

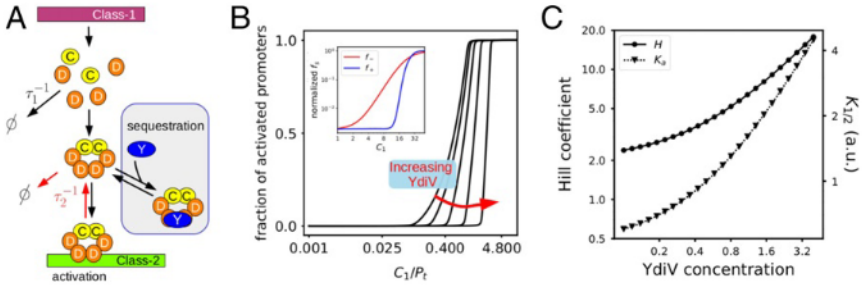

Fig. 3. The YdiV-mediated sequestration mechanism for switch like behavior. (*A*) Schematic representation of the mechanism of transcriptional regulation of class 2 promoters by FlhDC. The class 1 promoter (purple) initiates the expression of the monomers FlhC and FlhD (yellow and orange, respectively) that forms the heterohexamer $FlhD_4C_2$ (FlhDC). The timescale τ_1 may be related to the degradation rate of the monomers (FlhD and FlhC), while τ_2 may be related to the degradation rate of FlhDC or its unbinding rate to the promoters, both labeled by red arrows. The activation of class 2 genes, which depends on binding of the FlhDC complex to the class 2 promoter (green), is inhibited when FlhDC is sequestered by binding to YdiV (blue) instead (gray box). (*B*) Response curves obtained from the YdiV sequestration model for different values of YdiV concentrations (P_t is the total class 2 promoter concentration); see *Methods* for details. (*Inset*) The response functions, $f_+(C_1)$ (blue) and $f_-(C_1)$ (red), in the presence and absence of YdiV, respectively, from fitting experimental data (the same as in Fig. 1*D*). (*C*) Hill coefficient (H) and the half-maximum concentration ($K_{1/2}$) as a function of the YdiV concentration.

f_+ and f_- shown in Fig. 3 *B*, *Inset* for comparison. A response curve can be described by its effective Hill coefficient H and its half-maximum concentration $K_{1/2}$. From the experimental data and our model, both the Hill coefficient and $K_{1/2}$ are larger in wild-type cells than those in the $\Delta ydiV$ mutant. As shown in Fig. 3*C*, the YdiV-mediated sequestration model reproduces this general trend, with both H and $K_{1/2}$ increasing with YdiV concentration.

The steep response function $f_+(C_1)$ suggests that YdiV serves (approximately) as a digital filter. When the input signal, that is, FlhDC concentration (C_1), is lower than a threshold $C^* \equiv K_{1/2}(Y_t)$ set by the YdiV concentration Y_t, the normalized output $f_+(C_1)/max(f_+)$ is close to 0. When $C_1 > C^*$, the normalized output is close to 1.

One important ingredient in our model is the inhibitory effect of YdiV on class 2 promoter activity. Therefore, changing YdiV at a fixed average FlhDC level can modulate the class 2 promoter activity. In particular, the model predicts that the increase in YdiV levels raises the threshold of class 1 concentration that needs to be reached to have a nonnegligible class 2 response. More technically, the constant $K_{1/2,+}$ increases with the concentration of YdiV. This results in a shift in the class 2 distributions toward lower mean values for larger YdiV levels. We have tested this prediction by using flow cytometry to quantify the distributions of class 2 expression in two strains, ProBY and ProDY (36), which have higher levels of YdiV expression than in wild type (see *SI Appendix*, *Supplementary Information E* for details of the experiments). As shown in *SI Appendix*, Fig. S6, we find that, when the YdiV level is increased, there is a shift in the class 2 gene expression distributions toward lower values consistent with our model results.

A Filter-and-Integrate Mechanism for Stochastic Pulses without Feedback. Put together, our results show that *E. coli* uses a combined filter-and-integrate strategy (mechanism) to control its class 2 gene expression. As shown in Fig. 4, in the presence of YdiV (left side of Fig. 4), the input signal, that is, the FlhDC concentration $C_1(t)$ (blue line), which fluctuates relatively fast at a short timescale τ_1, first goes through a highly nonlinear ultrasensitive response function $f_+(C_1)$, which serves approximately as a digital filter that eliminates the FlhDC fluctuations lower than certain threshold C^* set by the YdiV concentration. The filtered signal $f_+(C_1(t))$ (green line) has a nearly digital stochastic pulse-like pattern.

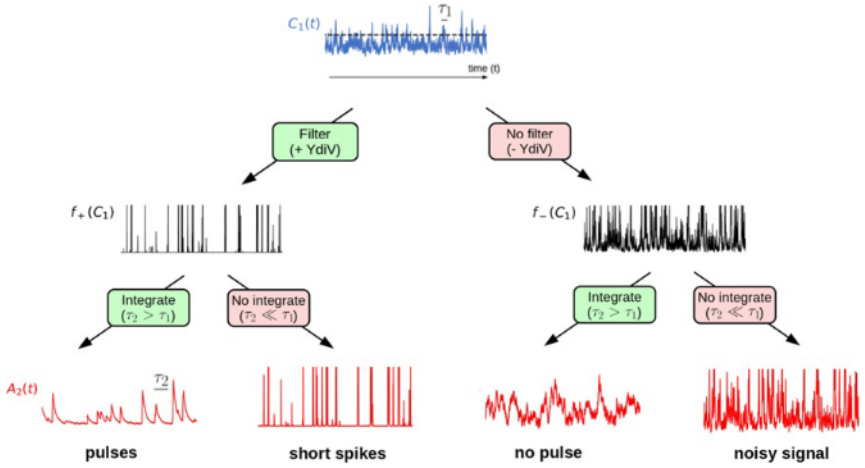

Fig. 4. The filter-and-integrate mechanism for pulse generation. In the presence of YdiV, the FlhDC concentration $C_1(t)$ (blue line), which fluctuates with a fast timescale τ_1, is filtered by the inhibitory effect of YdiV. The filtered signal $f_+(C_1(t))$ (black line), which has a pulse-like pattern, is then integrated over a longer timescale τ_2 to determine the class 2 activity $A_2(t)$ (red line). In the absence of YdiV, there is no filtering effect, and the resulting A_2 dynamics is noisier, with no pulse. Without integration ($\tau_2 \ll \tau_1$), A_2 would have many short spikes when the signal is filtered, and it resembles the original noisy signal (C_1) in the absence of filtering. Results presented here are obtained from simulations of our model using parameters fitting the *P4* strain.

The second part of the two-pronged process is to integrate the filtered signal over a timescale $\tau_2(>\tau_1)$, which gives rise to the intermittent "strong" pulses in A_2 (red line). Without the integration process $\tau_2 \ll \tau_1$, the A_2 activity would have many short spikes, too short to complete the flagellar synthesis process. The integration process acts as a filter in frequency space, as demonstrated in Fig. 2 *C* and *D*, to suppress those FlhDC fluctuations that are strong in intensity but short in time duration. As a result, class 2 gene expression is turned on only when there are persistently high levels of FlhDC signal. Due to the same integration (memory) effect, once turned on, the class 2 expression remains active for a time duration of $\sim\tau_2$ even without strong instantaneous FlhDC signal. This long duration of on time is necessary for the cell to finish the flagellar synthesis process that lasts for several generations.

Both filtering and integration in time are critical to generate the stochastic pulses observed in experiments (19) where A_2 alternates between having near-zero activity (quiet) for a long period and being highly active during a strong pulse that lasts for a long timescale set by τ_2 (bottom left in Fig. 4). In the absence of YdiV, the noisy $C_1(t)$ signal is not filtered, and, as a result, there is no pulse in $A_2(t)$, which fluctuates continuously (bottom right in Fig. 4). In the absence of integration, A_2 follows the same dynamics as the filtered signal f_+, which has many short spikes (bottom middle in Fig. 4).

Predicted Gene Expression Responses to Time-Varying Signals. Now that our model is fully developed with all its parameters determined by existing single-cell experiments (19), it provides a powerful tool to predict class 2 gene expression responses to more realistic time-dependent signals without any tuning parameter. As an example, we use our model to study the responses of class 2 activity to an increase of the class 1 promoter strength from a low value $\mu_0(=\mu(P1)\approx 2.6)$ to a series of higher values μ_1 at time $t=0$. We find that a cell will turn on the class 2 gene expression after a delay time τ_r, as shown in Fig. 5*A*. The delay time τ_r varies from cell to cell and follows a distribution $P(\tau_r|\mu_1)$ that depends on μ_1. In Fig. 5*B*, the delay time distributions for different values of μ_1 are shown. Both the average delay time and its variance decrease with μ_1 or, equivalently, the average FlhDC level as shown in Fig. 5*C*. Responses to other time-varying stimuli such as ramps and oscillatory signals can be studied by our model similarly.

Our results suggest that, even though FlhDC concentration does not control the class 2 gene expression deterministically, the filter-and-integrate mechanism allows the cell to control the timing statistics of the class 2 expression based on the average FlhDC signal intensity. As shown in Fig. 5*C*, a higher FlhDC concentration induces (on average) a shorter and more accurate (smaller coefficient of variation) onset time to turn on the class 2 genes. Similarly, an elevated average FlhDC level leads to more frequent occurrence of the transcriptional pulses. These model predictions can be tested in future single-cell experiments by using inducible class 1 promoters.

Summary and Discussion

In this paper, dynamics of the class 1 and class 2 promoters in *E. coli* flagellar synthesis are studied by using a minimal stochastic model that captures the essential characteristics of the underlying system. Applying our stochastic model to a large single-cell dataset quantitatively shows that the transcriptional pulses are generated by a filter-and-integrate mechanism without the need of feedback control. Both the YdiV-enabled filtering and the memory-enabled time integration are crucial to average out FlhDC fluctuations in intensity and in time so that an individual cell can make a "calculated" (informed) decision on whether to turn on the expensive class 2 gene expression. This mechanism or a variation of it may be at work in other pulse-generating systems without feedback. For example, in the reconstituted SinI–SinR circuit (30), the antagonist SinI may play the critical role of the filter (analogous to YdiV) by binding to and inhibiting the transcriptional regulator SinR (analogous to FlhDC). On the other hand, while the timescale of the class 2 activity (A_2) pulses in flagellar synthesis is provided by the integration time τ_2, the pulse timescale in the SinI–SinR system may be controlled directly by the timescale of the SinI fluctuations (30).

To better understand the feedbackless filter-and-integrate strategy in *E. coli*, we compare it with a closely related bacterium *Salmonella*, which does have an additional positive feedback loop in controlling its class 2 promoter activity. As illustrated in Fig. 6, FliZ, a class 2 protein, can inhibit YdiV in *Salmonella* but not in *E. coli*. This positive feedback loop in *Salmonella* leads to bistability in class 2 gene response (37). As a result, the hysteretic response function in *Salmonella* $f_{Sal}(C_1)$ has two critical points C_l^* and $C_h^*(>C_l^*)$, as illustrated in Fig. 6*B*. Due to the bistability, once the FlhDC level goes over the upper critical level $C_1>C_h^*$ even for a short time, the response signal $f_{Sal}(C_1)$ will stay on (green) for a much longer time, as long as FlhDC level is higher than the lower critical level $C_1>C_l^*$. This amplification of the "on" time duration allows *Salmonella* to activate class 2 gene expression upon detecting a large-but-short signal. In contrast, as shown in Fig. 6*A*, the response function for *E. coli*, though steep, does not have bistability, because there is no feedback. As a result, an elevated FlhDC level over the threshold (C^*) for a short time is not enough to trigger significant class 2 activity, given the large integration time τ_2.

Their different strategies may reflect the different purposes of motility for these two bacteria. *Salmonella* is a pathogen

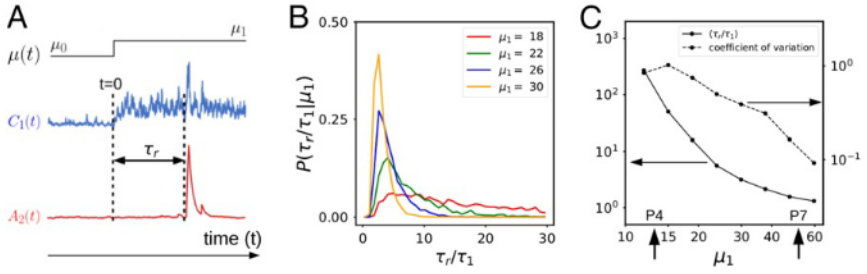

Fig. 5. The predicted responses to step stimuli. (*A*) Predicted dynamics of C_1 and A_2 in response to a time-dependent (step function) stimulus in which the class 1 promoter strength μ is changed at $t=0$ from a low value $\mu_0=\mu(P1)=2.6$ to a higher value μ_1. The delay time τ_r is the time duration from the stimulus ($t=0$) to the onset of class 2 expression, which is defined as the time when A_2 crosses a threshold $A_2^*=\max(f_+)/4$ (qualitative results do not depend on the choice of A_2^*). (*B*) Distributions of the delay time τ_r (normalized by τ_1) for different values of μ_1. (*C*) Both the average and variance of τ_r decreases with μ_1 or the average FlhDC concentration. The values of μ for strains *P4* and *P7* used in our study are marked for reference.

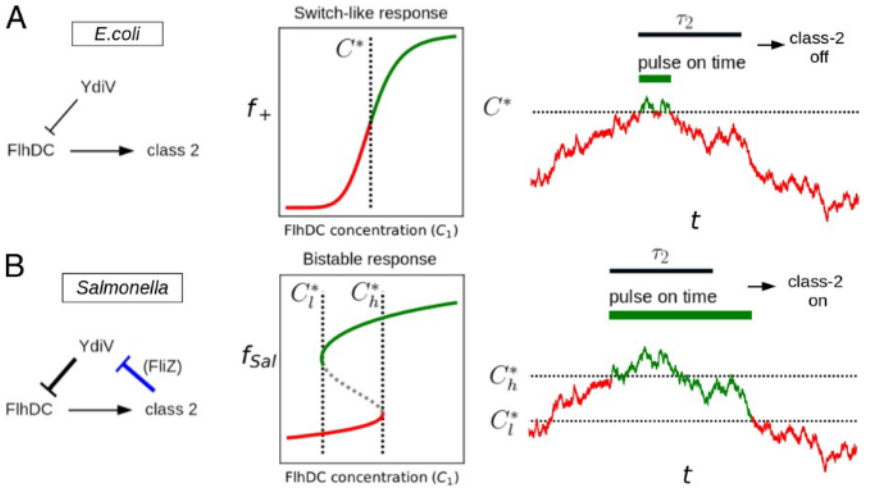

Fig. 6. The different gene regulation strategies for *E. coli* and *Salmonella*. (*A*) The *E. coli* network has the YdiV-mediated inhibition of the class 1 master regulator (FlhDC) but not the feedback mechanism from class 2 gene. As a result, *E. coli* has a steep but monotonic response function f_+ with a threshold C^*; that is, class 2 gene expression is activated (green) when $C_1 > C^*$ or inhibited (red) when $C_1 < C^*$. For an FlhDC signal that passes the threshold C^* briefly, the class 2 activity is only turned on for a short time, which is not enough to start the flagella synthesis, given the long integration time τ_2 for class 2 gene activation. (*B*) In *Salmonella*, there is a feedback (blue line) from a class 2 protein (FliZ) to suppress the inhibitor YdiV. This feedback mechanism leads to a bistable (hysteretic) response of class 2 genes in the range $C_l^* < C_1 < C_h^*$. As a result, a brief period of elevated FlhDC concentration over the (upper) threshold C_h^* can cause a prolonged class 2 gene activation as long as $C_1 > C_l^*$, which can lead to flagella synthesis.

and activates the flagellar gene expression cascade in nutrient-rich conditions. A short pulse of strong stimulus from the host may trigger the flagellar synthesis process. This aggressive strategy may be optimized for host colonization and infection. In *E. coli*, however, flagellar synthesis is activated in nutrient-poor conditions for the purpose of foraging and searching for nutrients. Therefore, *E. coli* may adopt a more prudent (filter-and-integrate) strategy that only triggers the expensive flagellar synthesis process when the cell experiences persistent poor nutrient conditions.

Our current analysis opens up a few directions for future investigation. Experimentally, effects of the filter-and-integration mechanism may be tested by perturbing the two key factors, for example, by tuning the strength of the interaction between FlhDC and class 2 promoters (32, 38) or by reducing the effective concentration of YdiV by adding *Salmonella* FlhC known to have weaker affinity for *E. coli* promoters but the same affinity for YdiV as *E. coli* FlhC (39). It would be highly informative to measure single-cell gene expression dynamics of *Salmonella* in the mother machine as done for *E. coli* (19), to understand its gene regulation strategy quantitatively. In our current model, YdiV level is assumed to be constant as we focus on the long timescale (low frequency) dynamics of the system. However, in the high-frequency regime, the A_2 power spectrum from experiments follows $P(f) \approx f^{-1}$ (f is the frequency), which decays slower than the power spectrum from our current model with a fixed YdiV level (*SI Appendix*, Fig. S10). A similar $1/f$ noise spectrum was observed in flagellar motor switching dynamics (40) and was subsequently explained by considering fluctuations of the response regulator CheY-P in a single cell (41). It would be interesting to include dynamics of YdiV in our model to explore whether it also leads to the observed $1/f$ spectrum in the high-frequency regime. Furthermore, to fully understand how the class 2 gene expression response depends on YdiV, molecular details of the FlhDC heterohexamer complex formation, its binding (unbinding) to DNA, and its degradation mediated by YdiV need to be considered in addition to the YdiV-mediated sequestration studied here.

Finally, our work here demonstrates that system-level stochastic modeling augmented by quantitative single-cell measurements for different strains and mutants provides a powerful tool to decipher principles of gene regulation from inherently noisy single-cell data. This tool becomes necessary when dealing with complex gene regulatory circuits such as the full bacterial flagellar gene expression cascade with multiple feedback controls. The general system-level stochastic modeling approach with coarse-grained variables that can be directly compared with experiments should be applicable to study stochastic expression dynamics in other gene regulatory systems with or without feedback (23–26, 30, 42).

Methods

Simulation of the Stochastic Differential Equations. For the numerical simulation of the stochastic equations, we write Eq. **2** in discrete form,

$$\Delta C_1 = -(C_1 - \mu)\frac{\Delta t}{\tau_1} + \gamma_\eta C_1 \sqrt{2\Delta_1 \Delta t}, \qquad [4]$$

where γ_η is a stochastic variable following a Gaussian distribution $\mathcal{N}[0, 1]$. We let $\tau_1 = 1$ to set the timescale, and the time step $\Delta t = 0.004\tau_1$. The discrete form of Eq. **1** is

$$\Delta A_2 = -(A_2 - f_s(C_1))\frac{\tau_1}{\tau_2}\Delta t + \gamma_\theta \sqrt{2\Delta_s \tau_1 \Delta t}, \qquad [5]$$

where γ_θ is a Gaussian random number as γ_η.

Bin-Averaged Response Curves and Flatness. To obtain the BARCs, we divide the full interval $10^{-1} \le C_1 \le 10^2$ into 20 logarithmically spaced bins and calculate the average of class 1 concentration and class 2 activity for all of the data points in each bin. This bin-averaged class 2 activity versus the bin-averaged class 1 concentration is the BARC. The BARCs in Fig. 2 are only shown at those bins that contain more than 5,000 data points, for better statistics.

Steady-State Solution of the Sequestration Model. Let P_t be the concentration of promoters, C_1 be the total concentration of FlhDC, and Y_t be the YdiV concentration. We denote C_{1p} and C_{1y} as the concentrations of FlhDC that are bound to the promoters and YdiV, respectively. In steady state, the balance of binding and unbinding reactions leads to $C_{1p} = C_{1f}P_t/C_{1f} + K_p$, and $C_{1y} = C_{1f}Y_t/C_{1f} + K_y$, where K_p and K_y are the dissociation constants of FlhDC to the promoters and YdiV, respectively. $C_{1f}(=C_1 - C_{1p} - C_{1y})$ is the free FlhDC concentration, and it satisfies the following equation (conservation of FlhDC):

$$C_{1f} + \frac{C_{1f}P_t}{C_{1f} + K_p} + \frac{C_{1f}Y_t}{C_{1f} + K_y} = C_1, \qquad [6]$$

which can be solved to determine C_{1f} as a function of P_t, C_1, Y_t, K_y, and K_p. The fraction of FlhDC bound promoters is $C_{1p}/P_t = C_{1f}/(K_p + C_{1f})$, which is used as a measure of the class 2 activity. The dependence of C_{1p}/P_t on C_1/P_t is plotted in Fig. 3*B*. Parameters $K_y/P_t = 10^{-2}$ and $K_p/P_t = 10^{-1}$ are used in Fig. 3 *B* and *C*.

Data Availability. All study data are included in the article and *SI Appendix*.

ACKNOWLEDGMENTS. We thank Drs. Y. Lin and M. Erhardt for discussions. The work by A.S.S. and Y.T. is supported by NIH Grant R35GM131734 to Y.T. The work by M.G.-A., M.J.K., and P.C. is supported by NIH Grant R01GM134275 to P.C. and NSF Grant 1615487 to P.C. M.G.-A. acknowledges support from Consejo Nacional de Ciencia y Tecnología (CONACYT) through a PhD scholarship and Dirección General de Asuntos del Personal Académico - Universidad Nacional Autónoma de México (DGAPA-UNAM) for support through Grant PAPIIT-IN226917.

1. A. Martinez-Antonio, S. Janga, D. Thieffry, Functional organisation of *Escherichia coli* transcriptional regulatory network. *J. Mol. Biol.* **381**, 238–247 (2008).
2. J. Bähler, A transcriptional pathway for cell separation in fission yeast. *Cell Cycle* **4**, 39–41 (2005).
3. D. W. Allan, S. Thor, Transcriptional selectors, masters, and combinatorial codes: Regulatory principles of neural subtype specification. *Wiley Interdiscip. Rev. Dev. Biol.* **4**, 505–528 (2015).
4. I. Spöring *et al.*, Regulation of flagellum biosynthesis in response to cell envelope stress in *Salmonella enterica* serovar typhimurium. *mBio* **9**, e00736-17 (2018).
5. O. A. Soutourina, P. N. Bertin, Regulation cascade of flagellar expression in Gram-negative bacteria. *FEMS Microbiol. Rev.* **27**, 505–523 (2003).
6. F. Jacob, J. Monod, Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–56 (1961).
7. S. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64–68 (2002).
8. M. Ghafari, A. Mashaghi, On the role of topology in regulating transcriptional cascades. *Phys. Chem. Chem. Phys.* **19**, 25168–25179 (2017).
9. G. S. Chilcott, K. T. Hughes, Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar typhimurium and *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **64**, 694–708 (2000).
10. S. Kalir *et al.*, Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* **292**, 2080–2083 (2001).
11. D. Apel, M. G. Surette, Bringing order to a complex molecular machine: The assembly of the bacterial flagella. *Biochim. Biophys. Acta* **1778**, 1851–1858 (2008).
12. T. Ikebe, S. Iyoda, K. Kutsukake, Promoter analysis of the class 2 flagellar operons of *Salmonella*. *Genes Genet. Syst.* **74**, 179–183 (1999).
13. X. Liu, P. Matsumura, Differential regulation of multiple overlapping promoters in flagellar class II operons in *Escherichia coli*. *Mol. Microbiol.* **21**, 613–620 (1996).
14. D. M. Fitzgerald, R. P. Bonocora, J. T. Wade, Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. *PLoS Genet.* **10**, e1004649 (2014).
15. Y. Komeda, Transcriptional control of flagellar genes in *Escherichia coli* K-12. *J. Bacteriol.* **168**, 1315–1318 (1986).
16. K. Kutsukake, Y. Ohya, T. Iino, Transcriptional analysis of the flagellar regulon of Salmonella typhimurium. *J. Bacteriol.* **172**, 741–747 (1990).
17. J. D. Brown *et al.*, The rate of protein secretion dictates the temporal dynamics of flagellar gene expression. *Mol. Microbiol.* **70**, 924–937 (2008).
18. S. Saini, J. D. Brown, P. D. Aldridge, C. V. Rao, FliZ Is a posttranslational activator of $FlhD_4C_2$-dependent flagellar gene expression. *J. Bacteriol.* **190**, 4979–4988 (2008).
19. J. M. Kim, M. Garcia-Alcala, E. Balleza, P. Cluzel, Stochastic transcriptional pulses orchestrate flagellum biosynthesis in *E. coli*. *Sci. Adv.* **6**, eaax0947 (2020).
20. 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).
21. P. Wang *et al.*, Robust growth of *Escherichia coli*. *Curr. Biol.* **20**, 1099–1103 (2010).
22. L. Cai, C. K. Dalal, M. B. Elowitz, Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature*, **455**, 485–490 (2008).
23. S. Tay *et al.*, Single-cell nf-κb dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–271 (2010).
24. E. Batchelor, L. Alexander, C. Mock, G. Lahav, Stimulus-dependent dynamics of p53 in single cells. *Mol. Syst. Biol.* **7**, 488 (2011).
25. N. Hao, E. K. O'Shea, Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat. Struct. Mol. Biol.* **19**, 31–39 (2012).
26. J. C. W. Locke, J. W. Young, M. Fontes, M. J. Hernández Jiménez, M. B. Elowitz, Stochastic pulse regulation in bacterial stress response. *Science* **334**, 366–369 (2011).
27. T. M. Norman, N. D. Lord, J. Paulsson, R. Losick, Stochastic switching of cell fate in microbes. *Annu. Rev. Microbiol.* **69**, 381–403 (2015).
28. J. H. Levine, Y. Lin, M. B. Elowitz, Functional roles of pulsing in genetic circuits. *Science* **342**, 1193–1200 (2013).
29. R. Martinez-Corral, E. Raimundez, Y. Lin, M. Elowitz, J. Garcia-Ojalvo, Self-amplifying pulsatile protein dynamics without positive feedback. *Cell Syst.* **7**, 453–462.e1 (2018).
30. N. D. Lord *et al.*, Stochastic antagonism between two proteins governs a bacterial cell fate switch. *Science* **366**, 116–120 (2019).
31. T. Wada, Y. Hatamoto, K. Kutsukake, Functional and expressional analyses of the anti-$FlhD_4C_4$ factor gene ydiV in *Escherichia coli*. *Microbiology* **158**, 1533–1542 (2012).
32. A. Takaya *et al.*, YdiV: A dual function protein that targets FlhDC for ClpXP-dependent degradation by promoting release of DNA-bound FlhDC complex. *Mol. Microbiol.* **83**, 1268–1284 (2012).
33. T. Wada, Y. Tanabe, K. Kutsukake, FliZ acts as a repressor of the ydiV gene, which encodes an anti-$FlhD_4C_2$ factor of the flagellar regulon in *Salmonella enterica* serovar typhimurium. *J. Bacteriol.* **193**, 5191–5198 (2011).
34. R. Hengge *et al.*, Systematic nomenclature for GGDEF and EAL domain-containing cyclic di-GMP turnover proteins of *Escherichia coli*. *J. Bacteriol.* **198**, 7–11 (2016).
35. Y. Sato, A. Takaya, C. Mouslim, K. T. Hughes, T. Yamamoto, FliT selectively enhances proteolysis of FlhC subunit in $FlhD_4C_2$ complex by an ATP-dependent protease, ClpXP. *J. Biol. Chem.* **289**, 33001–33011 (2014).
36. J. H. Davis, A. J. Rubin, R. T. Sauer, Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* **39**, 1131–1141 (2010).
37. S. Koirala *et al.*, A nutrient-tunable bistable switch controls motility in *Salmonella enterica* serovar typhimurium. *mBio* **5**, e01611-14 (2014).
38. G. P. Stafford, T. Ogi, C. Hughes, Binding and transcriptional activation of non-flagellar genes by the *Escherichia coli* flagellar master regulator $FlhD_2C_2$. *Microbiology* **151**, 1779–1788 (2005).
39. A. Albanna *et al.*, Driving the expression of the *Salmonella enterica* sv Typhimurium flagellum using *flhDC* from *Escherichia coli* results in key regulatory and cellular differences. *Sci. Rep.* **8**, 16705 (2018).
40. E. Korobkova, T. Emonet, J. M. G. Vilar, T. S. Shimizu, P. Cluzel, From molecular noise to behavioral variability in a single bacterium. *Nature* **428**, 574–578 (2004)
41. Y. Tu, G. Grinstein, How white noise generates power-law switching in bacterial flagellar motors. *Phys. Rev. Lett.* **94**, 208101 (2005).
42. L. Susman *et al.*, Individuality and slow dynamics in bacterial growth homeostasis. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5679–E5687 (2018).