A filter-and-integrate mechanism generates stochastic transcriptional pulses in *E.coli* flagellar synthesis

Alberto Stefano Sassi,[1] Mayra Garcia-Alcala,[2,3] Philippe Cluzel,[2] and Yuhai Tu[1]

[1]*IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, U.S.A*

[2]*Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA*

[3]*Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Colonia Chamilpa, Cuernavaca 62210, Morelos, Mexico*

Bacterial flagellar synthesis is a complex process that requires coordinated expressions of three classes of genes. Recent single cell experiments in *E. coli* showed that flagellar class-2 promoters are activated in stochastic pulses instead of following a deterministic program as suggested by previous population measurements. Here, by using a stochastic model to quantitatively analyze these single-cell expression data, we identify two main factors that are responsible for the stochastic pulsating expression of class-2 genes. First, YdiV, an inhibitor of the master regulator FlhDC (a class-1 protein complex), creates a steep response function that filters out the small FlhDC fluctuations, and the filtered signal approximates a digital random telegraph signal. Second, the timescale (τ_2) governing the class-2 activity dynamics is longer than that of FlhDC fluctuations (τ_1), which indicates an integration effect wherein class-2 activity depends on previous FlhDC levels within a time-window τ_2. Together, our results suggest that the noisy FlhDC input signal is first filtered with a threshold set by YdiV and then integrated over the long timescale τ_2 to yield the intermittent transcriptional pulses observed in *E. coli* flagellar synthesis. This filter-and-integrate strategy allows *E. coli* to avoid prematurely starting the expensive class-2 gene expression by small or large-but-short FlhDC fluctuations. Possible molecular mechanisms for the filter-and-integrate strategy are proposed. Our model is also used to make testable predictions on responses to realistic time-dependent stimuli. Finally, a comparison with a different gene regulation strategy enabled by an additional positive feedback loop in *Salmonella* is discussed.

I. INTRODUCTION

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 TF 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 chemo-sensory 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 *E. 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 turned on sequentially [10].

Recently, some of us (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, i.e., bursts of strong activity amid a quiet (near zero) background, rather than following a deterministic time-ordered program. Simliar stochastic pulsating gene expression dynamics have been observed in a wide range of organisms, e.g., transcription factor Msn2 in yeast [22], transcription factor p53 [23] and NF-κB [24] in mammalian cells, bacterial transcription factor σ_b in the bacterium Bacillus [25]. However, despite the ubiquity of the pulsatile expression pattern, the underlying molecular principles that govern the genesis and dynamics of pulses remain not fully understood. Here, using stochastic modeling constrained by the data from Kim et al [19], we wish to provide a

theoretical framework with a minimal number of parameters to explain the origin of the observed stochastic dynamics of flagellar synthesis in *E. coli*.

II. RESULTS

We first present a 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. A stochastic model for class-1 and class-2 gene expression dynamics

In the recent single-cell experiments [19], Kim et al used strains whose expression of class-1 genes was controlled by promoters of different strengths (see Sec. A in Supplemental Information (SI) for details of the experimental data). For each strain, long time series (~ 70 generations) of class-1 and class-2 promoter activity was 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 YdiV [26–28] (also known as RflP [29]), which inhibits the interaction between the master regulator and the class-2 promoter. To be consistent with experiments, we define the class-2 gene expression activity, A_2, as the time derivative of the fluorescence time series associated with the class-2 reporter gene in single cells, and use the class-1 reporter florescence signal as a proxy to represent the FlhDC level C_1.

To understand these single-cell expression data, we develop a minimal model that captures the essential ingredients of the system. Specifically, to describe the single cell dynamics of the class-2 activity in single cells, we propose a stochastic differential equation (SDE) based on the three most important features of the single cell gene expression process: activation (production), decay (degradation), and noise:

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

where θ_s is the single cell noise term with $s = (+, -)$ for cells with and without YdiV respectively. Without going into molecular details of activity decay (e.g., de-activation of the promoters due to unbinding of FlhDC or the inhibition of the transcription by other regulators, etc.), we define τ_2^{-1} as the decay rate with τ_2 the characteristic timescale associated with the class-2 expression. For convenience, we write activation rate of the class-2 promoter as $f_s(C_1)/\tau_2$ so that the average $\langle A_2 \rangle = \langle f_s(C_1) \rangle$. 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 $\Delta_s(\langle C_1 \rangle)$ depend 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:

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

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 $\langle C_1 \rangle = \mu$ 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 a constant for all promoter strengths. A multiplicative noise is used, i.e., the noise η is multiplied by C_1 in Eq. 2. This choice of noise comes directly from class-1 reporter fluorescence intensity measurement data, which can be better fitted by having a multiplicative noise than an additive noise in Eq. 2 (see Sec. B in SI for details).

Our full model consists of two SDE's (Eqs. 1-2) 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 $f_s(C_1)$. By fitting all 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 mechanisms.

B. YdiV controls the sensitivity of class-2 gene expression response

We first use our model to study all the steady state class-2 activity distributions from experiments (see Methods for details on simulations of our model). In Fig. 1A, experimen-

tally 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 Sec. A in SI 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 the left panel of Fig. 1A) is strongly asymmetric with a peak that is significantly shifted with respect to the mean, whereas the distributions for lower and higher promoter strength are more symmetric. This asymmetry is largely absent in strains without YdiV (the right panel of Fig. 1A). Our model can fit all the steady state A_2 distributions from different strains together as shown by the direct comparison between experimentally measured distributions (Fig. 1A) and those obtained from our model (Fig. 1B).

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. 1C, 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.

One of the main results obtained by fitting our model to the single-cell data (see Sec. C in SI for details of the fitting and values of all the fitted model parameters) is the determination of the response functions in the presence or absence of YdiV respectively, $f_+(C_1)$ and $f_-(C_1)$, which are shown in Fig. 1D. These response functions are difficult to determine directly from experiments due to their nonlinearity and the large fluctuations in C_1. One can appreciate this difficulty by noting in Fig. 1D that the response functions do not go through the centers of the data distributions because $\langle A_2\rangle = \langle f_s(C_1)\rangle \neq f_s(\langle C_1\rangle)$ due to their nonlinearity.

The difference between the two response functions ($f_+(C_1)$ and $f_-(C_1)$) informs us about the underlying gene regulation mechanism mediated by YdiV. In particular, we find that in the presence of YdiV the response function has a small constant value up to a threshold where it rises sharply (solid line in Fig. 1D). At a medium class-1 promoter strength, when the mean FlhDC concentration is close to the threshold, 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 (left panels in Fig. 1A&B) due to the large steepness of the gene expression response function f_+ near

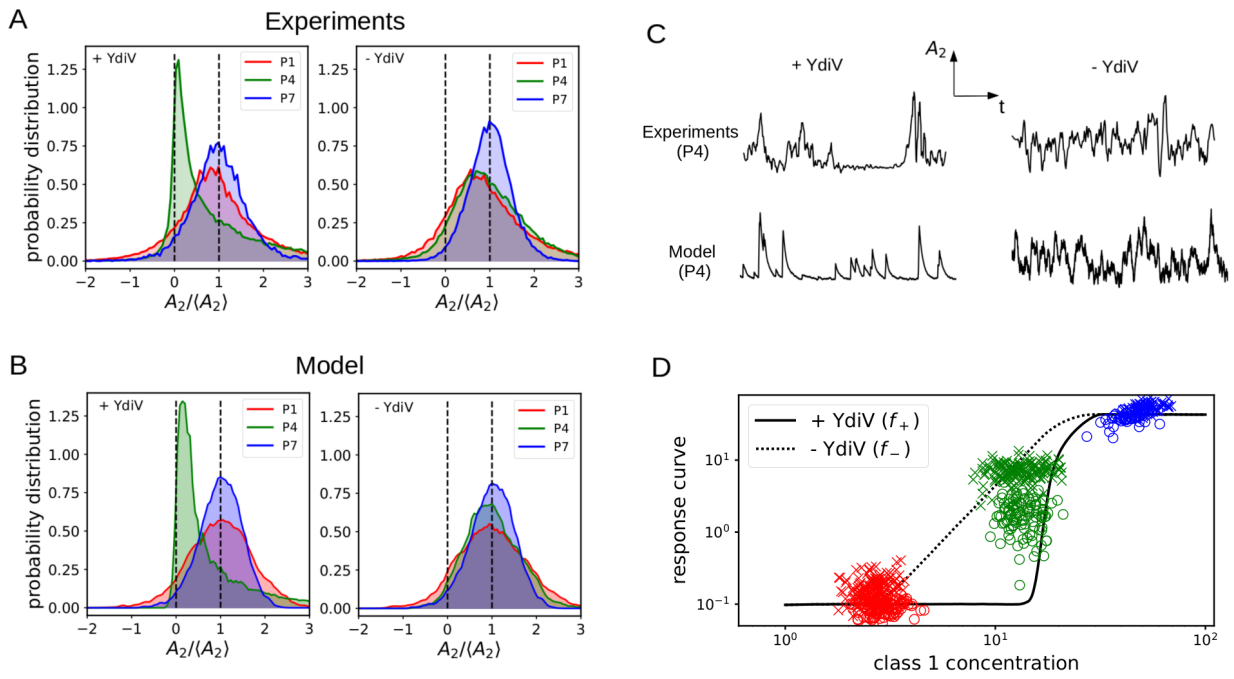

FIG. 1: Comparison between experimental data and model results. (A) Distributions of the normalized class-2 activity $\frac{A_2}{\langle A_2 \rangle}$ in the presence (left) and absence (right) of YdiV for small (P1, red), medium (P4, green) and large (P7, blue) class-1 promoter strengths. Data taken from [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 YdiV (f_-, dashed line). The circles and crosses indicate the mean value for different mother cells, in the presence and the absence of YdiV respectively. The color code for each promoter is the same as in A&B.

the threshold. In the absence of YdiV, the response function f_- is found to be less steep (dotted line in Fig. 1D) and the probability distributions are more symmetric (right panels in Fig. 1A&B). The agreements between our model and the experiments not only valid our model, but more importantly, they reveal that the presence of YdiV leads to a delayed but much steeper or more sensitive response of the class-2 gene expression to the class-1 protein (FlhDC) concentration.

C. Separation of timescales and the memory effect

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:

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

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 $t-\tau_2 \leq t' \leq t$, i.e., 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 $\langle A_2(t)\rangle \approx f_s(C_1(t))$, i.e., the class-2 activity depends on the instantaneous level of class-1 protein $C_1(t)$ with negligible memory effect. However, if $\tau_2/\tau_1 \gg 1$, the memory effect is dominant as $\langle A_2(t)\rangle \approx \langle f_s(C_1)\rangle_{C_1} \equiv \int f_s(C_1)P(C_1)dC_1$ becomes a constant independent of the instantaneous FlhDC level $C_1(t)$ ($P(C_1)$ is the steady state C_1 distribution function).

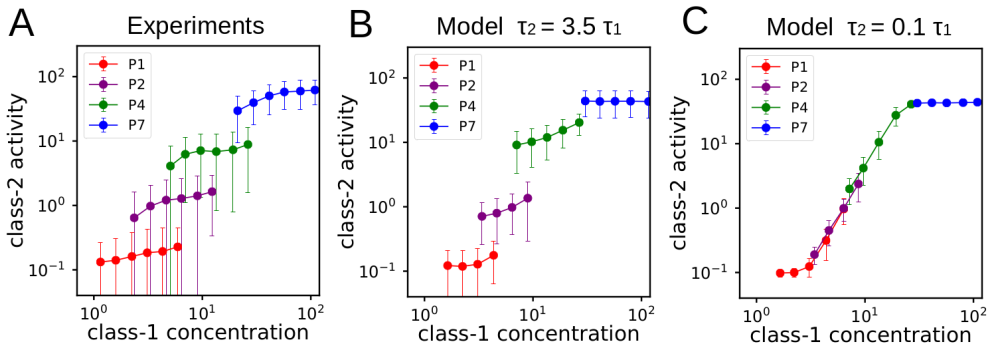

FIG. 2: The bin-averaged response curves and the ratio of two timescales τ_2/τ_1. (A) The bin-averaged response curves with different class-1 promoter strength from experimental data [19] for strains without YdiV (see SI for similar results for strains with YdiV). Different colors are used for different class-1 promoters. The corresponding model results for (B) $\tau_2 = 3.5\tau_1$ and (C) $\tau_2 = 0.1\tau_1$.

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. 2A, at the same (bin-averaged) FlhDC concentration the (bin-averaged) class-2 activities can be different for different strains, i.e, the BARC's for different strains can be shifted from each other. 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, simulations of our model confirm that BARC's for different strains are separated when $\tau_2 > \tau_1$ as shown in Fig. 2B, whereas when $\tau_2 \ll \tau_1$, they collapse onto the same curve as shown in Fig. 2C. Quantitatively, by comparing BARC's 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 Sec. D in SI for details).

D. A YdiV-mediated sequestration mechanism for filtering

So far, we have used our model to explain and understand the steady state distributions and dynamics of the noisy single-cell data. In this section, we propose possible molecular mechanisms underlying results from previous sections with a focus on the role of YdiV in regulating FlhDC signal.

We first breifly 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 is related to the decay time of the FlhDC-dependent class-2 promoter activation (see Fig. 3A for an illustration). Since FlhD&FlhC can only serve as functional transcription factor 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 [30], 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 the Discussion section for more details).

We now study the role of YdiV in creating a delayed but steeper switch-like response of class-2 gene expression. 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 [27]. Second, YdiV mediates the interaction between FlhDC and the degradation complex ClpXP [26, 27]. Here, we only consider the effect of competitive binding on the

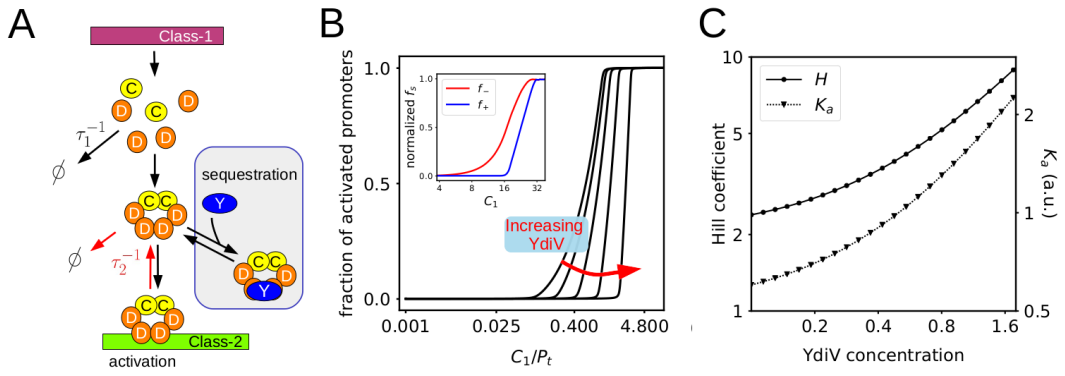

FIG. 3: The YdiV-mediated sequestration mechanism for switch like behavior. (A) Schematic representation of the mechanism of transcriptional regulation of class-2 promoters by FlhDC. The class-1 promoter (purple) initiates the expression of the monomers FlhC and FlhD (yellow and orange respectively) 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. 1D). (C) Hill coefficient (H) and the half maximum concentration (K_a) as a function of the YdiV concentration.

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 test this YdiV-mediated sequestration mechanism, we developed a model based on competitive binding of FlhDC by YdiV and the class-2 promoter as illustrated in Fig. 3A (see Methods for details of the sequestration model). In Fig. 3B, 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 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 f_+ and f_- shown in the inset of Fig. 3B for comparison. A response curve can be quantitatively described by its Hill coefficient H and its half maximum concentration K_a. From the experimental data, the Hill coefficient changes roughly from $H \sim 2$ in the $\Delta ydiV$ mutant to $H \sim 4$ in the wild type, and K_a almost doubles. As shown in Fig. 3C, the YdiV-mediated sequestration model reproduces the general trend with both H and K_a increasing with YdiV concentration in a range consistent with experiments.

Overall, the steep response function $f_+(C_1)$ suggests that YdiV serves (approximately) as a digital filter. When the input signal, i.e., FlhDC concentration (C_1) is lower than a threshold $C^* \equiv K_a(Y_t)$ 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. The (normalized) filtered signal switches between 0 and 1 randomly resembling a random telegraph signal.

E. The filter-and-integrate strategy to control noisy signal

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, i.e., the FlhDC concentration $C_1(t)$ (blue line), which fluctuates relatively fast at a short timescale τ_1, first goes through a highly nonlinear 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 akin to the random telegraph signal. The second part of the two-pronged process is to integrate the filtered signal over a timescale $\tau_2 (> \tau_1)$, which give rise to the pulses (intermittent burst pattern) in $A_2(t)$ dynamics (red line). The integration process acts as a filter in time to filter out those large-but-short (strong in intensity but short in time duration) FlhDC fluctuations. As result, class-2 is turned on only when the averaged filtered signal over the long memory timescale (τ_2) is high. Due to the same memory effect, once turned on, the class-2 expression remains active for a time duration τ_2 even without FlhDC signal. The combination of filtering and integration leads to the intermittent but strong pulses seen in the class-2 activity dynamics (lower left panel in Fig. 4). In the absence of YdiV, as shown

in the right side of Fig. 4, the noisy $C_1(t)$ signal is not filtered and there is no pulse in $A_2(t)$ as a result (lower right panel in Fig. 4).

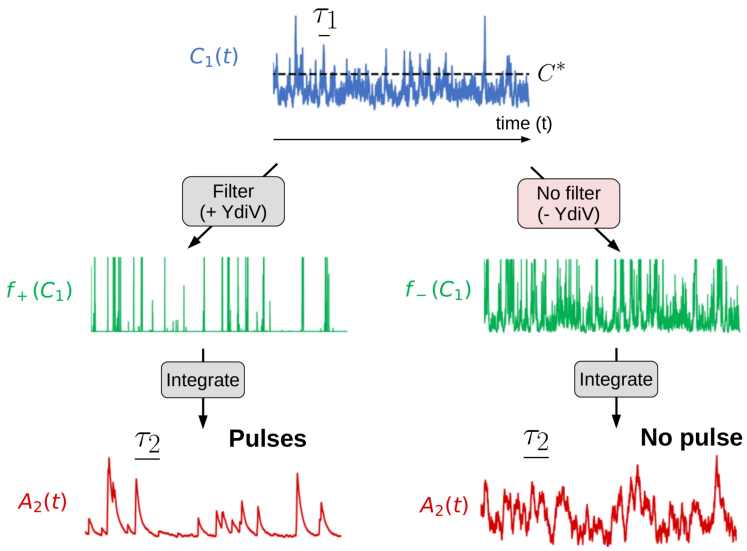

FIG. 4: Signal processing with and without YdiV. In the presence of YdiV, the FlhDC concentration $C_1(t)$ (blue line), which fluctuates with a fast timescale τ_1, is filtered by the inhibitory effect of YdiV. The filtered signal $f_+(C_1(t))$ (green line), which has a pulse-like pattern, is then integrated 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 and no pulse. Results presented here are obtained from simulations of our model using parameters for the P4 strain.

F. Model predictions: 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 $\mu_0(=\mu(P_1)\approx 2.6)$ 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. 5A. 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. 4B, 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. 4C. Responses to other time-varying stimuli such as ramps and oscillatory signals can be studied by our model similarly.

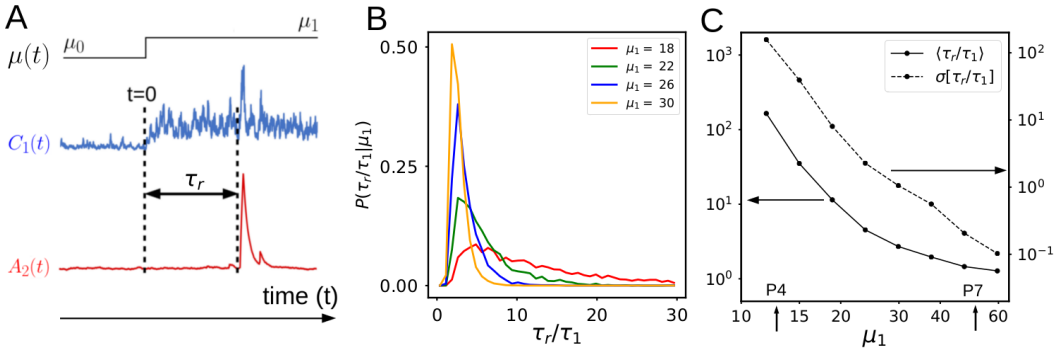

FIG. 5: The predicted responses to step stimuli. (A) Predicted dynamics of C_1 and A_2 in response to a time-dependent (step function) stimulus in which the class-1 promoter strength μ is changed at $t = 0$ from a low value $\mu_0 = \mu(P1) = 2.6$ to a higher value μ_1. The delay time τ_r is the time duration from the stimulus ($t = 0$) to the onset of class-2 expression, which is defined as the time when A_2 cross 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.

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 FlhDC signal. As shown in Fig. 4C, a higher FlhDC concentration induces (on average) a shorter and more accurate (smaller variance) onset time to turn on the class-2 genes. These model predictions can be tested in future single cell experiments by using inducible class-1 promoters.

III. SUMMARY AND DISCUSSIONS

In this paper, we study dynamics of the class-1 and class-2 promoters in *E. coli* flagellar synthesis by using a minimal stochastic model that captures the essential characteristics of the underlying system. Fitting our stochastic model to a large single-cell dataset quantitatively allows us to distill the underlying regulatory mechanism from the noisy data. Our

study suggests that *E. coli* adopts a sophisticated filter-and-integrate strategy to control noise for gene expression. 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.

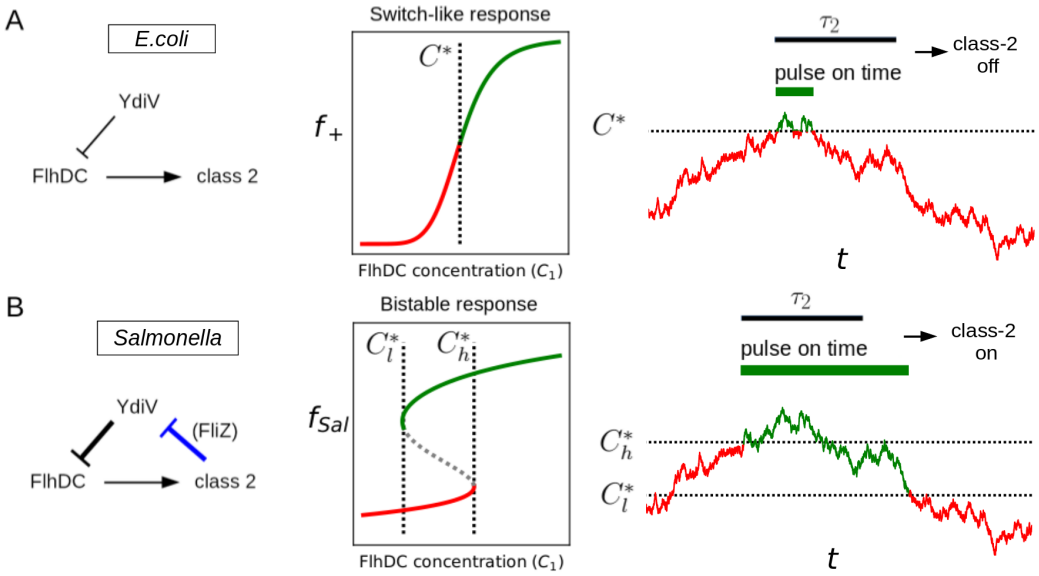

FIG. 6: The different gene regulation strategies for *E. coli* and *Salmonella*. (A) The *E. coli* network has the YdiV-mediated inhibition of the class-1 master regulator (FlhDC) but not the feedback mechanism from class-2 gene. As a result, *E. coli* has a steep but monotonic response function f_+ with a threshold C^*, i.e., class-2 gene expression is activated (green) when $C_1 > C^*$ or inhibited (red) when $C_1 < C^*$. For a 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.

To better understand the filter-and-integrate strategy in *E. coli*, we compare it with a different strategy that seems to be adopted by a closely related bacterium *Salmonella*. In both *E. coli* and *Salmonella*, YdiV inhibits the activation of class-2 promoters through a

direct interaction with FlhDC. However, 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 [31]. 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. 6B. 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 for *E.coli* as shown in Fig. 6A, 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.

The different strategies may reflect the different purposes of motility for these two bacteria. *Salmonella* is a pathogen and it 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 study suggests a few new directions for future investigation. On the experiment side, effects of the filter-and-integration mechanism may be tested by perturbing the two key factors, e.g., by tuning the strength of the interaction between FlhDC and class-2 promoters [27, 32] 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 [33]. 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. On the modeling side, YdiV level is assumed to be constant in our model for simplicity. However, YdiV level fluctuates with time and varies from cell to cell. We need to improve our model by including dynamics of YdiV level and its cell-to-cell variability to study gene expression dynamics in single cells. Furthermore, our modeling framework can be extended to include class-3 genes with the associated feedback regulations to study the full bacterial flagellar gene expression cascade as well as stochastic

dynamics in other gene regulatory systems in single cells [22–25, 34].

IV. METHODS

Simulation of the stochastic differential equations. For the numerical simulation of the stochastic equations, we write Eq. 2 in discrete form:

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

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:

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

where γ_θ is a Gaussian random number as γ_η.

Bin-averaged response curves and flatness. To obtain the bin-averaged response curves, we divide the full interval $10^{-1} \leq C_1 \leq 10^2$ into 20 logarithmically spaced bins and calculate the average of class-1 concentration and class-2 activity for all the data points in each bin. This bin-averaged class-2 activity versus the bin-averaged class-1 concentration is called the bin-averaged response curve (BARC). The BARC in Fig. 2 are only shown at those bins that contain more than 5000 data points for better statistics.

Steady-state solution of the sequestration model. Let P_t be the concentration of promoters, C_1 is the total concentration of FlhDC, Y_t 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: $C_{1p} = \frac{C_{1f}P_t}{C_{1f}+K_p}$, and $C_{1y} = \frac{C_{1f}Y_t}{C_{1f}+K_y}$, 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):

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

which can be solved to determine C_{1f} as a function of P_t, C_1, Y_t, and K_y, 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. 3B. Parameters $K_y/P_t = 10^{-2}$ and $K_p/P_t = 10^{-1}$ are used in Fig. 3B&C.

V. ACKNOWLEDGMENTS

We thank Drs. M. Kim and M. Erhardt for discussions. The work by ASS and YT are supported by a NIH grant (R35GM131734 to YT). The work by MG and PC are supported by a NIH grant (R01GM134275 to PC) and a NSF grant (1615487 to PC).

---

[1] A. Martinez-Antonio, S. Janga, and D. Thieffry. Functional organisation of Escherichia coli transcriptional regulatory network. *J. Mol. Biol.*, 381:238–247, Jun 2008.

[2] J. Bähler. A transcriptional pathway for cell separation in fission yeast. *Cell Cycle*, 4:39–41, Jun 2005.

[3] D. W. Allan and S. Thor. Transcriptional selectors, masters, and combinatorial codes: regulatory principles of neural subtype specification. *Wiley Interdiscip. Rev. Dev. Biol*, 4:318–56, 2015.

[4] I. Spöring, S. Felgner, M. Preuße, D. Eckweiler, M. Rohde, S. Häussler, S. Weiss, and M. Erhardt. Regulation of flagellum biosynthesis in response to cell envelope stress in Salmonella enterica serovar typhimurium. *MBio*, 9:e00736–17, 2018.

[5] O. A. Soutourina and P.N. Bertin. Regulation cascade of flagellar expression in Gram-negative bacteria. *FEMS Microbiology Reviews*, 27:505–523, Oct 2003.

[6] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, pages 318–56, Jun 1961.

[7] Shai S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics*, page 64–68, Jun 2002.

[8] M. Ghafari and A. Mashaghi. On the role of topology in regulating transcriptional cascades. *Phys. Chem. Chem. Phys.*, 19:25168–25179, 2017.

[9] G. S. Chilcott and 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, Dec 2000.

[10] S. Kalir, J. McClure, K. Pabbaraju, C. Southward, M. Ronen, S. Leibler, M. G. Surette, and U. Alon. Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science*, 292(5524):2080–2083, 2001.

[11] D. Apel and 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, and K. Kutsukake. Promoter analysis of the class 2 flagellar operons of Salmonella. *Genes. Genet. Syst.*, 74(4):179–83, Aug 1999.

[13] X. Liu and P. Matsumura. Differential regulation of multiple overlapping promoters in flagellar class II operons in Escherichia coli. *Mol. Microbiol.*, 21(3):613–20, Aug 1996.

[14] D.M. Fitzgerald, R.P. Bonocora, and J.T. Wade. Comprehensive Mapping of the Escherichia coli Flagellar Regulatory Network. *PLoS Genet.*, 10(10):e1004649, Oct 2014.

[15] Y Komeda. Transcriptional control of flagellar genes in escherichia coli k-12. *J. Bacteriol.*, 168(3):1315–1318, 1986.

[16] K Kutsukake, Y Ohya, and T Iino. Transcriptional analysis of the flagellar regulon of salmonella typhimurium. *J. Bacteriol.*, 172(2):741–747, 1990.

[17] J. D. Brown, S. Saini, C. Aldridge, J. Herbert, C. V. Rao, and P. D. Aldridge. The rate of protein secretion dictates the temporal dynamics of flagellar gene expression. *Mol. Microbiol.*, 70(4):924–937, 2008.

[18] S. Saini, J.D. Brown, P.D. Aldridge, and C.V. Rao. FliZ Is a posttranslational activator of $FlhD_4C_2$-dependent flagellar gene expression. *J. Bacteriol.*, 190(14):4979–4988, 2008.

[19] J. Mark Kim, M. Garcia-Alcala, E. Balleza, and P. Cluzel. Stochastic transcriptional pulses orchestrate flagellum biosynthesis in E. coli. *Sci. Adv.*, 6, Feb 2020.

[20] J. R. Moffitt, J. B. Lee, and P. Cluzel. The single-cell chemostat: an agarose-based, microfluidic device for high-throughput, single-cell studies of bacteria and bacterial communities. *Lab on a Chip*, 12(8):1487–1494, Apr 2012.

[21] P. Wang, L. Robert, J. Pelletier, W.L. Dang, F. Taddei, A. Wright, and S. Jun. Robust growth of Escherichia coli. *Curr. Biol.*, 20(12):1099–103, May 2010.

[22] E. Batchelor, L. Alexander, C. Mock, and G. Lahav. Stimulus-dependent dynamics of p53 in single cells. *Mol. Syst. Biol.*, 7(1):488, 2011.

[23] N. Hao and E. K. O'Shea. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nature Structural & Molecular Biology*, 19(1):31–39, 2012.

[24] S. Tay, J. J. Hughey, T. K. Lee, T. Lipniacki, S. R. Quake, and M. W. Covert. Single-cell nf-κb dynamics reveal digital activation and analogue information processing. *Nature*, 466(7303):267–271, 2010.

[25] J. C. W. Locke, J. W. Young, M. Fontes, M.J. Hernández Jiménez, and M.B. Elowitz. Stochastic pulse regulation in bacterial stress response. *Science*, 334(6054):366–369, 2011.

[26] T. Wada, Y. Hatamoto, and K. Kutsukake. Functional and expressional analyses of the anti-$FlhD_4C_4$ factor gene ydiV in Escherichia coli. *Microbiology*, 158:1533–1542, Mar 2012.

[27] A. Takaya, M. Erhardt, K. Karata, K. Winterberg, T. Yamamoto, and K. T. Hughes. YdiV: a dual function protein that targets FlhDC for ClpXP-dependent degradation by promoting release of DNA-bound FlhDC complex. *Mol. Microbiol.*, 83(6):1268–1284, Mar 2012.

[28] T. Wada, Y. Tanabe, and 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(19):5191–5198, Oct 2011.

[29] R. Hengge, M. Y. Galperin, J. Ghigo, M. Gomelsky, J. Green, K. T. Hughes, U. Jenal, and P. Landini. Systematic nomenclature for ggdef and eal domain-containing cyclic di-gmp turnover proteins of escherichia coli. *J. Bacteriol.*, 198(1):7–11, 2016.

[30] Y. Sato, A. Takaya, C. Mouslim, K.T. Hughes, and 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, Nov 2014.

[31] S. Koirala, P. Mears, M. Sim, I. Golding, Y.R. Chemla, P.D. Aldridge, and C.V. Rao. A nutrient-tunable bistable switch controls motility in salmonella enterica serovar typhimurium. *mBio*, 5(5), 2014.

[32] G.P. Stafford, T Ogi, and C Hughes. Binding and transcriptional activation of non-flagellar genes by the Escherichia coli flagellar master regulator $FlhD_2C_2$. *Microbiology*, 151:1779–88, Nov 2005.

[33] A Albanna, M Sim, P.A. Hoskisson, C. Gillespie, C.V. Rao, and P.D. Aldridge. 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, Nov 2018.

[34] L. Susman, M. Kohram, H. Vashistha, J.T. Nechleba, H. Salman, and N. Brenner. Individuality and slow dynamics in bacterial growth homeostasis. *Proc. Natl. Acad. Sci. U.S.A.*, 115(25):E5679–E5687, 2018.

Supplementary Information for "A filter-and-integrate mechanism generates stochastic transcription pulses in *E.coli* flagellar synthesis" by Sassi et al

Supplementary Information A: The single-cell experimental dataset

All the experimental data analyzed in this paper are from the published single cell experiments by Kim et al [19]. Briefly, to vary the class-1 expression strength, different constitutive artificial promoters are used to replace the native class-1 promoter (flhDp) and either T7 or mut4 sequence are used as the downstream ribosome binding site (RBS-1). The ratio of the effective number of class-1 proteins and fluorescent proteins produced is different for the mut4 and T7 strains. To account for this difference, we multiply the C_1 values by a factor (~ 0.22) for the strains with mut4 as RBS-1 (see SI in [19] for details).

The experimental dataset consists of ~ 100 independent single-cell time series for each class-1 promoter strain, each lasting for ~ 70 generations. The time interval between two consecutive time frames is 5 minutes. Given that the average division time is ~ 40 minutes, each time series has $\sim 70 \times 40/5 = 560$ time points on average, and there are on average $\sim 100 \times 560 = 56,000$ data points for each strain. In this paper, we consider four of the seven strains studied in [19]: P1, P2, P4 and P7 (we use the same strain names as in [19]). These four strains are chosen to cover the full range ($\sim$ 20-fold) of class-1 promoter strength change in [19] .

Supplementary Information B: Distribution of class-1 concentration and multiplicative noise

In order to model the class-1 protein concentration we used a multiplicative noise in Eq. 2 in the main text rather than an additive noise. This choice is motivated by the form of the probability distribution. We can consider for example the class-1 distribution when the promoter strength is P4 (green data points in Fig. S1). In case of additive noise, the concentration follows an Ornstein-Uhlenbeck process (OU) whose steady state distribution is Gaussian. In Fig. S1, we plot a Gaussian distribution (black curve) with the same mean and variance as the experimental data as well as the distribution (red curve) obtained with a multiplicative noise. Fig. S1 clearly shows that the fluctuations in C_1 are better described

by a stochastic model (Eq. 2 in the main text) with a multiplicative noise.

Supplementary Information C: Model parameters and response functions determined by fitting model to the single-cell experimental dataset

In this paper, we develop a simple stochastic model with a minimal set of biologically meaningful parameters and response functions, which can be determined by fitting our model to the large experimental data set described in Sec. A.

For the C_1 dynamics described in Eq. 2 in the main text, there are three parameters: the timescale τ_1, the multiplicative noise strength Δ_1, and the promoter strength μ, which is different for different strains. In our study, τ_1 can be set to unity to set the timescale of the problem, all other timescales are scaled by τ_1. By fitting our model to the experimental data, we found that the noise strength $\Delta_1 \approx 0.1\tau_1^{-1}$, and the promoter strength for different strains are: $\mu(P1) = 2.6$, $\mu(P2) = 6.06$, $\mu(P4) = 13.5$, $\mu(P7) = 51.7$, in the unit of florescence intensity.

The dynamics of class-2 activity A_2 is determined by a timescale τ_2, the response function $f_s(C_1)$, and a noise strength Δ_s that is different for different strains. The parameters and response functions are obtained by fitting our model to the experimental data (class-2 activity dynamics). The range of τ_2/τ_1 is determined as shown in Fig. 2 in the main text. The noise strength Δ_+ and Δ_- for different strains in the presence and absence of YdiV are also determined and their values are shown in Fig. S2.

For the response function f_s (with $s = +, -$), we parameterize it with a piece wise power law function that approximate a sigmoid function with a low and a high plateau. In particular, we set a few anchor points (x_i, y_i) $(i = 1, 2, .., N)$ for f_s such that $f_s(x_i) = y_i$. Between the anchor points $x_i \leq x \leq x_{i+1}$, $f_s(x) = a_i x^{b_i}$ follows a power law, and we have $f_s(x \leq x_1) = y_1$ and $f_s(x \geq x_N) = y_N$ to enforce the two plateaus. We determine the coordinates of the anchor points by fitting our model (with the piece wise power law response function) to the data. For the response function f_+ in the presence of YdiV, we used $N = 4$ anchor points with their fitted coordinates: $f_+(C_1 \leq 17) = 0.1$, $f_+(19) = 0.2$, $f_+(22) = 10$, $f_+(C_1 \geq 24) = 43$. For the response function f_+ in the absence of YdiV, we used only $N = 2$ anchor points with their fitted coordinates: $f_-(C_1 \leq 3) = 0.1$ and $f_-(C_1 \geq 22) = 43$.

In Fig. S3, we show the average C_1 versus average A_2 for a single cell over many generations from experiments and model. The response functions (smoothed around the anchor points) are also shown (the same as in Fig. 1D in the main text).

Supplementary Information D: Estimating the timescale ratio

The separation of timescales seems to be independent of the presence or absence of YdiV. In Fig. S4, we show the BARC's from experimental data for strains with YdiV and those from our model, which are similar to the results for strains without YdiV (Fig. 2 in the main text). Quantitatively, BARCs can be used to estimate the timescale ratio τ_2/τ_1. To quantify the shape of BARC, we define a flatness parameter as the absolute value of the inverse of the average slope of linear fits to BARC's of different strains such as those shown in Fig. 2 in the main text and Fig. S4 here. A larger flatness means that the class-2 activity has a stronger dependence on the average FlhDC level than its instantaneous value and thus the memory effect is stronger. We computed the flatness for different values of the relative timescale ratio τ_2/τ_1 in our model. As shown in Fig. S5, a larger τ_2/τ_1 ratio leads to a larger flatness, i.e., a stronger memory effect. Based on a quantitative comparison with the flatness values obtained from experimental data with and without YdiV, we estimate the ratio τ_2/τ_1 to be in the interval, $2.1 \leq \tau_2/\tau_1 \leq 3.6$, shown as the gray region in Fig. S5. In addition, in Fig. S6, we compare the autocorrelation function for wild type cells shown in [19] with the autocorrelation functions obtained from the simulations of our model (Eq. 1 and 1). The difference in the timescales between class-1 and class-2 is in line with the interval estimated by the analysis of the flatness shown in Fig. S5.

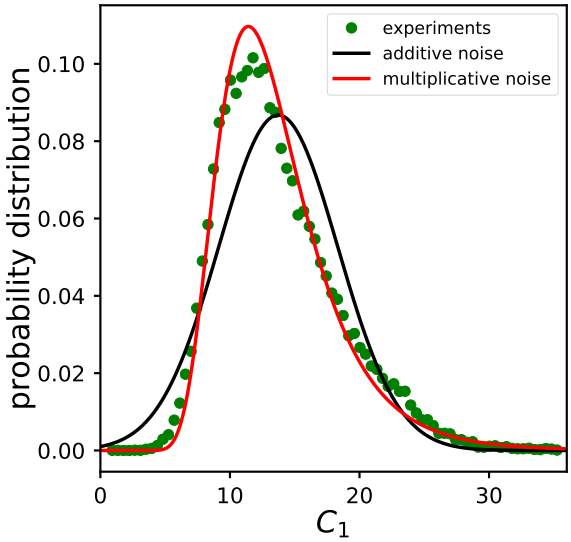

FIG. S1: Comparison of the model with additive and multiplicative noise to fit the experiments. In the case of multiplicative noise the values are distributed following an inverse Gamma function (red). This curve fits the experimental distribution (green) more accurately than the Gaussian distribution (black), corresponding to a model with additive noise.

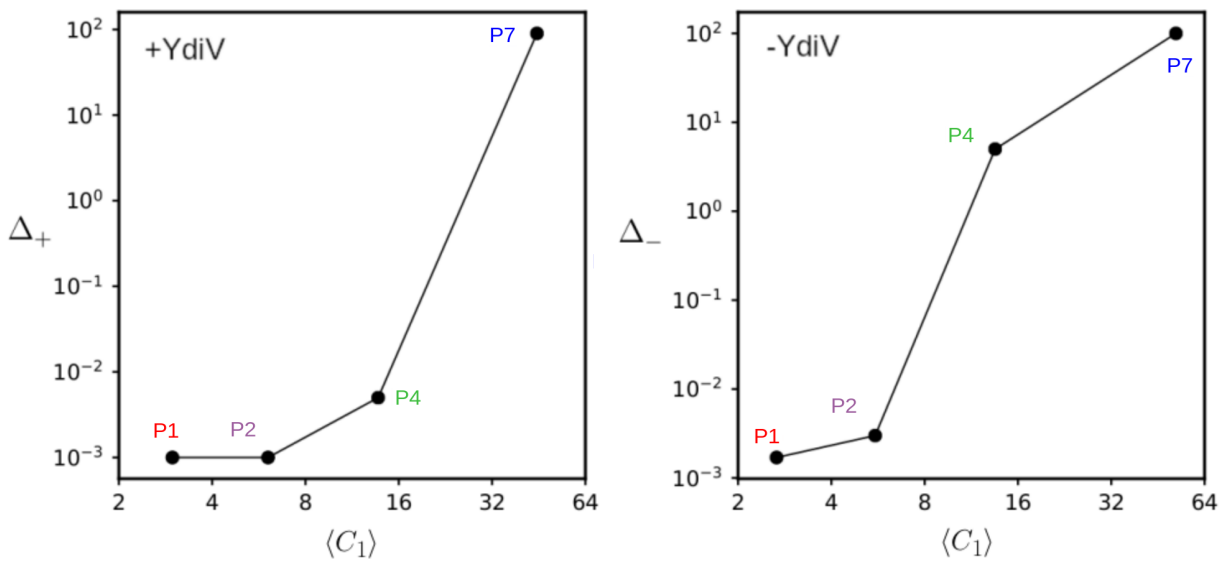

FIG. S2: Noise strength $\Delta_{+,-}$ (in the presence and absence of YdiV) as a function of the average class-1 protein level $\langle C_1 \rangle$ obtained by fitting the stochastic model (Eq. (1-2) in the main text) to experimental data.

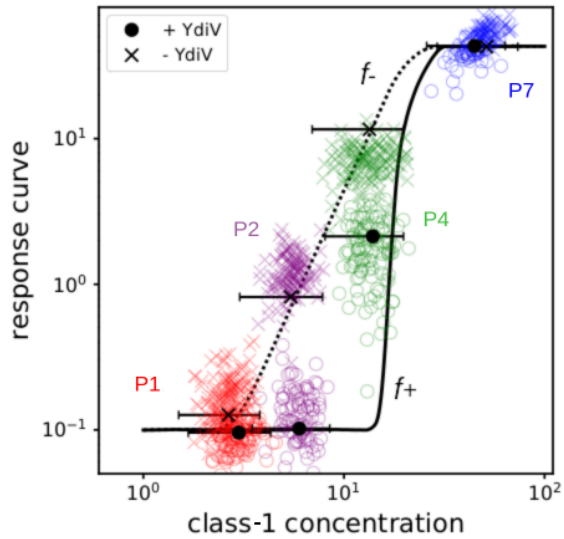

FIG. S3: Response functions $f_{+,-}(C_1)$ used in the model, in the presence (solid) and absence (dashed) of YdiV. The black circles and crosses indicate the mean value obtained with the stochastic model. The horizontal error bars (for C_1) are given by $\sigma = \langle C_1 \rangle \sqrt{2\Delta_1}$, where Δ_1 is the strength of the noise η for class-1 protein level.

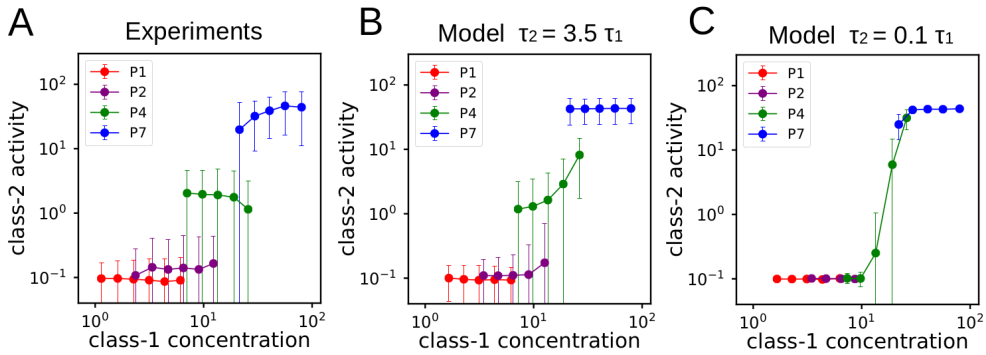

FIG. S4: The bin-averaged response curves and the ratio of two timescales τ_2/τ_1. (A) The bin-averaged response curves with different class-1 promoter strength from experimental data [19] for strains with YdiV. Different colors are used for different class-1 promoters. The corresponding model results for (B) $\tau_2 = 3.5\tau_1$ and (C) $\tau_2 = 0.1\tau_1$.

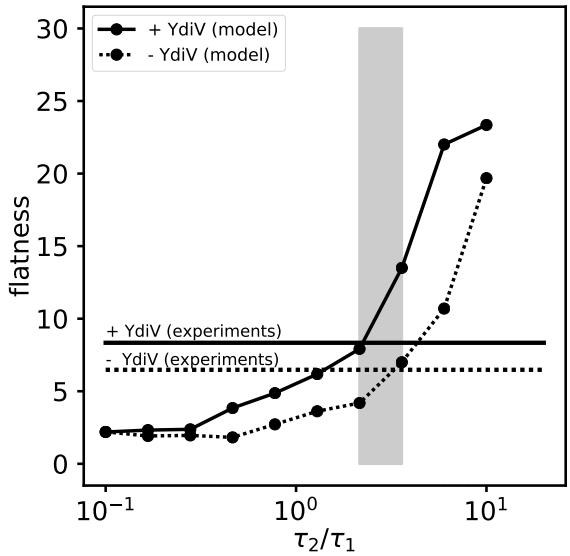

FIG. S5: The estimated range of the timescale ratio τ_2/τ_1. Flatness of the bin-averaged response curves as a function of τ_2/τ_1 in the absence (dashed line) and presence (solid line) of YdiV. The horizontal lines indicate the experimental results. Comparing the model results with the experiments, we estimate the ratio τ_2/τ_1 to be within the gray region: $2.1 \leq \tau_2/\tau_1 \leq 3.6$.

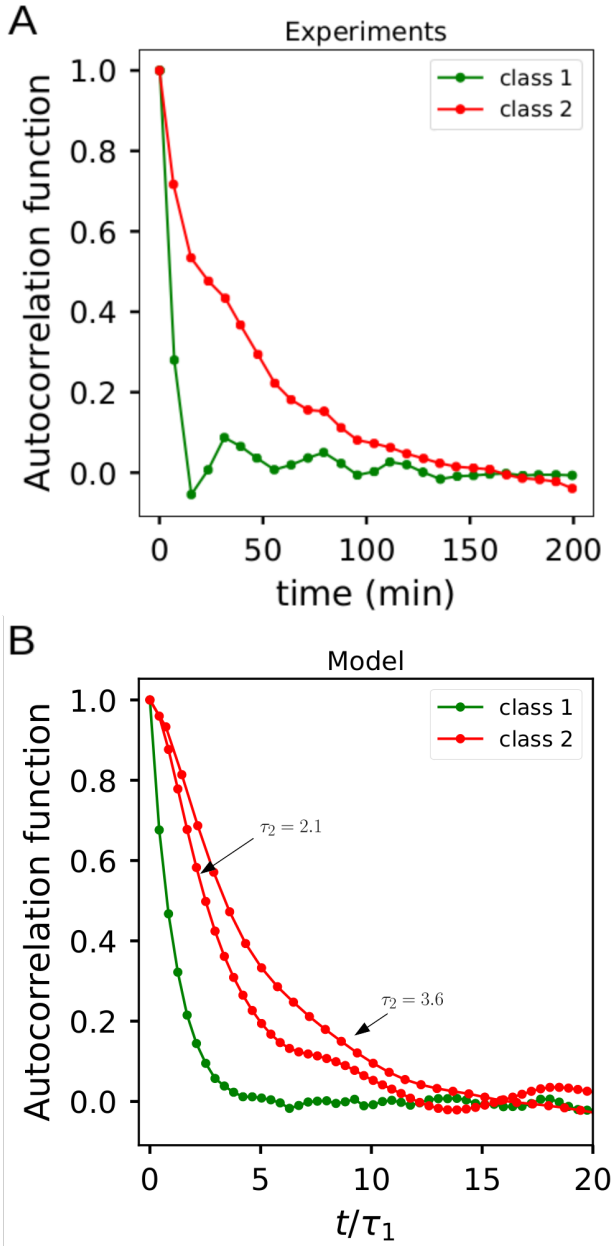

FIG. S6: (A) Autocorrelation functions for the fluorescence of class-1 and class-2 reporters in the case of wild type promoters (taken from Fig. 2D in [19]). (B) Autocorrelation functions for C_1 and A_2 from the stochastic simulations of our model (Eq.1 and Eq.2) for P4 strain.