Plasmid-borne prokaryotic gene expression: Sources of variability and quantitative system characterization

Sangram Bagh, Mostafizur Mazumder, Tharsan Velauthapillai, Vandit Sardana, Guang Qiang Dong, Ashok B. Movva, Len H. Lim, and David R. McMillen*

Institute for Optical Sciences and Department of Chemical and Physical Sciences, University of Toronto Mississauga, 3359 Mississauga Road North, Mississauga, Ontario L5L 1C6, Canada

(Received 8 October 2007; published 28 February 2008)

One aim of synthetic biology is to exert systematic control over cellular behavior, either for medical purposes or to "program" microorganisms. An engineering approach to the design of biological controllers demands a quantitative understanding of the dynamics of both the system to be controlled and the controllers themselves. Here we focus on a widely used method of exerting control in bacterial cells: plasmid vectors bearing gene-promoter pairs. We study two variants of the simplest such element, an unregulated promoter constitutively expressing its gene, against the varying genomic background of four *Escherichia coli* cell strains. Absolute protein numbers and rates of expression vary with both cell strain and plasmid type, as does the variability of expression across the population. Total variability is most strongly coupled to the cell division process, and after cell size is scaled away, plasmid copy number regulation emerges as a significant effect. We present simple models that capture the main features of the system behavior. Our results confirm that complex interactions between plasmids and their hosts can have significant effects on both expression and variability, even in deliberately simplified systems.

DOI: 10.1103/PhysRevE.77.021919 PACS number(s): 87.18.Tt, 87.18.Vf, 87.17.Aa, 87.17.Ee

I. INTRODUCTION

Synthetic biology has as one of its aims the development of a discipline of biological engineering, allowing systematic control to be exerted over living cells [1–4]. This ability would have profound implications, allowing medical interventions to be carried out at the cellular level. Cellular engineers dream of applications such as in-cell cancer fighters that detect oncogenesis and shunt the cell into apoptosis before the disease can start.

The fundamental requirement for the rational design of control systems is an understanding of the dynamics of the system to be controlled (the "plant"), the dynamics of the control mechanism (the "controller"), and the nature of the coupling between the two. Control system design aims to construct controllers that alter the dynamics of the plant in desired ways. Inside cells, the system to be controlled is a complex biochemical environment whose dynamics arise from the kinetics of reactions between species (e.g., DNA, mRNA, proteins). To exert control in this context requires a controller able to receive inputs in that biochemical language, and produce suitable outputs.

Extensive work on building synthetic controllers inside cells has already been done: oscillators [5], bistable toggle switches [6], feedback networks [7,8], logical operators [4,9,10], and cell-to-cell communication systems for pattern formation [11] and intercellular coordination [12,13] have all been implemented, among others [1–3,14–17]. Working in single-celled organisms has allowed researchers to explore a simplified environment while developing the background understanding necessary for future biological engineering projects. Unicellular organisms are also of interest in their own right, as a platform for the creation of robust, self-replicating, microscopic "machines" programmed to carry out human-designed tasks [1–4].

Most of the above-mentioned controllers have been implemented using plasmids (DNA loops that replicate in parallel with the host geneome) as the mechanism for inserting designed regulatory networks into cells, and the behavior of gene expression systems carried on such plasmids is the focus of our study. Understanding plant-controller dynamics in cells has largely taken the form of formulating mathematical models based on biochemical reaction schemes. Issues to be addressed in such models include the following.

(1) The rates and rate constants associated with the reactions representing biological processes such as transcription, translation, and regulatory binding.

(2) The source and nature of cell-to-cell variability in gene expression. Biochemical reactions may take place in low-rate regimes where the fluctuations arising from the discrete nature of molecular interactions are not averaged away; cells may also vary in their individual biochemical composition, causing cell-to-cell differences in gene expression.

(3) The effect of the plant (the host's metabolism and genomic content) on the behavior of the controller [the gene(s) being expressed from a plasmid].

(4) The ability of the controller to generate consistent outputs.

Our study addresses these issues using a deliberately simplified system, the simplest we could construct: an unregulated promoter ($P_{LtetO-1}$) directly expressing a reporter protein [enhanced green fluorescent protein (EGFP)] [18] from a multiple-copy plasmid vector. Although the longer-term aim is to set the stage for feedback and regulatory control, here we address this simplified system and examine its behavior in detail. We consider the following.

(1) Two different types of plasmid vector. Our medium-copy plasmid incorporates feedback control of its own repli-

*mcmillen@utm.utoronto.ca

cation, limiting its copy number and narrowing the copy number variability. The high-copy plasmid lacks this negative feedback, increasing the average copy number but also yielding results consistent with an increased intercell variability in plasmid number.

(2) Four strains of *Escherichia coli*, each representing a variant on the consensus genome; we insert the two plasmid types into each strain.

In each of these eight cases, we examine exponentially growing, asynchronous populations of *E. coli* cells. We use fluorimetry, calibrated against solutions of known EGFP concentration, to determine the average number of proteins per cell, and the average rate of protein production. Quantification of the absolute protein numbers present in cells has been surprisingly neglected in the literature, though there are several recent examples [19–22] in which the question of protein number quantification is carefully addressed; we view our work as complementary to the alternative approaches pursued in this earlier work. The average protein numbers and production rates we obtain vary substantially across our four strains, pointing out that descriptions of "the behavior of plasmid *X*" are not sufficient; we must consider "the behavior of plasmid *X* in genotype *Y*." Qualitative changes in the behavior of synthetic systems as a function of cell strain have been noted previously [9]; here, we consider the phenomenon quantitatively and in further detail, though in a simpler situation.

We employ flow cytometry and fluorescence microscopy to measure the distribution of protein expression levels across populations. The results from the two methods agree closely, and show that the fluorescence intensity distributions have very similar shapes, within each cell strain, for the high- and medium-copy plasmid types. Variation in our system may be partitioned into two categories, following the consensus nomenclature [20,23–40]: intrinsic noise represents variability arising from fluctuations in the process of gene expression itself (including, for example, RNA polymerase binding, transcription of messenger RNA, and translation of mRNA into proteins); extrinsic noise represents all other effects causing cell-to-cell variations in gene expression levels (including cell size, plasmid copy number, and the availability of enzymes such as RNA polymerase and ribosomes). Our quantification of mean protein numbers enables us to formulate simple stochastic models (discussed below) which predict that the amount of intrinsic noise present should be minimal, suggesting that extrinsic noise dominates in our system.

Moving on to investigate the nature of this extrinsic noise, we find that the dominant contribution arises from progress of the cell through its cell division cycle; this matches a result previously reported in yeast [40]. Growth and division creates intercell variability, and size effects must be scaled away before other effects on variability may be observed. Scaling intensities by image sizes from microscopy greatly reduces the variability, and reveals a strong pattern: within each cell strain; the size-scaled variability of the high-copy plasmid is substantially greater than that of the medium-copy plasmid. The medium-copy plasmid also shows very consistent size-scaled variability across all strains, while the high-copy plasmid varies much more substantially from strain to strain. These results reinforce the extrinsic noise origin of the observed variability, and also highlight a secondary source of variations: the plasmid copy number differences between cells. Although we do not have a direct measure of single-cell plasmid copy numbers, the results suggest that the medium-copy plasmid, with its negative feedback regulation, has substantially lower intercell variability in copy number, while the unregulated high-copy plasmid varies more significantly from cell to cell.

The study offers a detailed characterization of a system with surprising complexity despite its deliberately simplified construction. The results underscore the importance of selecting an appropriate cell strain and plasmid type when constructing networks in bacteria, offer quantitative values to be used in constructing models of the systems behavior, and confirm that highly consistent behavior can be obtained. We view this as a step toward the detailed understanding of cellular dynamics required to implement the designs envisioned in synthetic biology.

II. RESULTS AND DISCUSSION

A. Plasmid construction and cell strains

We have studied gene expression from an unregulated promoter system, directly expressing the fluorescent protein EGFP under the control of the promoter $P_{LtetO-1}$ [41]. This promoter is repressed by the TetR protein, which is not present in the *E. coli* strains and plasmids employed, and thus the promoter provides access to the simplest case in gene expression: unregulated (constitutive) expression. With this promoter-gene system we constructed two different plasmids, high copy (from the pEGFP vector backbone) and medium copy (from the pBR322 vector backbone); full plasmid details are provided below in Table III and Fig. 7. The medium-copy plasmid incorporates feedback control of its own replication (the PMB origin of replication with the *rop* gene present), while this feedback is disrupted in the high-copy version (the pUC origin of replication, with a mutation in the *rop* gene as discussed further in Sec. II E, below), yielding average copy numbers per cell reported to be in the ranges of 40–60 [42–44] and 400–500 [43,45], respectively.

The two plasmid versions were incorporated into four different cell strains of the bacterium *Escherichia coli*: DH5α, Top10, B/r, and BL21*. Each of these strains incorporates a set of alterations to the consensus wild-type *E. coli* genome, and full details of the genomic changes are listed in Table III. The use of multiple cell strains allows us to test the same plasmid-borne gene expression system against varying genomic backgrounds.

Without access to quantitative polymerase chain reaction (qPCR) data, reliable measurements of the plasmid copy numbers in our specific strain-plasmid combinations have not been achievable. We have obtained rough estimates of average numbers by densitometry: quantifying the brightness of stained plasmid DNA in gels and comparing the intensity to DNA ladder bands with known masses of DNA present (see Sec. IV). Using this technique, we have obtained the following values (with errors of plus or minus one standard deviation): DH5α (high, 600 ± 140; medium, 80 ± 60);

TABLE I. Average protein expression levels and rates. Protein numbers were determined by calibrated fluorimetry, and proteins produced per cell division deduced as described in the text. Errors represent one standard deviation, over a minimum of five samples from independent colonies.

Cell strain	Plasmid copy number	Percent dead	Mean EGFP/cell	Cell division time (min)	Proteins/min
DH5α	High	21.7 ± 2.5	$165\,000 \pm 6000$	36.8 ± 2.6	3100 ± 250
	Medium	18.5 ± 1.2	$22\,500 \pm 2000$	30.5 ± 2.4	510 ± 60
Top10	High	24.0 ± 2.0	$128\,000 \pm 11\,000$	29.5 ± 1.0	3000 ± 280
	Medium	18.1 ± 2.1	$17\,200 \pm 400$	29.8 ± 1.4	400 ± 20
B/r	High	19.5 ± 6.8	$140\,000 \pm 4000$	27.9 ± 1.4	3500 ± 200
	Medium	13.1 ± 0.7	$13\,000 \pm 1100$	33.1 ± 2.0	270 ± 30
BL21*	High	23.1 ± 3.6	$51\,800 \pm 4600$	49.9 ± 20.3	720 ± 300
	Medium	13.3 ± 1.8	6000 ± 200	31.6 ± 1.3	130 ± 10

Top10 (high, 400 ± 60; medium, 50 ± 10); B/r (high, 460 ± 60; medium, 60 ± 10); and BL21* (high, 300 ± 200; medium, too low to quantify). The values match at least roughly with the literature values noted above [42–45].

B. Protein numbers

Fluorescent proteins allow the measurement of a signal whose intensity is proportional to the number of proteins being expressed in an individual cell. By calibration, we may convert that proportionality into an approximation of the actual numbers of proteins in each cell. Such numbers are important in modeling natural genetic regulatory networks [20,23–27,30,32–34,38–40,46–48] and in the design and construction of novel versions of such networks [1–7,11–15,17]. A recent study [20] used fluorescence intensities as a measure of protein abundance, related them to previously published average protein numbers per cell [19], and showed a scaling relationship between the average amount of protein produced per cell and the variability in the yeast *Saccharomyces cerevisiae*. Another recent method deduced protein numbers by observing the partitioning of fluorescent proteins through a cell division event [22,36].

Here, we use a straightforward fluorescence intensity calibration to obtain the number of proteins per cell. This method has been used elsewhere to quantify protein expression in mammalian cells [21], and we view it as complementary to the above-mentioned quantification techniques. We used fluorimetry on EGFP solutions of known concentrations to generate a standard curve of EGFP number versus fluorescence intensity (data not shown). The solution was excited at 488 nm, and the fluorescence emission spectrum collected over the range 500–650 nm, using the intensity at the emission maximum of 507 nm for calculations. The total cell fluorescence of cultures of EGFP-expressing cells was measured under the same conditions; background autofluorescence levels were obtained by measuring the emission from the same cell strain without the EGFP-expressing plasmid, and this background was subtracted. No difference in emission maxima between pure EGFP solution and cellular EGFP was observed. The fluorescence intensity from EGFP-expressing cells was converted into a total equivalent number of EGFP proteins present in the culture, and mean numbers of EGFP per cell were calculated using the concentration of cells in the cellular culture, obtained by optical absorbance measurements at 600 nm. A minimum of five samples from independent colonies was obtained for each cell strain and plasmid type, yielding the results shown in Table I. The high-copy plasmids produced averages on the order of hundreds of thousands protein molecules per cell whereas the medium-copy plasmids produced tens of thousands of protein molecules per cell. The protein numbers we obtain with this technique represent a lower bound on the average number of proteins actually present in the cells: effects such as quenching, absorption of emitted fluorescence by intracellular material, and the inability to see immature or misfolded proteins imply that additional EGFP proteins could be present without contributing to the observed fluorescence intensity.

The four variants on the consensus wild-type *E. coli* genome represented by our four cell strains yield strikingly different levels of protein expression from the same plasmids, varying three-to fourfold across the strains (see Table I and Fig. 1). To convert the experimentally measured average numbers of proteins per cell into a rate of protein expression requires some information about the variation of protein numbers with time. We take cell growth and protein expression to be exponential with time [49], and note that the steady state for each cell is one in which the number of proteins doubles (on average) in one cell division time. Given the slow degradation rate of EGFP [50,51] and the lack of regulatory feedback, this yields $G(t)=G_0 e^{(\ln 2)t/\tau}$, where G_0 is the initial EGFP level in each cell cycle and τ is the duration of the cell division cycle. The mean EGFP is then $\langle G \rangle = G_0/\ln 2$, and since $G_0 \rightarrow 2G_0$ (on average) at steady state, G_0 is also approximately the number of proteins produced per cell division. Using the measured cell division duration for each strain then yields the number of proteins expressed per minute for each strain and plasmid type.

C. Variability

The intercell variability in protein expression levels was measured with both flow cytometry and fluorescence microscopy, yielding the results shown in Fig. 1 and Table II. By flow cytometry, the fluorescence was measured from 200 000

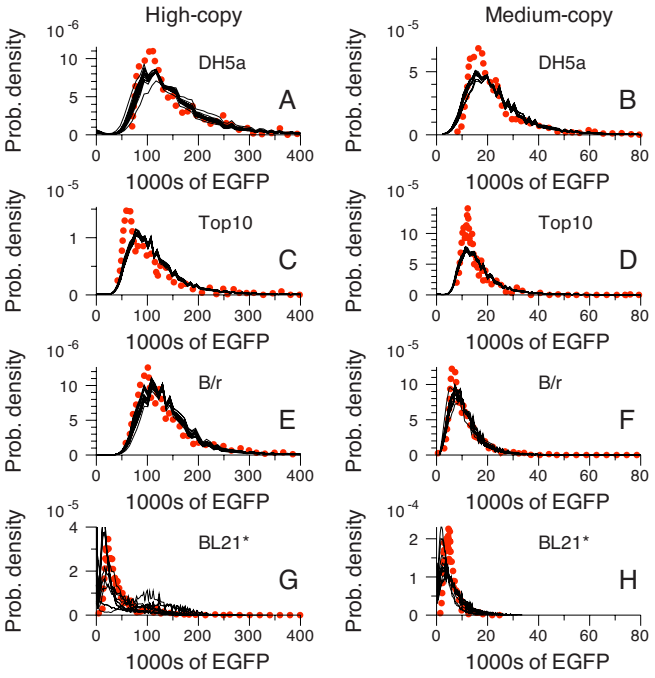

FIG. 1. (Color online) Probability density functions for protein expression levels across the population for each cell strain and plasmid type, from microscopy (dotted) and flow cytometry (solid), each scaled by the measured mean numbers of EGFP. Cytometry results superimpose multiple independent colonies, as follows: DH5α, high (14 colonies) (a) and medium (11 colonies) (b); Top10, high (15 colonies) (c) and medium (13 colonies) (d); B/r, high (15 colonies) (e) and medium (16 colonies) (f); and BL21*, high (10 colonies) (g) and medium (15 colonies) (h).

cells per sample (100 000 in the slow-growing BL21* high-copy case), and 10–15 independent colonies were selected, grown, and analyzed in a single session, for each strain-plasmid combination. By microscopy, 350–1300 cells were imaged, from at least two independent colonies for each plasmid-strain combination. In Fig. 1, the mean EGFP values described above have been used to place the fluorescence intensities obtained by cytometry and microscopy on the same scale: the average intensity across all colonies was used to create a conversion factor from fluorescence intensity to number of EGFP, and this constant was applied to each measured distribution. Since absolute intensity values varied from day to day owing to optical alignment and fluidics fluctuations in the instrument, a single intensity-to-EGFP factor could not be applied across all strains; each strain-plasmid combination was measured in a single day, however, so a consistent conversion factor applied within each of the plots in Fig. 1.

The cytometer and microscope distributions match well, except that the cytometry consistently shows a less sharp onset at the low end of the scale. This is an artifact of the image processing used to identify cells in microscopy: very small cells could not be reliably distinguished from debris, so objects of that size were excluded from the distribution in microscopy. The slight differences in coefficient of variation (CV, standard deviation divided by the mean) values obtained by the two methods (Table II) arise because the effect

TABLE II. Variability in protein expression over cellular populations, determined by flow cytometry and microscopy, given as coefficient of variation expressed as a percentage (% CV). Error ranges for flow cytometry represent one standard deviation, over a minimum of ten independent colonies (colony numbers listed in the caption to Fig. 1). For BL21* microscopy, error is one standard deviation over three independent colonies; in all other cases the microscopy data for two colonies were combined into a single data set. The final column reports variability when each cell's intensity is divided by its size (in pixels) from the microscope images.

Cell strain	Plasmid type	% CV (flow cytometry)	% CV (microscopy)	% CV (microscopy, size scaled)
DH5α	High	55.0 ± 1.0	58.4	20.2
	Medium	52.9 ± 0.7	56.0	12.6
Top10	High	55.7 ± 1.6	76.4	25.0
	Medium	57.7 ± 1.9	74.1	12.8
B/r	High	51.5 ± 2.8	66.8	24.2
	Medium	58.6 ± 2.5	61.0	12.4
BL21*	High	86.0 ± 12	78.4 ± 11	27.2 ± 2.1
	Medium	75.2 ± 1.5	51.7	12.1

on the CV of the small percentage of outliers in the tail is larger in the smaller sample sizes available through microscopy than in cytometry. Within each cell strain, we observe that the distributions for high- and medium-copy plasmids have essentially identical shapes, and nearly overlap when normalized by their respective means (shown in Fig. 6 after the introduction of our computational model).

Within each cell strain, the distributions for multiple independent colonies overlap nearly exactly (Fig. 1), yielding CVs differing by only a few percent (Table II), with the exception of the high-copy plasmid in strain BL21* (discussed further, below). This confirms that consistent behavior can be obtained from our simple system (given a fixed genomic background), which is of course a key requirement for the design of synthetic genetic regulatory systems.

D. Cell size

One significant component of the observed variability arises from the nominal doubling of both cell size and protein number over each cell division [40]; this effect would be present even for purely deterministic gene expression and cell division, since we measure absolute protein numbers and deterministic growth and doubling in an asynchronously dividing population creates a distribution of protein numbers across the population (see Sec. II F, below). In fact, there is known to be significant stochasticity in both gene expression [20,23–40] and cell division [52,53], and here we have attempted to separate cell size effects from other sources of variability.

Obtaining accurate size information from scattered light in flow cytometry can be problematic [54], and although a recent study [55] has established the possibility of making a connection between cell volumes and scattering information, we lack such a calibration for our instrument and cell strains.

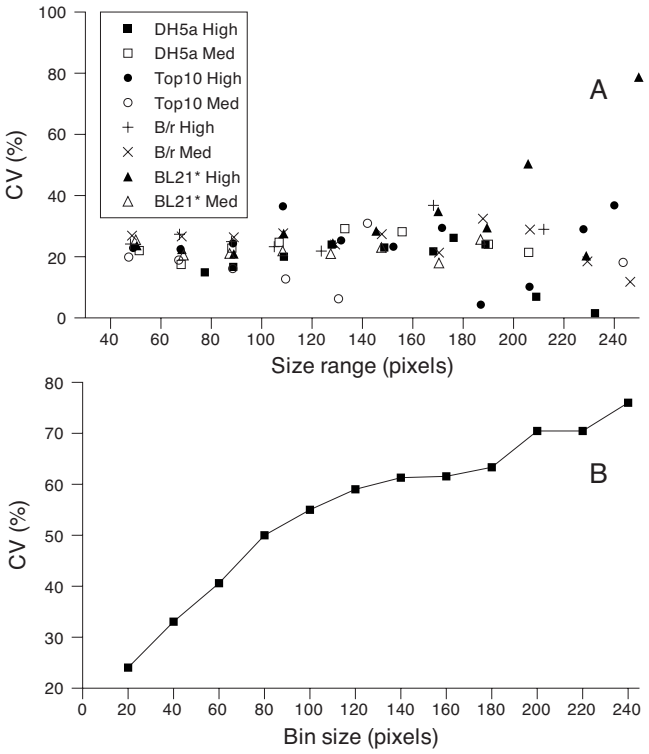

FIG. 2. Dependence of coefficient of variation (CV) of EGFP expression in the cellular population on microscopy bin size. (a) Binning the cell sizes and finding the CV within each bin shows that CV remains relatively constant for all cell sizes, implying constant noise at all points in the cell division cycle. Results are shown for all cell strains and plasmid types: DH5α-High and -Medium (filled and unfilled squares); Top10-High and -Medium (filled and unfilled circles); B/r-High and -Medium (pluses and crosses); and BL21*-High and -Medium (filled and unfilled triangles). (b) The average CV over the population is shown as a function of the bin size selected in (a). This plot shows strain Top10 containing our high-copy plasmid; results are similar for all strains.

We therefore report cell sizes only for microscopy, using the image area in pixels as a size measure (cells were distinguished from background by thresholding using automated software, hand checked, as described in Sec. IV, below). Plotting fluorescence intensity against cell size shows, as expected, a strong correlation (raw data not shown), and dividing each cell's total intensity by its size significantly narrows the distribution (see Table II for the CVs of the size-scaled distributions and Fig. 6 for the distributions themselves). The size-scaled CVs remain roughly constant throughout the cell cycle, as may be seen by plotting the CVs derived from cells only within a narrow range of pixel sizes, and sweeping this range across all cell sizes [see Fig. 2(a)]. Figure 2(b) illustrates the effect of changing this bin size: bin size is selected, swept across the range of cell sizes, and the resulting CV is computed. For very small bin sizes, the resulting CV converges to the value obtained by scaling the intensity by each cell's size, while large enough bin sizes encompass the entire population and thus converge to the unscaled CV. Between these extremes, the CV increases rapidly with bin size, suggesting that cell size effects cannot be trivially removed by restricting measurements to a range of cell sizes, unless very precise size exclusion can be achieved. In our case, convergence between the scaled and unscaled estimates of CV occurs only when the cell sizes are restricted to less than 10% of the total size range.

One strain-plasmid permutation, the BL21* strain containing the high-copy plasmid (BL21*-High), displayed very different behavior from the others. Independent BL21*-High colonies grew at a wide variety of rates (thus the large error in mean cell division time in Table I), and showed very different ranges of gene expression and variability from one to another; see the widely varying distributions in Fig. 1(g). These changes from one colony to another appeared to arise mainly through alterations in the cell division process: multiple colonies examined by microscopy showed very different size distributions (data not shown). It is known that the presence of plasmids interacts with cellular metabolism [56,57], and it appears that the genomic modifications inherent in the BL21* strain render it particularly sensitive to the presence of our high-copy plasmid, though given the complexity of such interactions we cannot point to a reason for this particular behavior. The result points out the need for caution in the choice of cell strain in synthetic biology, since some choices will lead to poor, inconsistent results.

E. Plasmid copy number regulation

After scaling away cell sizes, intercellular variability in our populations is significantly reduced, but still nonzero (see Table II). There is a striking difference between the high-copy and medium-copy plasmids, in the remaining variability after size scaling: while the high-copy plasmids yield CVs ranging from 20% to 30% across the strains, the medium-copy plasmids give CVs consistently in a narrow range, between 12% and 13%. Since the plasmids contain identical promoter-gene pairs for gene expression, the key difference between them is their origin of replication, controlling the regulation of plasmid copy numbers. Plasmid copy number variation is known to be a source of variability in gene expression [8,27,30,31].

Some information is available about the molecular basis for the regulation of plasmid copy numbers in the plasmids we have studied. The backbone of our medium-copy plasmid is pBR322, which contains a ColE1-type replication origin [42,58,59]. Single-stranded RNA II initiates plasmid replication by serving as a primer for DNA polymerase. This initiation is regulated by RNA I: when RNA I hybridizes to RNA II, it destroys the priming conformation of the RNA II molecule and halts replication. The binding of RNA I to RNA II is favored by the Rop/Rom protein [42]. Our medium-copy plasmid, incorporating the Rop protein, exhibits negative feedback in the control of its plasmid copy number: when many copies of the plasmid are present, they transcribe more RNA I, and this acts to repress additional plasmid replication, while few copies of the plasmid present allows replication to proceed with less repression. While negative feedback does not universally lead to decreased noise [8], its effect in the case of our medium-copy plasmid appears to be tight regulation of the number of copies of the plasmid present in each cell [42,43,58]. Our high-copy plasmid is from a family

of vectors (pUC) developed from the pBR322 family [45]. They have the same type of origin of replication, but a point mutation in RNA II weakens the RNA I-RNA II interaction [60], and the plasmids lack the Rop/Rom protein that would partially suppress the effect of this mutation [60]. The result is a near-elimination of negative feedback on the plasmid copy number, and a significant increase in the number of copies of the plasmid per cell [43,45,60]. We observe increased variability and wider distributions in the high-copy cases (Table II and Fig. 6), suggesting that the plasmid copy number variability may be larger for the high-copy plasmids. We do not currently have data on the plasmid copy numbers in individual cells.

F. Formulation of simple models

The experimental data clearly identifies two key factors in the behavior of the system: the pattern of cell division; and the degree of regulation of the copy numbers of the two plasmid types. Here, we formulate highly simplified biochemical kinetic models to explore the minimal framework required to represent the system.

Addressing cell growth and division, we note that bacteria are known to increase their volume exponentially with time [49]; this may be represented as a schematic "reaction" $V \xrightarrow{k} 2V$ where species V represents the cellular volume. If cells grow and divide perfectly from an initial size V_0 to a final size $2V_0$, we obtain the deterministic relation $V(t) = V_0 \exp(kt/\tau)$, where τ is the cell division time. Treating the volume as a random variable, we find its cumulative distribution function to be $F(V)=t(V)/\tau=(1/k)\ln(V/V_0)$ for $V \in [V_0, 2V_0]$, where $t(V)$ represents the time at which a volume V is reached. Taking the derivative with respect to V yields the density function $f(V)=(kV)^{-1}$, $V \in [V_0, 2V_0]$. This gives the distribution of cell volumes expected in an asynchronously growing (and thus uniformly distributed over all times in the cell cycle) population of bacterial cells. Figure 4 plots the deterministic model distribution, normalized by mean volume, against the experimental cell size distributions from microscopy, where each distribution has also been normalized by its mean to allow direct comparisons. Note that the experimental distributions do favor smaller cells, as the deterministic distribution does, but the real distributions drop off much more rapidly than the $1/V$ dependence predicted by the deterministic first approximation, and lack a sharp cutoff at some maximum size.

The experimental cell size distributions have an exponential tail, and we incorporate that into a simple model as follows. Assume that the cells have a threshold volume V_{thresh}, and that they have a constant probability of dividing at each volume past V_{thresh}, effectively following a Poisson process in volume rather than time; this yields an exponential distribution of cell division volumes for $V > V_{\text{thresh}}$. We have incorporated this into a biochemical model by using the original exponential-growth "reaction" for the volume, $V \xrightarrow{k} 2V$, modifying the cell division process so that rather than always dividing at a division volume $V_{\text{div}}=2V_0$, a new cell division volume is selected randomly after each division event. By setting $V_{\text{div}}=(2V_0-\lambda)+X$, where X is a random variable chosen from an exponential distribution with mean λ, we obtain a mean cell division size of $2V_0$, but allow for the possibility of cells dividing at smaller or larger volumes. This simple system has been implemented using the BIONETS software for simulating stochastic chemical reaction systems [61], which provides a convenient user interface for implementing the Gillespie stochastic simulation algorithm [62]. By treating volume as a species in a stochastic reaction, additional randomness is introduced into the cell size distribution by making the growth into a series of Poisson processes with mean step time $1/(kV)$; smaller numbers of volume steps lead to larger variations in the time to cell division for a given V_{div}. Upon reaching the cell division volume threshold for each cycle, the species V is divided approximately in half according to a binomial distribution, using a feature of BIONETS specifically intended to simulate segregation of biomolecules between two newly divided cells. Using $V_0=250$ and $\lambda=200$ (and thus $V_{\text{thresh}}=2V_0-\lambda=300$) yields the $V(t)$ time series shown in Fig. 3 and the stochastic distribution shown in Fig. 4. These values (and all others in this section) were obtained by a combination of parameter sweeping and inspection, with no formal fitting procedure.

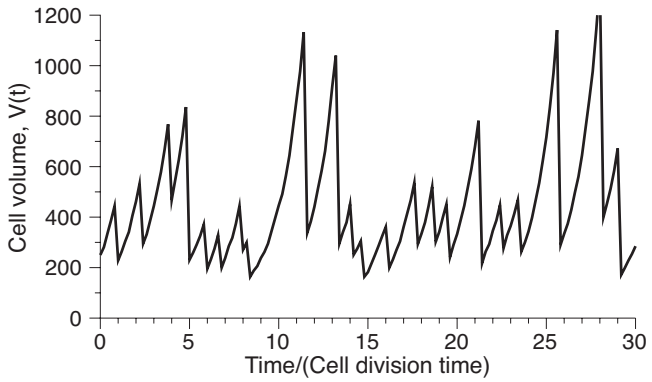

FIG. 3. Cellular volume versus time for the stochastic cell division model. Time has been scaled by the mean cell division time. After each cell division event, a new cell division volume is chosen from an exponential distribution, as described in the text. The nominal initial volume is $V_0=250$ and the mean of the exponential distribution is $\lambda=200$. The distribution of cell sizes from the model is compared to experimental results in Fig. 4.

Next, we incorporate gene expression of the EGFP protein into the model. Combining transcription and translation into a single step, we write

$$V \xrightarrow{k} 2V, \tag{1}$$

$$D \xrightarrow{k_d} 2D, \tag{2}$$

$$D \xrightarrow{k_p} D + G, \tag{3}$$

where V represents the cellular volume, as above, D (for DNA) represents the plasmid (high or medium copy), and G represents the EGFP protein. Cell division occurs when V reaches a threshold, with the threshold varying from one cell

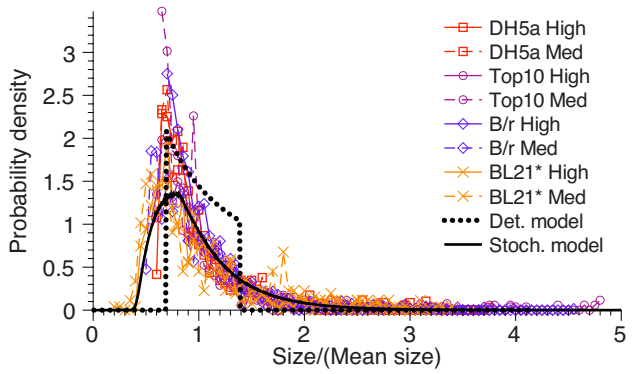

FIG. 4. (Color online) Cell size distribution scaled by mean size, for experiment and models. Experimental cell size distributions (pixel areas from fluorescence microscopy) are plotted against the size normalized by the mean size of each population (symbols given in the legend). A simple deterministic model of cell division gives the distribution shown as a heavy dotted line, with its too-sharp transitions and lack of the experimentally observed tail. Modifying this simple model with cell division sizes chosen from an exponential distribution after each division event yields the stochastic model shown as a heavy solid line.

cycle to the next, as described above and illustrated in Fig. 3. Species D and G are also divided binomially at each cell division event [61].

We approximate the increase in species D as exponential growth, assuming that replication of plasmids proceeds asynchronously with a rate proportional to the number of plasmids present. However, completely unconstrained exponential growth of plasmid numbers is unrealistic for two reasons: there is no acknowledgment of the physical limitations on plasmid growth imposed by the finite resources available to the cell's replication machinery; and there is no natural steady state for the plasmid copy number under these conditions: the numbers can double from 100 to 200 or from 1000 to 2000 in each cell division, with no tendency to leave any particular range once fluctuations have driven the system into it. Bacteria do in fact show consistent average numbers of plasmids per cell [42–45], suggesting that an unconstrained exponential cannot provide a full description of plasmid replication.

As noted above, our two plasmid types vary in their regulatory feedback, with the medium-copy plasmid exerting control over its own copy number while the high-copy plasmid has those regulatory mechanisms disrupted. We thus formulate two different simple models in which we treat plasmid regulation differently. For the high-copy plasmid, we simulate the limited resources present in the cell by having k_d in reaction (2) monotonically decrease with D using a sigmoidal function, $k_d = k_0 f(D)$ where

$$f(D) = 1 - (1 + e^{-(D-\hat{D})/T})^{-1}. \quad (4)$$

Taking $\hat{D}=800$ and $T=100$ yields a function that is nearly unity for the first 500 plasmid copies, then falls off between 500 and 1000 copies, approaching zero by 1500. Implementing this form of constrained exponential growth, where the constraint represents resource limitations, allows the plasmid copy number to undergo wide excursions, but prevents unrealistic explosions to infinity, as shown in Figs. 5(a) and 5(b). The resulting distributions of the total intensity and the intensity scaled by cell size are shown in comparison to the experimental results in Fig. 6; once again, we normalize by the mean of each distribution to allow direct comparisons.

Simply shifting the same exponential growth down to a lower average plasmid copy number does not capture the difference in experimental behavior between our two plasmid types; recall that both types show similar distributions in total fluorescence intensity (and thus, by inference, in total EGFP content), but the medium-copy plasmid shows significantly smaller CVs when the cell size is scaled away (Table II). Direct simulation of the feedback mechanism in our medium-copy plasmid would involve more biochemical detail than we aim to include in these schematic models, including a number of unknown rate constants [42,43,58]. To represent the self-regulation of plasmid copy number in the medium-copy case in an admittedly crude fashion, we use reactions (1)–(3) but impose the constraint that the plasmid copy number is reset to a constant value D_0 after each cell division. This permits excursions during periods of cell growth, but prevents the exponential growth dynamics from causing the plasmid numbers to "wander" too far from the nominal number of copies, as shown in Figs. 5(c) and 5(d).

These simple assumptions about plasmid copy number (resource-constrained exponential growth for the high-copy plasmid and strongly regulated exponential growth for the medium-copy plasmid) yield good agreement with the experimentally observed EGFP distributions. The models are deliberately highly schematic, but it is interesting to note that just a few assumptions can yield a good approximation to the observed experimental behavior: the total intensity distributions are very similar for the high-copy and medium-copy strains [see Figs. 6(a) and 6(c)], but when the intensity is scaled by cell size [Figs. 6(b) and 6(d)], the medium-copy plasmid distribution is narrowed more significantly than the high-copy distribution: the CVs are 24% for the high-copy version of the model, and 17% for the medium-copy version.

The model shows nearly deterministic production of EGFP, when the rates are scaled to yield average numbers of EGFP (species G in the model) matching the experimentally measured values reported above; that is, the curves within each cell division are a good match for those that would be obtained by neglecting stochastic fluctuations in the production process. By combining transcription and translation into a single process, we have neglected a potential additional source of fluctuations from bursts of translation from small, highly fluctuating populations of transcripts [27,29,30]. We have considered a slightly more complex model in which mRNA is produced and proteins are translated from the mRNA. The essential observation from the transcription plus translation model was that the mRNA numbers were not small enough to disrupt the near-deterministic production of the EGFP. Experimental estimates suggest that each mRNA transcript is translated roughly 50 times before it degrades, and even using a conservative 100 translations per transcript still gives us populations of hundreds to thousands of mRNAs per cell, to yield the tens to hundreds of thousands

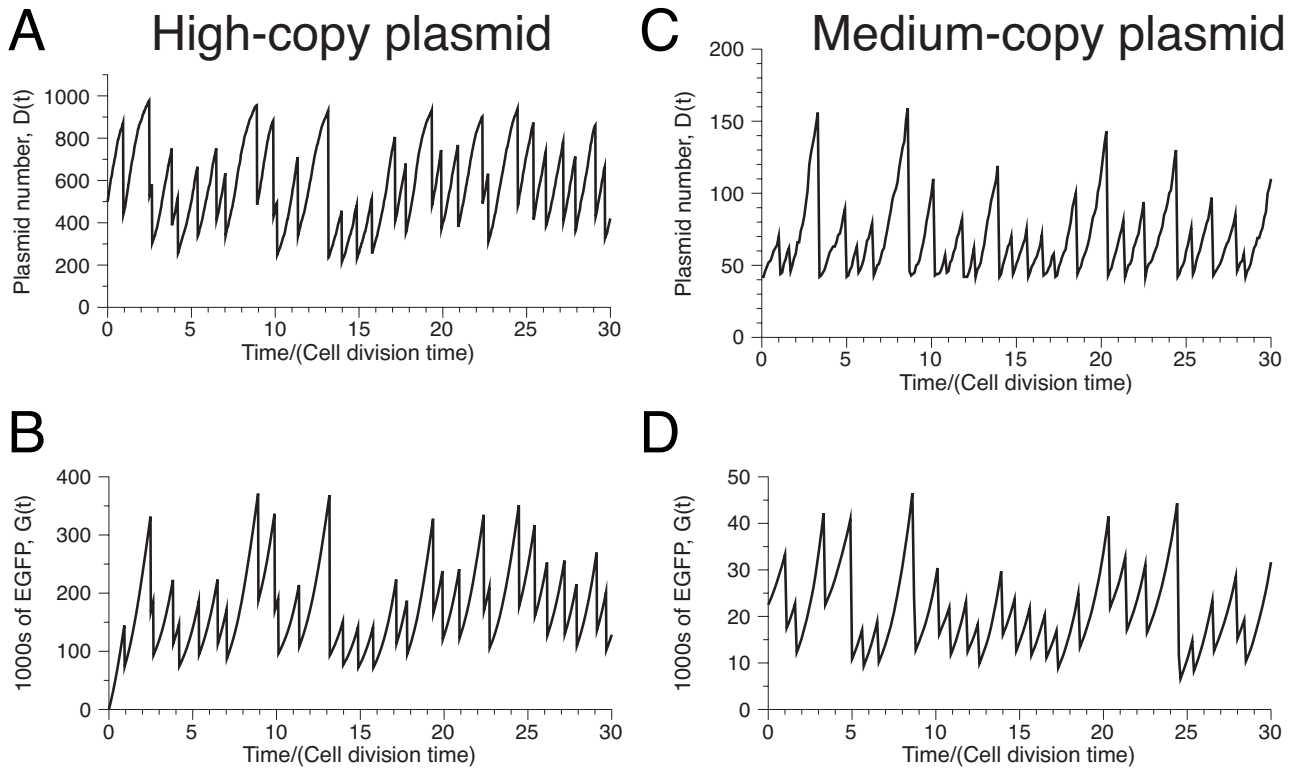

FIG. 5. Time series for plasmid copy number and EGFP number from the model, showing approximately the first 30 cell divisions from longer runs. (a) and (b), curves for species D and G, respectively, from the high-copy plasmid version of the model. (c) and (d), curves for species D and G, respectively, from the medium-copy plasmid version of the model. All times in the model have been scaled so that one time unit equals the mean cell division. Parameters (given in Sec. IV) have been chosen to approximately match the plasmid and EGFP numbers from the DH5α strain. Note the smooth shape of the $G(t)$ lines within each cell division, indicating near-deterministic behavior; the time sampling used in the plots is fine enough that internal fluctuations would be visible.

of proteins per cell we observe experimentally. mRNAs present in these quantities did not significantly add to the overall variability of species G (results not shown), prompting us to prefer the more highly simplified single-step production process.

As has been previously noted [4], variations in plasmid behavior across different cell strains are not generally incorporated into models because of a lack of sufficient information to characterize the differences between strains. We face the same situation, here: the complexity of the full bacterial genome is such that it is not currently tractable for us to formulate a model of how the specific alterations in each strain's genome lead to the observed differences in gene expression levels and variability.

III. CONCLUSIONS

The work presented here provides experimental evidence of the importance of cell size, plasmid copy number regulation, and genomic variations on the level and variability of protein expression in bacterial cells. We have used highly simplified models to capture the main features of our experimental observations. Somewhat more elaborate models could be formulated to incorporate more detailed pictures of cell division [27,40,52,53] and plasmid copy number control [31], while still remaining relatively simple and high level. Since changes in the genomic background necessarily have implications across the entire cell, attempting to form an explicit model of the effects of altering the genome on expression of our gene interest is likely to require a more complex model, incorporating highly detailed information about the internal environment of the cell.

The ability to implement novel regulatory networks offers the possibility of designing human medical interventions at the cellular level, and also of programming microorganisms to carry out desired tasks [1–7,11–15,17]. For the latter, plasmid-borne genes of the type considered here represent one of the simplest methods currently under consideration. Two of our results are of particular interest to engineers approaching synthetic biological designs: that the choice of cell strain (and thus, the genomic background against which the synthetic system operates) is important, leading to substantial changes in both expression levels and levels of intercell variability; and that near-deterministic behavior can be achieved (for some combinations of strain and plasmid type), even at relatively low mean numbers of proteins per cell (for example, the BL21* medium-copy plasmid, where the size-scaled variability is just 12% percent while it expresses 6000 proteins per cell). Fundamental to any engineering discipline is system characterization, the study of the underlying dynamics of the system to be modified and controlled. This study and other recent work

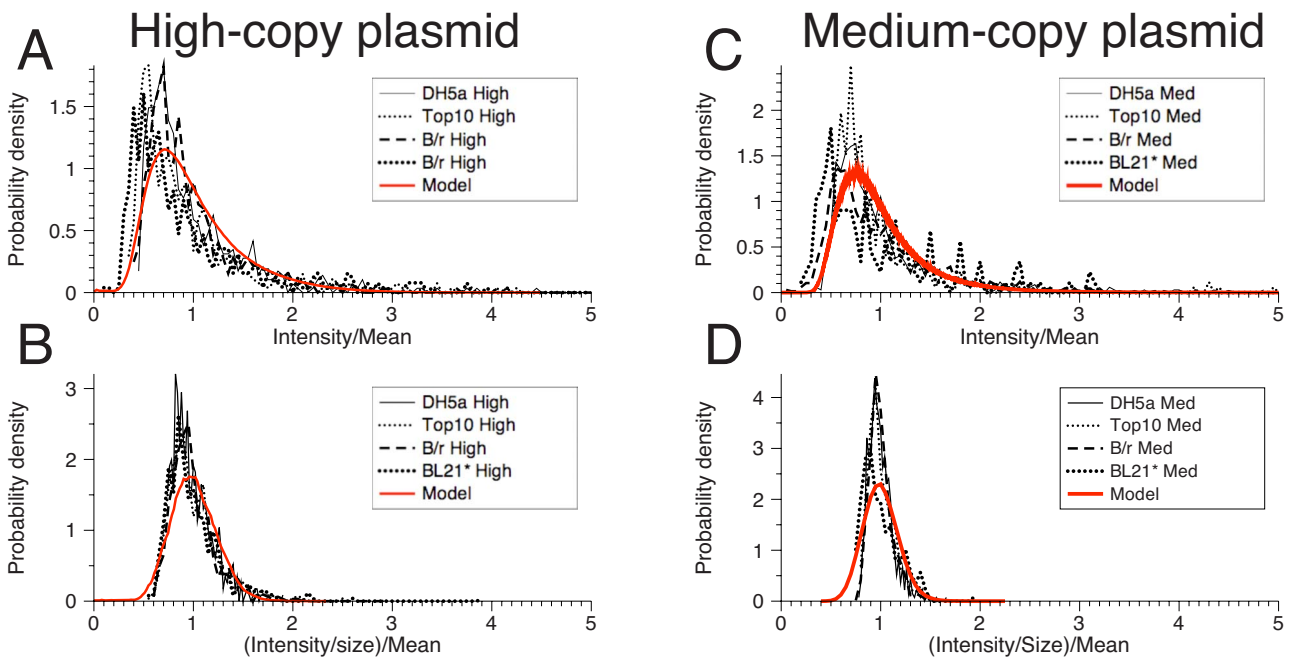

FIG. 6. (Color online) Cellular population distributions of EGFP fluorescence intensity and of EGFP fluorescence intensity scaled by cell size. Experimental results report total fluorescence intensity from microscopy or total intensity divided by cell size in pixels, normalized by the mean in each case. Model results show the distributions for species G or for the quantity G/V, also normalized by the mean in each case. (a), (b) Total intensity and size-scaled intensity distributions for the high-copy plasmid experiments and model. (c), (d) Total intensity and size-scaled intensity distributions for the medium-copy plasmid experiments and model. Note that the total intensity distributions are similar for the high- and medium-copy cases, while the size-scaled distributions show a greater narrowing for the medium-copy case, for both the experiment and the simple models described in the text.

[4–7,11,12,17,20,23,25,32,36,40,46,63,64] represent steps toward the level of system characterization required to implement systematic design in the emerging field of synthetic biology.

IV. MATERIALS AND METHODS

Genes, plasmids, and cell strains. All *Escherichia coli* strains and plasmids used in this study are listed in Table III. High- and medium-copy plasmids were constructed by rearranging promoters and genes using standard molecular biology techniques (Sambrook and Russell [65]), placing the $P_{LtetO-1}$ promoter (from plasmid pZA31-luc, Lutz and Bujard [41]) upstream of the egfp gene on two plasmid backbones: the high-copy (pUC ORI) pEGFP plasmid (Clontech, Mountain View, CA) with the lac promoter removed; and the medium-copy (ColE1 ORI) pBR322 plasmid (Clontech) with the tetR gene truncated. Maps of the two plasmids are shown in Fig. 7. Plasmid DNA was purified using QIAPrep Spin Miniprep kits (Qiagen Canada, Mississauga, ON), sequenced (The Centre for Applied Genomics, Toronto, ON), and transformed by electroporation (ECM 399, BTX, San Diego, CA) into four cell strains: DH5α, Top10, B/r, and BL21* (Invitrogen Canada, Burlington, ON). All other enzymes and buffers were purchased from New England Biolabs Canada (Pickering, ON).

Cell growth and sample preparation. Cultures were grown overnight (16 h) in Lauria-Bertani (LB) medium (BioShop Canada, Burlington, ON) plus 100 μg/ml ampicillin (BioShop) at 37 °C from single colonies and diluted 1:100 in LB+100 μg/ml ampicillin. Growth curves of optical density at 600 nm (OD_{600}) values for each cell-strain–plasmid combination were obtained (Ultrospec 100 pro spectrophotometer, Biochrom, Cambridge, England) and doubling times determined (see Table II). For analysis, cultures were grown to log (exponential growth) phase using incubation times from the growth curves, confirmed by measuring the final OD_{600} value. Cells were spun down and washed twice with phosphate-buffered saline (PBS, pH 7.4) to minimize background fluorescence from the medium. Cultures from the PBS wash were diluted in PBS by varying degrees: one- to threefold for fluorimetry, tenfold for microscopy, and roughly 100-fold for flow cytometry, adjusted to achieve desired cell count rates in the cytometer.

EGFP quantification. The average number of EGFP expressed per cell was determined by fluorimetry (QuantaMaster; Photon Technology, Birmingham, NJ), exciting the samples at 488 nm and measuring the emission at 507 nm. Standard curves were created immediately prior to cell culture analysis by measuring the fluorescence intensity of purified EGFP (BioVision, Mountain View, CA) in PBS buffer (pH 7.4) over a range of known concentrations. Concentration of the EGFP solutions was measured by uv absorbance at 260 and 280 nm (uv-visible Cary-14 spectrophotometer; Olis, Bogart, GA). Fluorescence intensities of PBS-suspended cell samples were measured, and cellular autofluorescence was subtracted; autofluorescence for each cell strain was found by measuring the fluorescence output of cells con-

TABLE III. Bacterial strains and plasmids used in this study. Genomic modifications relative to *E. coli* wild type for each of our cell strains are listed, along with information on each of the plasmids used. Plasmids pEGFP, pBR322, and pZA31-luc provided components used to construct the high-copy and medium-copy plasmids pTEGFP and pTLEGFP; only these last two plasmids were experimentally studied.

	Characteristics	Source
	Cell strains	
DH5α	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG ϕ80dlacZΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ-	Invitrogen
Top10	F^- endA1 mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 deoR nupG recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(Str^R) λ-	Invitrogen
B/r	F26 his thy	Professor J. Grimwade, Florida State University
BL21*	F^- ompT gal dcm lon hsdS_B(r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Invitrogen
	Plasmids	
pEGFP	pUC ORI (high copy), A_p^R, EGFP expressed from P_{lac}	Clontech
pBR322	pMB1 ORI with *rop* (medium copy), A_p^R, and T_c^R	NEB
pZA31-luc	C_m^R, luciferase gene expressed from $P_{LtetO-1}$; source of promoter used in our study	Lutz and Bujard [41]
pTEGFP	pUC ORI (high copy), A_p^R, EGFP expressed from $P_{LtetO-1}$	This study
pTLEGFP	pMB1 ORI with *rop* (medium copy), A_p^R, EGFP expressed from $P_{LtetO-1}$	This study

taining no plasmids. The standard curve allowed conversion of the cellular fluorescence intensity to an equivalent total number of EGFP molecules per milliliter. Dividing by the concentration of cells/ml then yielded the number of EGFPs per cell. Cellular concentration was determined by calibrating OD_{600} values to cell counts (using standard colony-counting methods to determine the number of colony-forming-units per ml of culture, as a function of OD_{600}), corrected by the percentage of live cells in the sample. Live and dead percentages were determined by flow cytometry of propidium iodide (Sigma Aldrich Canada, Oakville, ON) stained populations to detect cells with perforated membranes.

Flow cytometry. All flow cytometry was performed using an Epics Altra cell sorter (Beckman Coulter Canada, Mississauga, ON). Cells were excited at 488 nm at a power of 300 mW with an Innova 70 laser (Coherent, Santa Clara, CA). Data were analyzed using DATATANK (Visual Data Tools, Chapel Hill, NC). A region centered on the median of the forward and side scatter outputs was defined, set to include 95% of the cells; the remaining 5% of cells along each scatter axis were excluded to minimize noncell events. Fluorescence intensity was measured with a 500–550 nm bandpass filter. Fluorescence intensities were thresholded in two ways. At the low end, control samples of cells from each strain containing no plasmids were analyzed, and a lower intensity threshold was established; a single threshold was used for all strains, eliminating 95–99.9% of control cells (percentages varied by cell strain). At the upper end, outliers were excluded by comparison with microscopy: for each strain and plasmid, the brightest cell observed in the microscopy data was used to define an upper limit on the brightest cytometry events, and events beyond this threshold were excluded. Events excluded by this threshold were at least six standard deviations above the mean of the truncated distribution, except in the BL21* high-copy runs (four standard deviations above).

Microscopy. Epifluorescence was measured using a TE 2000U microscope (Nikon Canada, Mississauga, ON) with a dry objective (40× 0.95, Nikon Canada). The microscope was equipped with a Piston GFP filter cube (HQ470/40x, Q495LP, HQ515/30m; Chroma Technology, Rockingham, VT). Images were collected with a charge-coupled device camera (Cascade 650, Photometrics, Tucson, AZ) with fixed on-chip multiplier gain [66], with a 100 ms integration time to reduce photobleaching. Image analysis was performed with METAMORPH v6.2r6 (Molecular Devices, Sunnyvale, CA). Cell sizes were determined using a fixed intensity threshold to define cell boundaries, and cells were distinguished from debris using METAMORPH image analysis tools to define regions of total size, shape factor, and total and average fluorescence intensity that constituted probable cells rather than noncells. In each case, the software-identified cells were double checked manually; phase contrast images

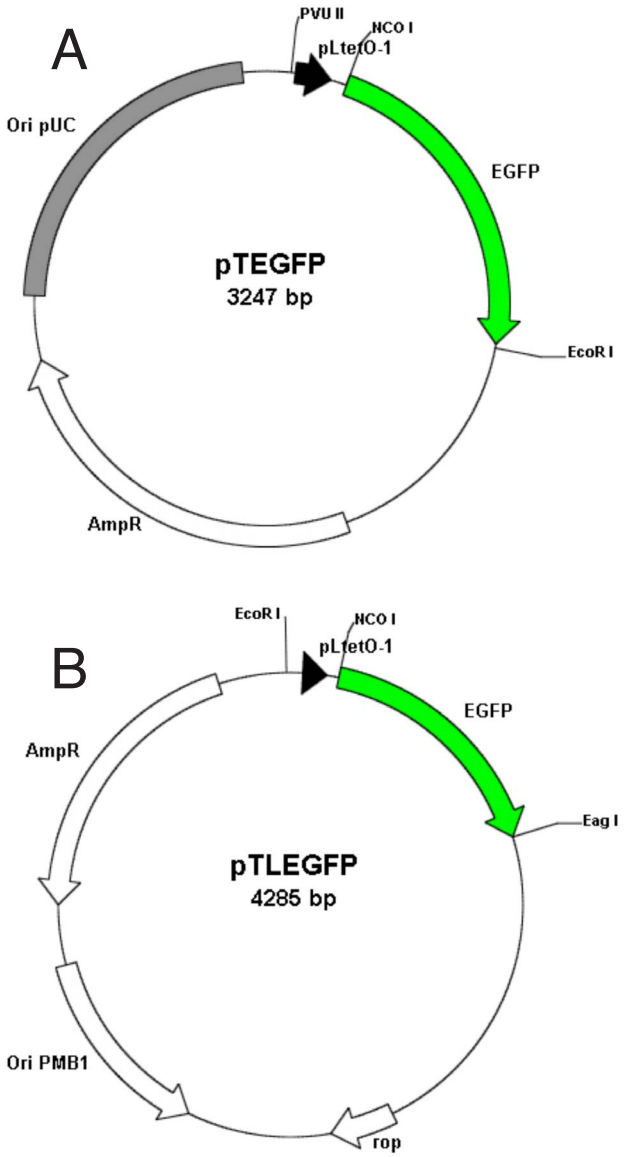

FIG. 7. (Color online) (a) High-copy plasmid pTEGFP, constructed using the pEGFP plasmid backbone. Restriction enzymes PvuII and NcoI can be used to excise the $P_{LtetO-1}$ promoter. The gene *egfp* can be excised using NcoI and other restriction sites in the 3 multiple cloning site. (b) Medium-copy plasmid pTLEGFP, constructed using the pBR322 plasmid backbone. Restriction enzymes *Eco*RI and NcoI can be used to excise the $P_{LtetO-1}$ promoter. The gene *egfp* can be excised using NcoI and EagI.

of the cells were used to gain a sense of the range of cell sizes present in the population, and this was used to allow the human operator to make final decisions about including or excluding borderline objects.

Plasmid copy number determination. The average number of plasmids per cell was determined by extracting plasmid DNA, quantifying the mass of DNA present, dividing this by concentration of cells in the culture (determined by OD_{600}, calibrated to the number of colony-forming units present, as described above), and converting the mass of plasmid DNA per cell into an equivalent number of plasmids using the known molecular weight of each plasmid type. Plasmid DNA was extracted from 5 ml of 3 h cultures using QIAPrep Spin Miniprep kits (Qiagen Canada, Mississauga, ON). Running the extracted plasmid DNA in an agarose gel stained with ethidium bromide, the amount of DNA was quantified through comparison with the stated mass present in a standard DNA ladder (New England Biolabs Canada, Pickering, ON). Intensity comparisons were performed using METAMORPH image processing software (Molecular Devices, Sunnyvale, CA). Results were averaged over two to three colonies per strain. In the case of the BL21* medium-copy populations, too little DNA was extracted to be quantifiable by this technique.

Modeling. Models were implemented using modified versions of source code generated by BIONETS [61]; code freely available upon request. Simulation runs were generated and analyzed using DATATANK (Visual Data Tools, Chapel Hill, NC). All times and rates were scaled to make one time unit equal to the mean cell division time, and a minimum of 100 000 cell divisions were simulated to generate the distributions shown in the figures. For the high-copy plasmid version of the model, parameter values used were $k=\ln 2$, $k_p=(5/4)\ln 2$, $k_p=220$, $V_0=250$, $\lambda=200$, $\hat{D}=800$, and $T=100$. The medium-copy plasmid version of the model used the following parameters: $k=k_d=\ln 2$, $k_p=200$, and $D_0=42$.

ACKNOWLEDGMENTS

We thank Professor Julia Grimwade of Florida State University for providing us with the B/r cell strain and Professor Hermann Bujard of the Zentrum fur Molekulare Biologie Heidelberg for the $P_{LtetO-1}$ promoter. This work was funded by the Natural Science and Engineering Research Council (NSERC) of Canada, the Canada Foundation for Innovation (CFI), and the Ontario Photonics Consortium (OPC).

[1] J. Hasty, D. McMillen, and J. J. Collins, Nature (London) **420**, 224 (2002).
[2] M. Kaern, W. Blake, and J. J. Collins, Annu. Rev. Biochem. Eng. **5**, 179 (2003).
[3] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, Mol. Syst. Biol. **2**, 2006.0028 (2006).
[4] R. Weiss, S. Basu, S. Hooshangi, A. Kalmbach, D. Karig, R. Mehreja, and I. Netravali, Natural Comput. **2**, 47 (2003).
[5] M. B. Elowitz and S. Leibler, Nature (London) **403**, 335 (2000).
[6] T. Gardner, C. R. Cantor, and J. J. Collins, Nature (London) **403**, 339 (2000).
[7] A. Becskei and L. Serrano, Nature (London) **405**, 590 (2000).
[8] Y. Dublanche, K. Michalodimitrakis, N. Kummerer, M.

Foglierini, and L. Serrano, Mol. Syst. Biol. **2**, 133 (2006).
[9] C. C. Guet, M. B. Elowitz, W. Hsing, and S. Leibler, Science **296**, 1466 (2002).
[10] J. C. Anderson, C. A. Voigt, and A. P. Arkin, Mol. Syst. Biol. **3**, 133 (2007).
[11] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, Nature (London) **434**, 1130 (2005).
[12] Y. Yokobayashi, C. H. Collins, J. R. Leadbetter, R. Weiss, and F. H. Arnold, Adv. Complex Syst. **6**, 37 (2003).
[13] L. You, R. S. Cox, R. Weiss, and F. H. Arnold, Nature (London) **428**, 868 (2004).
[14] M. Ramachandra *et al.*, Nat. Biotechnol. **19**, 1035 (2001).
[15] C. A. Voigt, Curr. Opin. Biotechnol. **17**, 548 (2006).
[16] C. Grilly, J. Strickler, W. L. Pang, M. R. Bennett, and J. Hasty, Mol. Syst. Biol. **3**, 127 (2007).
[17] W. Weber, J. Stelling, M. Rimann, B. Keller, M. Daoud-El Baba, C. C. Weber, D. Aube, and M. Fussenegger, Proc. Natl. Acad. Sci. U.S.A. **104**, 2643 (2007).
[18] A. B. Cubitt, R. Heim, S. R. Adams, A. E. Boyd, L. A. Gross, and R. Y. Tsien, Trends Biochem. Sci. **20**, 448 (1995).
[19] S. Ghaemmaghami, W.-K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, and J. S. Weissman, Nature (London) **425**, 737 (2003).
[20] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, Nat. Genet. **38**, 636 (2006).
[21] Y. Gerena-Lopez, J. Nolan, L. Wang, A. Gaigalas, A. Schwartz, and E. Fernandez-Repollet, Cytom. Part A **60**, 21 (2004).
[22] N. Rosenfeld, T. J. Perkins, U. Alon, M. B. Elowitz, and P. S. Swain, Biophys. J. **91**, 759 (2006).
[23] K. Baetz and M. Kaern, Nat. Genet. **38**, 610 (2006).
[24] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[25] H. B. Fraser, A. E. Hirsh, G. Giaever, J. Kumm, and M. B. Eisen, PLoS Biol. **2**, e137 (2004).
[26] J. Hasty, D. McMillen, F. Isaacs, and J. J. Collins, Nat. Rev. Genet. **2**, 268 (2001).
[27] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Nat. Rev. Genet. **6**, 451 (2005).
[28] J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman, Nature (London) **441**, 840 (2006).
[29] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).
[30] J. Paulsson, Nature (London) **427**, 415 (2004).
[31] J. Paulsson and M. Ehrenberg, Q. Rev. Biophys. **34**, 1 (2001).
[32] J. M. Pedraza and A. van Oudenaarden, Science **307**, 1965 (2005).
[33] J. M. Raser and E. K. O'Shea, Science **304**, 1811 (2004).
[34] J. M. Raser and E. K. O'Shea, Science **309**, 2010 (2005).
[35] D. Reanney, Nature (London) **307**, 318 (2002).
[36] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, Science **307**, 1962 (2005).
[37] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proc. Natl. Acad. Sci. U.S.A. **99**, 12795 (2002).
[38] P. S. Swain and A. Longtin, Chaos **16**, 026101 (2006).
[39] M. Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. U.S.A. **98**, 8614 (2001).
[40] D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty, Nature (London) **439**, 861 (2006).
[41] R. Lutz and H. Bujard, Nucleic Acids Res. **25**, 1203 (1997).
[42] P. Balbas and F. Bolivar, in *Recombinant Gene Expression: Reviews and Protocols*, 2nd ed., edited by P. Balbas and A. Lorence, Methods in Molecular Biology Vol. 267 (Humana Press, Totowa, NJ, 2004), pp. 77–90.
[43] C. L. Lee, D. S. W. Ow, and S. K. W. Oh, J. Microbiol. Methods **65**, 258 (2006).
[44] J. R. Lupski, S. J. Projan, L. S. Ozaki, and G. N. Godson, Proc. Natl. Acad. Sci. U.S.A. **83**, 7381 (1986).
[45] J. Vieira and J. Messing, Gene **19**, 259 (1982).
[46] W. J. Blake and G. Balazsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, and J. J. Collins, Mol. Cells **24**, 853 (2007).
[47] W. J. Blake, M. Kaern, C. R. Cantor, and J. J. Collins, Nature (London) **422**, 633 (2003).
[48] R. Kemkemer, S. Schrank, W. Vogel, H. Gruler, and D. Kaufmann, Proc. Natl. Acad. Sci. U.S.A. **99**, 13783 (2002).
[49] F. J. Trueba and L. J. H. Koppes, Arch. Microbiol. **169**, 491 (1998).
[50] P. Corish and C. Tyler-Smith, Protein Eng. **12**, 1035 (1999).
[51] A. Varshavsky, Genes Cells **2**, 13 (1997).
[52] M. A. Henson, Curr. Opin. Biotechnol. **14**, 460 (2003).
[53] Z. Kutalik, M. Razaz, and J. Baranyi, J. Theor. Biol. **232**, 285 (2005).
[54] H. M. Shapiro, *Practical Flow Cytometry*, 3rd ed. (John Wiley & Sons, New York, 1995).
[55] O. Julia, J. Comas, and J. Vives-Rego, J. Microbiol. Methods **40**, 57 (2000).
[56] S. Birnbaum and J. E. Bailey, Biotechnol. Bioeng. **37**, 736 (1991).
[57] Z. Wang, L. Xiang, J. Shao, A. Wegrzyn, and G. Wegrzyn, Microbial Cell Factories **5**, 34 (2006).
[58] P. Balbas, X. Soberon, E. Merino, M. Zurita, H. Lomeli, F. Valle, N. Flores, and F. Bolivar, Gene **50**, 3 (1986).
[59] F. Bolivar, R. L. Rodriguez, P. J. Green, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow, Gene **2**, 75 (1977).
[60] S. Lin-Chao, W.-T. Chen, and T.-T. Wong, Mol. Microbiol. **6**, 3385 (1992).
[61] D. Adalsteinsson, D. McMillen, and T. C. Elston, BMC Bioinf. **5**, 24 (2004).
[62] D. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[63] X.-j. Feng, S. Hooshangi, D. Chen, G. Li, R. Weiss, and H. Rabitz, Biophys. J. **87**, 2195 (2004).
[64] T. J. Perkins, J. Jaeger, J. Reinitz, and L. Glass, PLOS Comput. Biol. **2**, e51 (2006).
[65] J. Sambrook and D. W. Russell, *Molecular Cloning: A Laboratory Manual*, 3rd ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
[66] S. Bagh and M. F. Paige, Can. J. Chem. **83**, 435 (2005).