

Comparison and Calibration of Different Reporters for Quantitative Analysis of Gene Expression

Hernan G. Garcia,[†] Heun Jin Lee,[‡] James Q. Boedicker,[‡] and Rob Phillips^{†*}

[†]Department of Physics and [‡]Department of Applied Physics, California Institute of Technology, Pasadena, California

ABSTRACT Absolute levels of gene expression in bacteria are observed to vary over as much as six orders of magnitude. Thermodynamic models have been proposed as a tool to describe the expression levels of a given transcriptional circuit. In this context, it is essential to understand both the limitations and linear range of the different methods for measuring gene expression and to determine to what extent measurements from different reporters can be directly compared with one aim being the stringent testing of theoretical descriptions of gene expression. In this article, we compare two protein reporters by measuring both the absolute level of expression and fold-change in expression using the fluorescent protein EYFP and the enzymatic reporter β -galactosidase. We determine their dynamic and linear range and show that they are interchangeable for measuring mean levels of expression over four orders of magnitude. By calibrating these reporters such that they can be interpreted in terms of absolute molecular counts, we establish limits for their applicability: autofluorescence on the lower end of expression for EYFP (at ~ 10 molecules per cell) and interference with cellular growth on the high end for β -galactosidase (at $\sim 20,000$ molecules per cell). These qualities make the reporters complementary and necessary when trying to experimentally verify the predictions from the theoretical models.

INTRODUCTION

In recent years, our understanding of transcriptional regulation has increased dramatically. This is true both in terms of the number of regulatory circuits that have been dissected and of the precision with which they have been characterized (1–7). As illustrated in Fig. 1, quantitative measurements of gene expression have determined that the mean absolute level of expression of different promoters range over more than six orders of magnitude. The majority of gene products regulated under bacterial and viral promoters is present at levels from 0.1 to 10^5 molecules per cell.

In addition to being concerned with these absolute levels of expression, it is also of interest to know the range over which these promoters can be regulated. As shown in Fig. 2, *C* and *D*, the level of expression of a given promoter can in turn be regulated to vary over several more orders of magnitude. These results make it clear that a quantitative, genome-wide characterization of transcriptional regulation requires techniques with a broad dynamic range and for which the experimental uncertainties have been precisely characterized.

Quantitative experiments like those described above are making it possible to directly compare measurements of regulatory response with the predictions of an increasingly sophisticated host of theoretical ideas for describing regulatory circuits (8–17). This poses an experimental challenge: Is there a technique or techniques that can reliably span the many orders of magnitude in expression and fold-change?

There is a wide variety of different methods for carrying out measurements of gene expression like those described

above (18–21). One classic scheme for measuring the level of gene expression is based upon the enzyme action of β -galactosidase (LacZ) as a reporter in which a substrate for this enzyme can be detected colorimetrically upon cleavage (18). However, the use of fluorescent reporters is increasingly becoming the method of choice, especially with the construction of a variety of libraries in which nearly each and every gene in a model organism can be read out fluorescently (1,3,7). In certain cases, this idea has been pushed all the way to the single-molecule limit where individual molecules in regulatory circuits are detected through their fluorescence (22). mRNA counting both in bulk through quantitative PCR and at the single molecule level is also becoming widespread as a means for quantifying levels of gene expression (7,23). In the cases where antibodies against the protein of interest are available, Western or immunoblotting can provide a quantitative measure of the protein contents of a cell (24,25). Finally, another popular enzymatic technique is based on reporting gene expression levels through bioluminescence (19,26–28).

The quest to quantitatively dissect regulatory networks of all types (4,5,23,29) raises questions about the relative merits of these different measurement techniques. For example, one important question is whether they are linearly related, thus rendering them able to report reliably on the level of expression. Additionally, it is important to determine whether the use of reporters affects cellular processes in any observable way. To that end, in this work, we use systematic experimentation in the context of a well-characterized regulatory network to compare enzymatic and fluorescent reporters as a measure of level of gene expression. Similarly, recent measurements have begun to systematically explore the relation between the amount of

Submitted November 17, 2010, and accepted for publication June 9, 2011.

*Correspondence: phillips@pboc.caltech.edu

Editor: Andre Levchenko.

© 2011 by the Biophysical Society
0006-3495/11/08/0535/10 \$2.00

doi: 10.1016/j.bpj.2011.06.026

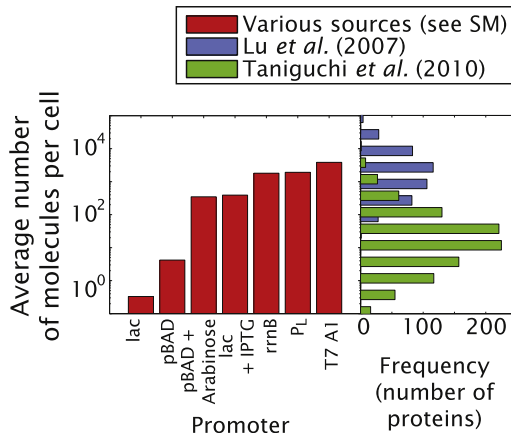


FIGURE 1 Gene expression levels in *E. coli*. (Red) The estimated absolute expression level of several bacterial and viral promoters obtained from the literature are shown (see the Supporting Material and Table S4 for the corresponding references and assumptions made to determine the level of expression). For comparison, the results from two recent cell censuses of *E. coli* are also shown as histograms of the number of proteins (7,69). Note that the range of expression spans greater than six orders of magnitude for a given set of measurements illustrating the wide dynamical range associated with bacterial promoters. The discrepancy between the two cell censuses of *E. coli* is further explored in Fig. S9.

expressed protein and the level of mRNA (7,23,30) and the quantification of protein levels through Western blotting or immunoblots (24,25). Luminescence has an advantage related to its low background levels (31). As an enzymatic reporter, it requires the addition of a substrate to the medium or the encoding of genes that can produce the substrate within the cell. It has been established that a constant luminescence level per cell can be used to quantify the number of cells in culture with a very high dynamic range (26).

In the context of using luminescence as a reporter for gene expression, luciferase has been compared over a limited range with respect to a different reporter, to the number of

plasmids bearing a constitutive promoter expressing luciferase and, recently, to absolute luciferase protein counts (28,32,33). The necessity for providing the cells with a substrate has certainly diminished its use with respect to the widespread fluorescent protein reporters. Additionally, there have been reports of the bioluminescence genes potentially affecting their own expression (34). As such we did not address this technique in this work, though studies similar in spirit to those performed here such as those by Suter et al. (28) would be a useful complement to our own study.

Our aim was to compare enzymatic and fluorescence reporters for the same promoters in a way that spans the large absolute dynamic range found in bacterial and viral promoters. In analogy to previous work (7,35) the main strategy consists in engineering a promoter regulated by Lac repressor into *Escherichia coli* which we subsequently induce to a variety of different levels with constructs harboring either a fluorescent or an enzymatic reporter. The choice of comparing the enzymatic reporter LacZ and the fluorescence reporter EYFP is based mainly on their wide use in the literature. Additionally, the particular choice of EYFP as a fluorescent reporter has been suggested to be the best compromise between the spectral properties of the available fluorescent proteins and of those molecules associated with cellular autofluorescence (36). It is clear, however, that similar systematic experiments will be required to characterize all the above-mentioned alternative schemes for measuring gene expression.

Theoretical models of gene expression predict the fold-change in gene expression (9,11,12). In particular, this fold-change is given as the ratio of the level of expression in the presence of a transcription factor to the level of gene expression in the absence of that same transcription factor. Further, this level of expression is a function of the concentrations and interaction energies of the different molecular players. Contrasting such relative predictions with experimental data

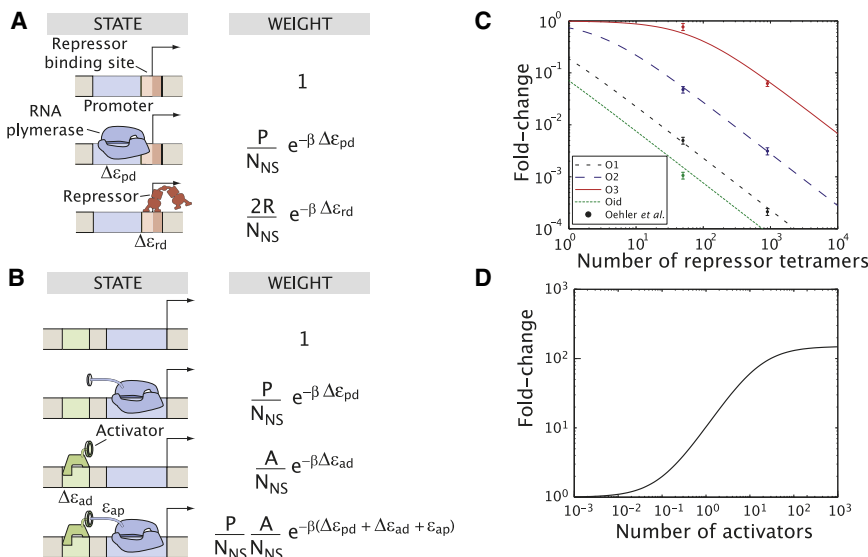


FIGURE 2 Fold-change of different regulatory motifs. The states and weights from the thermodynamic models are shown for the case of (A) simple repression by Lac repressor and (B) simple activation by CRP. The corresponding fold-change in gene expression as a function of transcription factor concentration predicted by the model is shown (C) for repression and (D) for activation. The fold-change values span over several orders of magnitude. Refer to Bintu et al. (12) for a derivation of the respective formulas and their parameters which are characteristic of bacterial promoters. The data for Lac repressor in panel C has been taken from Oehler et al. (70).

relies heavily on the linearity of the response of the reporter being used. As such, we require not only that reporters span a high dynamic range, but that they also be linear over the fold-change range of the theoretical predictions.

To examine the fold-change, we constructed a variety of different realizations of the network in which the binding affinities for Lac repressor are varied in a way that leads to different fold-change levels that differ over several orders of magnitude. Using these schemes, we can explore the presumed linearity of response of the enzymatic assays and their fluorescent reporter protein counterparts.

Though measurements like those described above focus on the mean level of gene expression, recent years have also seen an exciting set of experiments that address issues of cell-to-cell variability. These efforts have relied on exploring different parts of the central dogma by using fluorescent proteins and luminescence, counting mRNA molecules, and measuring single-cell enzymatic activity (2,7,23,28,37–41). Though in this work we address several strategies to quantify mean levels of gene expression, similar work that explores how these different techniques compare in terms of their ability to report cell-to-cell variability will be of great interest as well.

Another measure by which to judge all these different approaches for quantifying gene expression levels is to ask about their ability to report on its dynamics. As discussed in detail in the [Supporting Material](#), issues about differences in dynamics of expression of the different reporters can affect the steady-state level of expression significantly.

In the following sections we show a comparison and absolute calibration of enzymatic and fluorescence reporters both in terms of their absolute mean levels and the fold-change in gene expression. We show that they are interchangeable over several orders of magnitude of expression, but each method has a limited dynamic range either due to limitations of the reporter or to how the reporter acts on the cells. We conclude that they are both complementary and necessary if a systematic characterization of the predictions generated by thermodynamic models spanning over multiple orders of magnitude is to be achieved.

MATERIALS AND METHODS

DNA constructs and strains

Please refer to the [Supporting Material](#) for a detailed description of the construction of all plasmids and strains. All sequences, plasmids, and strains are available upon request.

Growth conditions and gene expression measurements

Growth conditions are detailed in the [Supporting Material](#). Induction and single cell microscopy was performed on the YFP samples as described in the [Supporting Material](#). Our protocol for measuring LacZ activity is basically a slightly modified version of the one described in Miller (18) and Becker et al. (42). Details are given in the [Supporting Material](#).

RESULTS

In the following sections we show a strategy for obtaining an absolute calibration of our two protein reporters. We then compare these reporters side-by-side and determine their ranges of applicability. Finally, we take these experiments one step further by characterizing the fold-change in gene expression measured with both reporters for a simple transcriptional network. This final analysis allows us to determine a range over which thermodynamic models of gene regulation can be tested using this approach.

Absolute calibration of the reporters

Absolute measurements of gene expression are often reported in arbitrary units, especially for fluorescence measurements where the signal depends on the particular details of the microscope used. Such a quantification of fluorescence makes it hard, if not impossible, to compare results between setups and establish unified standards. On the other hand, having a simple way to turn these arbitrary units into an absolute number of molecules would be helpful in both the context of taking the census of cellular proteins (1,7,43) and of characterizing the limits of each reporter.

In the following sections we obtain an absolute calibration for both the enzymatic and fluorescent reporters characterized throughout this work. In turn, this calibration will allow us to set absolute bounds on the interchangeability of these reporters as well as their effectiveness as reporters of the level of gene expression.

Calibration of the absolute number of EYFP molecules

Several previous experiments have performed absolute calibrations of fluorescence levels by looking at a bulk solution of purified fluorophore in buffer (6,7,44,45) or in cell extract (46). These approaches require a known volume of illumination which can be achieved, for example, by performing either confocal microscopy (44,45) or two-photon microscopy (6).

These methods should be considered in light of at least two caveats:

First, they rely on the extinction coefficient of the fluorophore to determine its concentration. However, the solution will be comprised of active and bleached fluorophores. Therefore the effective extinction coefficient of the solution will be an unknown combination of the extinction coefficients for the active and bleached fluorophores.

Second, they are performed outside the cell. Even in the case of cell extract, the local environment the protein sees might be different than that of the unperturbed cellular interior.

Counting fluorescent proteins inside the cell is, however, not straightforward. Because of the fast diffusion time of free fluorescent proteins in the cytoplasm (47) the signal

of individual fluorescent molecules gets blurred over the cell on the timescale of tens of milliseconds. As a result the fluorescence per unit area of a single fluorophore in the cytoplasm becomes comparable to the cell autofluorescence and, hence, not detectable under common continuous illumination conditions. A way to circumvent this is by immobilizing the fluorophore. For example, membrane proteins fused to fluorescent reporters present a much slower diffusion on the membrane than that of proteins in the cytoplasm. Single fluorophores can be then imaged in this way (7,48,49).

Our main approach for calibrating the fluorescence of a single EYFP molecule consists in immobilizing EYFP molecules *in vivo* by fusing them to a transcription factor which is in turn strongly bound to the genomic DNA of *E. coli*. Though this method has the advantage of being *in vivo*, one caveat is that in this case we are not imaging free cytoplasmic EYFP like in the gene expression measurements in this work. The fact that EYFP is fused to another protein that is in turn bound to the DNA could result in a difference of fluorescence.

Puncta of EYFP fused to Tet repressor could be observed inside the cells despite the poor signal/noise ratio of ~ 1.5 . In some cases these puncta could be observed to disappear in a single step, as shown in the traces in Fig. 3, A and B. We associate this with the photobleaching step of a single

EYFP molecule. More often, though, the puncta would correspond to multiple EYFP molecules. These fluorescence traces manifest multiple discrete levels as shown in Fig. 3, C and D. By integrating the fluorescence of the steps over a small area we can obtain the distributions of steps shown in Fig. 3 E. Please refer to the Supporting Material for a detailed discussion of the data analysis process.

We compared the fluorescence per EYFP molecule to the total fluorescence coming from a particular strain, HG105::galK (Δ 25O2+11-YFP, under the same conditions. This strain expresses cytoplasmic EYFP. As a result we estimate the number of EYFP molecules in this strain to be 2600 ± 600 . The reasoning behind choosing a strain where we directly measure the number of EYFP molecules is that all further gene expression measurements with EYFP as a reporter will be measured with respect to this “reference” strain. In this way we can easily estimate the number of EYFP molecules in any other strain we measure. Finally, we also quantified the fluorescence of single purified EYFP molecules and obtained a consistent result within 15%. Please refer to the Supporting Material for details of this *in vitro*, single-molecule fluorescence quantification.

As a sanity check on these results, we estimate the expected number of EYFP molecules. The average number of EYFP molecules in steady state can be approximated by

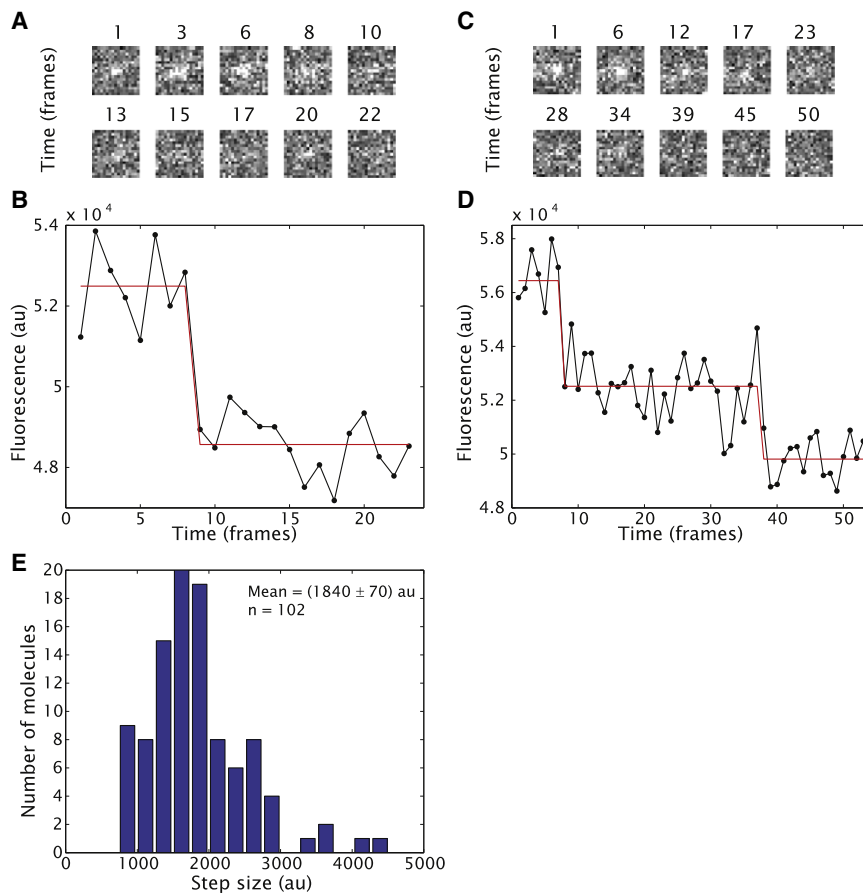


FIGURE 3 Absolute *in vivo* fluorescence calibration. (A) Representative fluorescence snapshots of a single molecule seen as a diffraction limited spot and (B) their corresponding fluorescence traces for a single bleaching event of the EYFP-Tet repressor fusion bound to the genomic DNA. The size of a pixel corresponds to ~ 143 nm. A total of 200 frames with a 250-ms interval were taken for all traces. (C and D) Snapshots and fluorescent traces for multiple bleaching events of the EYFP-Tet repressor fusion. (Red lines) Least-squares fit to a single or multiple step function. (E) Distribution of fluorescence of bleaching steps for the *in vivo* sample.

$$\langle \text{EYFP} \rangle = \frac{\alpha \times b}{\beta}, \quad (1)$$

where α is the mRNA production rate, b is the number of proteins made per mRNA molecule, and β is the protein decay rate (50). Due to the long lifetime of EYFP, the “decay rate” is actually nothing more than the cell doubling time because each cell division effectively halves the number of proteins. For the experiments considered here, we have a cell division time of ~ 1 h. The number of proteins per mRNA has been measured for the *lac* operon to be ~ 20 protein molecules per mRNA molecule (40) and hence we take $b = 20$ proteins/mRNA. This number is within the range of the various protein/mRNA measurements performed previously (7). Finally, the transcription rate for the fully induced *lac* operon has been reported to be between 1 min^{-1} and 20 min^{-1} (51,52). However, the *lacUV5* promoter is $\sim 30\%$ weaker than the fully induced, wild-type *lac* promoter (53) resulting in a range of $\alpha = 0.7\text{--}14 \text{ min}^{-1}$. When combining the decay rate β , the translation rate b and the transcription rate α we obtain an expected number of EYFP molecules of 1200–20,000 per cell, a range comparable to the intracellular number of EYFP molecules we measure.

Calibration of the absolute number of LacZ molecules

A simplified version of the reaction describing the breakdown of ONPG into ONP by β -galactosidase is given by



From this reaction scheme we can derive the rate equation for the production of the yellow compound ONP, which is given in turn by

$$\frac{d[\text{ONP}]}{dt} = k[\text{ONPG}][\text{LacZ}] \quad (3)$$

and a rate equation for the decay in ONPG concentration due to its hydrolysis,

$$\frac{d[\text{ONPG}]}{dt} = -k[\text{ONPG}][\text{LacZ}]. \quad (4)$$

We wish to obtain the concentration of β -galactosidase, $[\text{LacZ}]$, in our reaction to calculate the number of LacZ molecules per cell in the culture that was used to perform it.

If we assume that we have an excess concentration of ONPG and that the time of the reaction is short compared to $1/(k[\text{LacZ}])$, we can neglect its depletion during the reaction. As a result we take $[\text{ONPG}]$ as a constant in Eq. 3. The reaction described by Eq. 3 is the one we perform in the β -galactosidase assay to measure the amount of LacZ molecules per cell in Miller Units (MU). In this assay we monitor the production of ONP over time given by the increase in absorbance at 420 nm of the solution. The standard definition of the Miller Units (18) is

$$\text{MU} = 1000 \frac{\text{OD}_{420} - 1.75 \times \text{OD}_{550}}{t \times v \times \text{OD}_{600}}, \quad (5)$$

where v is the volume of cells used in mL at a cell density given by OD_{600} and t is the reaction time in minutes. These Miller Units were defined such that the fully induced wild-type *lac* operon has an activity of 1000 MU, and its non-induced level would yield ~ 1 MU. We seek to relate these arbitrarily-defined Miller Units defined in Eq. 5 to Eq. 3 to obtain an actual number of LacZ molecules inside the cell.

First, the term $\text{OD}_{420} - 1.75 \times \text{OD}_{550}$ in Eq. 5 is a measure of the amount of ONP, the product of the breakdown of ONPG by β -galactosidase, in the reaction corrected for the cell debris (see Materials and Methods). We relate the absolute concentration of ONP in the reaction to the absorption reading through this term such that $\gamma[\text{ONP}] = \text{OD}_{420} - 1.75 \times \text{OD}_{550}$. From an experimental point of view, the key assumption is that of a linear increase in the amount of ONP over time. Given that at the moment the experiment starts, there is as yet no ONP, we can obtain $d[\text{ONP}]/dt$ simply by taking the accumulated ONP at time t and dividing by this elapsed time; that is,

$$\frac{d[\text{ONP}]}{dt} \approx \frac{[\text{ONP}]}{t}. \quad (6)$$

We also invoke a relation between the OD_{600} reading and the density of cells such that $\text{OD}_{600} \times v \times \delta = N_{\text{cells}}$, where N_{cells} is the number of cells. Finally, we wish to obtain the number of LacZ tetramers present in the reaction N_{LacZ} from this previous equation. This can be done by rewriting the concentration as $[\text{LacZ}] = N_{\text{LacZ}}/V$, where V is now the reaction volume of the standard Miller LacZ assay. If we insert this in our definition of MU, we get

$$\text{MU} = 1000 \gamma k[\text{ONPG}] \delta \frac{1}{V} \frac{N_{\text{LacZ}}}{N_{\text{cells}}}. \quad (7)$$

We determined δ for our strains to be $(8.9 \pm 0.8) \times 10^8/\text{mL}$ through plating and counting of colonies. The relation between ONP absorption at 420 nm and concentration is $\sim 0.0045/\mu\text{M}_{\text{ONP}}$ (54,18). The volume of the reaction in the standard Miller assay is $V = 1.2 \text{ mL}$. However, before the ONP reading the sample gets diluted to $\sim 1.7 \text{ mL}$ by the addition of Na_2CO_3 . Therefore we define $\gamma = 0.0045/\mu\text{M}_{\text{ONP}} \times (1.7 \text{ mL}/1.4 \text{ mL})$. Finally, we need to obtain the turnover rate of LacZ given by k . Wallenfels and Weil (55) report a turnover rate of $138 \times 10^6 \text{M}_{\text{ONP}}/\text{min} \times \text{M}_{\text{LacZ}} \times \text{M}_{\text{ONPG}}$, where we are referring to LacZ tetramers. Similar values have been reported by Kennell and Riezman (52) and Craven et al. (56). Because the initial concentration of ONPG in the reaction is 1.86 mM, we get

$$\text{MU} \times \text{mL} \times \text{min} \times 0.5 \approx \frac{N_Z}{N_{\text{cells}}}. \quad (8)$$

Although this is only a rough estimate because of a lack of error bars associated with the reported values for the specific activity of LacZ, this gives us a direct connection between Miller Units and number of LacZ molecules per cell.

This LacZ calibration that we have just calculated is consistent with previous experimental results on the *lac* operon. For example, the expression level of the repressed operon is ~ 0.6 MU (4). Our calibration suggests that this corresponds to 0.3 LacZ tetramers/cell. Using single molecule techniques, the average number of LacZ tetramers under repressed conditions was estimated to be 1.2 tetramers/cell (40). The internal consistency of these different estimates is encouraging.

Limits of LacZ and YFP as absolute reporters of gene expression

Recall that our aim was to compare enzymatic and fluorescence reporters for the same promoters in a way that spans the large dynamic range found in natural bacterial and viral promoters.

As we have already noted, the level of expression in such bacterial promoters spans over six orders of magnitude as shown in Fig. 1 and Table S4 in the Supporting Material. Our interest was to design an experiment that would permit us to capture a similar dynamical range in a way that would result in a systematic comparison between the enzymatic and fluorescent reporters. To that end, we use an approach based on induction. We use an inducible *lacUV5* promoter with a single binding site for Lac repressor (Oid) located directly downstream from its transcriptional start (see Fig. S7). Two versions of this construct regulating the expression of either the *lacZ* or EYFP genes were created. These constructs were either located on the bacterial chromosome or a low copy plasmid in strains that bear wild-type levels of Lac repressor (*lacI*+), high levels of Lac repressor (*lacI*++), or no Lac repressor (*lacI*-). By growing the different combinations of resulting strains at different concentrations of the inducer IPTG we were able to compare the total EYFP fluorescence and LacZ enzymatic activity per cell. These induction curves are shown in Fig. S8.

In Fig. 4 we present the corresponding expression levels measured using the two reporters over four orders of magnitude. For comparison, these results are juxtaposed with the literature expression levels of some naturally occurring promoters such as those presented in Fig. 1 and Table S4. The blue line corresponds to a fit to a linear model showing that the data is consistent with a linear relation between the two reporters. This observation is consistent with recently published results (7). The slope or conversion factor is $(9.6 \pm 0.7) \times 10^{-5}$ arbitrary fluorescence units/MU. Even if we fit the relation between the two reporters with a more general functional form such as a power law, we find a linear dependence as shown in Fig. S10.

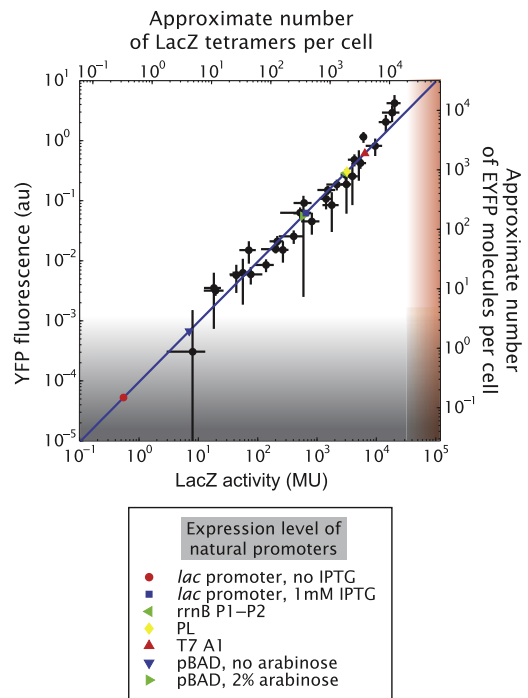


FIGURE 4 Relation between the mean cell fluorescence and β -galactosidase activity. The fluorescence per cell is plotted against the β -galactosidase activity. Each point corresponds to the same construct bearing either EYFP or *lacZ* as a reporter in the same strain background and at the same concentration of IPTG. (Blue line) Linear fit fixing the intercept to zero with a slope of $(9.6 \pm 0.7) \times 10^{-5}$ fluorescence units/MU or an estimated 0.1 YFP molecules/LacZ monomer. (Gray-shaded area) Range of YFP where the fluorescence signal is comparable to the cell autofluorescence (see discussion in the main text and Fig. S12). (Red-shaded area) Range where our assay can detect LacZ expression affecting cell growth (refer to the main text and to Table S1). (Blue line) The expression values of several natural promoters, some of which are also shown in Fig. 1, are plotted.

Although β -galactosidase activity is measured in absolute units, the fluorescence intensity depends strongly on details of the experimental apparatus used for the measurement such as the illumination intensity and transmission of the optical elements. The calibrations mentioned above that convert YFP arbitrary fluorescent units and LacZ Miller Units into a number of molecules allows for our expression levels to be converted into an approximate absolute number of molecules of each reporter as shown by the labeling on the alternative axes in Fig. 4. We estimate the EYFP-LacZ relation to be (roughly) ~ 0.1 EYFP molecules/LacZ monomer. This value seems to be at odds with the fact that they are being expressed from the same promoter. If the transcription rate is the same for both genes, then that leaves some difference at the translation initiation and translation levels (please refer to the Supporting Material for more information on this issue). However, we lack sufficient information to estimate those differences. Alternatively, an underestimation of the number of EYFP molecules inside the cells could be due to issues related to the fluorescence

of the molecule itself such as quenching, misfolding, and the presence of immature fluorophores (57,58).

Fluorescence measurements are fundamentally limited for low levels of gene expression. When the fluorescence signal becomes comparable to the autofluorescence level (<10 molecules/cell) the determination of the level of gene expression has a high associated error. In contrast to free cytoplasmic fluorescent proteins, this limitation is less stringent in the case of fluorescent proteins that are immobilized on the cell membrane (48,57) or DNA by a fusion (this work and (59)). The high error in the determination of low expression levels is reflected in the fluorescence distributions shown in Fig. S12. For the lowest expression levels, the dominant error comes from variations in the autofluorescence. For example, we observe a slight systematic bias toward overestimating the level of autofluorescence. Given the size of the autofluorescence variation, we do not regard the mean value of fluorescence as statistically significant.

This limitation is indicated as a gray-shaded area in Fig. 4. To give a sense of the scale, the expression level of the repressed wild-type *lac* promoter could not be measured with fluorescence unless a more sophisticated technique to visualize single fluorescent proteins is invoked (48). By way of contrast, no significant analogous background was observed in any of our LacZ measurements, showing that this method is more reliable for quantifying very low levels of gene expression. In fact, linearity of the LacZ activity has been reported down to 0.03 MU (4). Of course, this autofluorescence limit is related to the particular choice made of fluorescent protein, growth media, organism, and strain (36). As such it will be important to perform similar comparisons in these different contexts to generalize the results presented here.

When performing a measurement of gene expression using reporters it is important to demonstrate that the presence of the reporter itself is not affecting the state of the cell. We choose the growth rate as an indicator of the cellular state. For all the expression levels shown in Fig. 4 the doubling rate is ~ 1 h regardless of the reporter. However, strain *lacI*⁻ bearing a plasmid with LacZ as a reporter showed a longer doubling time of (74 ± 1) min, which was not the case for the corresponding EYFP strain. These growth rates are shown in Table S1 and the corresponding growth curves are shown in Fig. S11. We confirm previous observations that expression levels above 20,000 LacZ tetramers/cell start affecting the cell significantly (60). Unlike the low end of expression, where EYFP was limited by the autofluorescence, we find that for the high end of expression LacZ becomes limiting not because of signal issues, but because the cell is affected by the fact that high levels of LacZ are being expressed. Interestingly, even some of the stronger promoters such as *rrnB* and the T7 A1 promoter have levels below this threshold.

Although our measurements primarily focused on the use of microscopy to quantify EYFP fluorescence, it is by no

means the only option. An alternative, for example, is to use a plate reader. Although this method is limited to only reporting the mean level of gene expression of a population of cells (it does not provide single-cell information), it is able to produce data in much higher throughput than microscopy. On the other hand, plate readers will be more limited in terms of the minimum level of fluorescence they can quantify reliably. We perform a comparison between fluorescence measurements on the same strains using microscopy and a plate reader in the Supporting Material leading to Fig. S6. We reach the conclusion that they are completely interchangeable, but that the lower limit of detection is now ~ 50 molecules/cell—roughly five-times more than with microscopy.

Limits of LacZ and YFP as reporters of the fold-change in gene expression

The fold-change in gene expression due to regulation by a transcription factor is defined as the level of gene expression in the presence of that molecule divided by the level of gene expression in its absence. In particular, it is the key magnitude predicted by thermodynamic models of transcriptional regulation (11,12). These models can predict fold-changes in gene expression that span over multiple orders of magnitude for both repression (fold-change < 1) and activation (fold-change > 1).

To test these models, it is then necessary to be able to decide which reporter will be the best to assay a particular type of regulatory architecture. For example, in the previous section we found that we can reliably measure EYFP fluorescence down to 10 molecules/cell. If we are dealing with a promoter with a basal expression level of ~ 3000 YFP molecules/cell like the *lacUV5* promoter integrated on the chromosome used in this work, this means that the lowest fold-change we can measure with YFP is $10/3000 \approx 10^{-3}$. On the other hand, the maximum LacZ activity attainable before cell growth starts being compromised is $\sim 20,000$ LacZ tetramers/cell. This means that we can only increase the number of LacZ tetramers beyond the basal level up to this level before the cell senses the presence of these molecules as measured by its growth rate. Because the basal level of our promoter corresponds to 4000 LacZ tetramers/cell, this translates into a maximum measurable fold-change of $20,000/4000 \approx 10^1$ using LacZ as a reporter.

To test part of this assertion about the maximum fold-change in repression, we performed fold-change measurements on constructs bearing the operators O1, O2, or Oid, and the reporters LacZ or EYFP. Fig. 5 shows the fold-change measured using EYFP as a function of the fold-change measured using LacZ for the different single binding site constructs (O1, O2, and Oid) in two different Lac repressor backgrounds: *lacI*⁺ and *lacI*⁺⁺. We see that the fold-change levels measured with both reporters are in good correspondence. As expected, when the

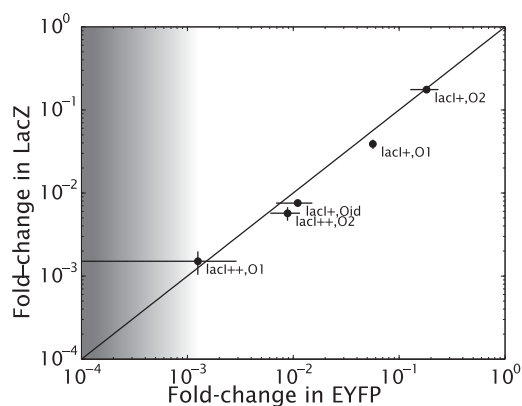


FIGURE 5 Fold-change in gene expression measured by LacZ and EYFP. The fold-change of a construct bearing a single Lac repressor binding site (Oid, O1, and O2) in the *lacI+* and *lacI++* backgrounds is compared when *lacZ* and EYFP are used as reporters. The line has a slope of one. The point in the plot displaying the lowest fold-change corresponds to fluorescence levels that are near the detection limit. This results in the very large error bar shown.

fold-change in gene expression reaches 10^{-3} , the EYFP readings start becoming too noisy to determine the fold-change in gene expression reliably, setting a limit on the range of fold-change that can be measured using EYFP as a reporter.

DISCUSSION

In this work we explored the feasibility of testing theoretical models of gene regulation using two reporters of protein expression: EYFP and LacZ. The calibration between EYFP and LacZ levels shown here is an important methodological prerequisite for testing quantitative models of gene expression. One important outcome is that it makes it possible to compare previously available data, generally taken using LacZ as a reporter, with single cell expression data obtained using EYFP over most of the range of expression of bacterial promoters. EYFP and LacZ are just a subset of the reporters commonly used. Our choice to characterize the relative merits of these two reporters stems from the vast literature available in terms of this enzyme and fluorescent proteins and from the fact that EYFP has been suggested to give the best signal/background with respect to cell autofluorescence (36).

Fluorescent molecules have generally been the method of choice recently because they allow for live imaging of single cells. Our work establishes a clear absolute boundary for these measurements: the autofluorescence level. The intuitive expectation that autofluorescence will contaminate fluorescent gene expression measurements is converted into a concrete and precise numerical criterion. We expect this absolute boundary to be dependent on the particular fluorescent protein used as they can vary widely in their spectral properties and as the autofluorescence is also measurably

different at different wavelengths (36). Additionally, this autofluorescence can be affected by the particular choice of growth media and bacterial strain or cell type (61,62).

Interestingly, the enzymatic activity of LacZ shows no such limitation. However, for high levels of expression, the presence of LacZ affects cell growth in a detectable way before any similar effect from EYFP can be detected. The experimental capacity to use both methods and to switch between one reporter and the other presented in this work makes it possible to obtain the best of both worlds: very low expression levels can be measured accurately in bulk using LacZ in absolute units, whereas slightly higher levels of expression can be measured at the single cell level using fluorescence. Because of fundamental limitations associated with each reporter, we conclude that both techniques need to be used together if the full range of absolute gene expression is to be measured. The outcome of this work has direct consequences on the fold-change in gene expression detectable with each reporter and, in turn, on the range of predictions that these measurements can be contrasted against.

Finally, the work presented here shows a relatively straightforward way to perform similar calibrations for different reporters. In the case of *E. coli*, for example, it is straightforward to swap the fluorescent protein in our transcription factor-fusion with another one to calibrate it at the single molecule level. Fusions of these transcription factors and their corresponding binding arrays have been shown to work in bacteria, yeast and in eukaryotic cells grown in culture (63–66). The inducible promoters can be easily engineered in such a fashion to provide a graded level of expression of the protein of interest. Again, such inducible promoters are commonplace in the context of bacteria, yeast, and cell culture and can span multiple orders of magnitude (67,68). By performing this titration with a reporter whose background is much lower (LacZ, for example) and making use of the absolute calibration of the fluorescent protein, one can then determine the point at which the level of fluorescence becomes nondetectable because of the cell autofluorescence.

Similar experiments where each technique is compared systematically will be necessary to judge the relative merits of each reporter. These comparisons need to be made not only with respect to the mean value of gene expression, but also as reporters of the dynamics of these mean values and of cell-to-cell variability.

SUPPORTING MATERIAL

Additional sections with equations, 12 figures, and four tables are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(11\)00755-7](http://www.biophysj.org/biophysj/supplemental/S0006-3495(11)00755-7).

We are grateful to a number of people for instruction, helpful discussions, and reading the manuscript. In particular, we are grateful to Nicole Becker, Jim Maher, Bob Schleif, Terry Hwa, James Locke, Ron Milo, Tom Kuhlman, Alvaro Sanchez, and Franz Weinert. We thank Roberto Chica for help with the EYFP purification.

This work was supported by National Institutes of Health Pioneer award No. DP1 OD000217 (to H.G.G., H.J.L., and R.P.), National Institutes of Health award No. R01 GM085286 and No. R01 GM085286-01S (to H.G.G., J.Q.B., and R.P.), La Fondation Pierre Gilles de Gennes (to R.P.), and the Caltech Division of Biology (to H.J.L.).

REFERENCES

- Ghaemmaghami, S., W. K. Huh, ..., J. S. Weissman. 2003. Global analysis of protein expression in yeast. *Nature*. 425:737–741.
- Rosenfeld, N., J. W. Young, ..., M. B. Elowitz. 2005. Gene regulation at the single-cell level. *Science*. 307:1962–1965.
- Zaslaver, A., A. Bren, ..., U. Alon. 2006. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods*. 3:623–628.
- Kuhlman, T., Z. Zhang, ..., T. Hwa. 2007. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*. 104:6043–6048.
- Ben-Tabou de-Leon, S., and E. H. Davidson. 2007. Gene regulation: gene control network in development. *Annu. Rev. Biophys. Biomol. Struct.* 36:191–212.
- Gregor, T., D. W. Tank, ..., W. Bialek. 2007. Probing the limits to positional information. *Cell*. 130:153–164.
- Taniguchi, Y., P. J. Choi, ..., X. S. Xie. 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*. 329:533–538.
- Gerland, U., J. D. Moroz, and T. Hwa. 2002. Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc. Natl. Acad. Sci. USA*. 99:12015–12020.
- Buchler, N. E., U. Gerland, and T. Hwa. 2003. On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. USA*. 100:5136–5141.
- Vilar, J. M., and S. Leibler. 2003. DNA looping and physical constraints on transcription regulation. *J. Mol. Biol.* 331:981–989.
- Bintu, L., N. E. Buchler, ..., R. Phillips. 2005. Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* 15:116–124.
- Bintu, L., N. E. Buchler, ..., R. Phillips. 2005. Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* 15:125–135.
- Klumpp, S., Z. Zhang, and T. Hwa. 2009. Growth rate-dependent global effects on gene expression in bacteria. *Cell*. 139:1366–1375.
- Hermesen, R., S. Tans, and P. R. ten Wolde. 2006. Transcriptional regulation by competing transcription factor modules. *PLoS Comput. Biol.* 2:e164.
- Morelli, M. J., P. R. Ten Wolde, and R. J. Allen. 2009. DNA looping provides stability and robustness to the bacteriophage λ -switch. *Proc. Natl. Acad. Sci. USA*. 106:8101–8106.
- Hermesen, R., B. Ursem, and P. R. ten Wolde. 2010. Combinatorial gene regulation using auto-regulation. *PLoS Comput. Biol.* 6:e1000813.
- Sneppen, K., S. Krishna, and S. Semsey. 2010. Simplified models of biological networks. *Annu. Rev. Biophys.* 39:43–59.
- Miller, J. H. 1972. *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Van Dyk, T. K., E. J. DeRose, and G. E. Gonye. 2001. LuxArray, a high-density, genomewide transcription analysis of *Escherichia coli* using bioluminescent reporter strains. *J. Bacteriol.* 183:5496–5505.
- Bongaerts, R. J., I. Hautefort, ..., J. C. Hinton. 2002. Green fluorescent protein as a marker for conditional gene expression in bacterial cells. *Methods Enzymol.* 358:43–66.
- Larson, D. R., R. H. Singer, and D. Zenklusen. 2009. A single molecule view of gene expression. *Trends Cell Biol.* 19:630–637.
- Raj, A., and A. van Oudenaarden. 2009. Single-molecule approaches to stochastic gene expression. *Annu. Rev. Biophys.* 38:255–270.
- Golding, I., J. Paulsson, ..., E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell*. 123:1025–1036.
- Borggrefe, T., R. Davis, ..., R. D. Kornberg. 2001. Quantitation of the RNA polymerase II transcription machinery in yeast. *J. Biol. Chem.* 276:47150–47153.
- Wu, J. Q., C. D. McCormick, and T. D. Pollard. 2008. Chapter 9: Counting proteins in living cells by quantitative fluorescence microscopy with internal standards. *Methods Cell Biol.* 89:253–273.
- Kishony, R., and S. Leibler. 2003. Environmental stresses can alleviate the average deleterious effect of mutations. *J. Biol.* 2:14.
- Cox, 3rd, R. S., M. G. Surette, and M. B. Elowitz. 2007. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* 3:145.
- Suter, D. M., N. Molina, ..., F. Naef. 2011. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 332:472–474.
- Oates, A. C., N. Gorfinkiel, ..., C. P. Heisenberg. 2009. Quantitative approaches in developmental biology. *Nat. Rev. Genet.* 10:517–530.
- Guet, C. C., L. Bruneaux, ..., P. Cluzel. 2008. Minimally invasive determination of mRNA concentration in single living bacteria. *Nucleic Acids Res.* 36:e73.
- Hautefort, I., and J. C. Hinton. 2000. Measurement of bacterial gene expression in vivo. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355:601–611.
- Pfeifer, C. G., and B. B. Finlay. 1995. Monitoring gene expression of *Salmonella* inside mammalian cells: comparison of luciferase and β -galactosidase fusion systems. *J. Microbiol. Methods*. 24:155–164.
- Lutz, R., and H. Bujard. 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1–I2 regulatory elements. *Nucleic Acids Res.* 25:1203–1210.
- Forsberg, A. J., G. D. Pavitt, and C. F. Higgins. 1994. Use of transcriptional fusions to monitor gene expression: a cautionary tale. *J. Bacteriol.* 176:2128–2132.
- Scholz, O., A. Thiel, ..., M. Niederweis. 2000. Quantitative analysis of gene expression with an improved green fluorescent protein. p6. *Eur. J. Biochem.* 267:1565–1570.
- Harms, G. S., L. Cognet, ..., T. Schmidt. 2001. Autofluorescent proteins in single-molecule research: applications to live cell imaging microscopy. *Biophys. J.* 80:2396–2408.
- Elowitz, M. B., A. J. Levine, ..., P. S. Swain. 2002. Stochastic gene expression in a single cell. *Science*. 297:1183–1186.
- Raser, J. M., and E. K. O’Shea. 2004. Control of stochasticity in eukaryotic gene expression. *Science*. 304:1811–1814.
- Zenklusen, D., D. R. Larson, and R. H. Singer. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* 15:1263–1271.
- Cai, L., N. Friedman, and X. S. Xie. 2006. Stochastic protein expression in individual cells at the single molecule level. *Nature*. 440:358–362.
- Larson, D. R., D. Zenklusen, ..., R. H. Singer. 2011. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science*. 332:475–478.
- Becker, N. A., J. D. Kahn, and L. J. Maher, 3rd. 2005. Bacterial repression loops require enhanced DNA flexibility. *J. Mol. Biol.* 349:716–730.
- Wu, J. Q., and T. D. Pollard. 2005. Counting cytokinesis proteins globally and locally in fission yeast. *Science*. 310:310–314.
- Hirschberg, K., C. M. Miller, ..., J. Lippincott-Schwartz. 1998. Kinetic analysis of secretory protein traffic and characterization of Golgi to plasma membrane transport intermediates in living cells. *J. Cell Biol.* 143:1485–1503.
- Piston, D. W., G. H. Patterson, and S. M. Knobel. 1999. Quantitative imaging of the green fluorescent protein (GFP). *Methods Cell Biol.* 58:31–48.
- Sourjik, V., and H. C. Berg. 2002. Binding of the *Escherichia coli* response regulator CheY to its target measured in vivo by fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA*. 99:12669–12674.

47. Elowitz, M. B., M. G. Surette, ..., S. Leibler. 1999. Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* 181:197–203.
48. Yu, J., J. Xiao, ..., X. S. Xie. 2006. Probing gene expression in live cells, one protein molecule at a time. *Science*. 311:1600–1603.
49. Choi, P. J., L. Cai, ..., X. S. Xie. 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*. 322:442–446.
50. Friedman, N., L. Cai, and X. S. Xie. 2006. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* 97:168302.
51. Young, R., and H. Bremer. 1975. Analysis of enzyme induction in bacteria. *Biochem. J.* 152:243–254.
52. Kennell, D., and H. Riezman. 1977. Transcription and translation initiation frequencies of the *Escherichia coli lac* operon. *J. Mol. Biol.* 114:1–21.
53. Deuschle, U., W. Kammerer, ..., H. Bujard. 1986. Promoters of *Escherichia coli*: a hierarchy of in vivo strength indicates alternate structures. *EMBO J.* 5:2987–2994.
54. Lederberg, J. 1950. The β -*d*-galactosidase of *Escherichia coli*, strain K-12. *J. Bacteriol.* 60:381–392.
55. Wallenfels, K., and R. Weil. 1972. Beta-galactosidase. *Enzyme*. 7: 617–663.
56. Craven, Jr., G. R., E. Steers, Jr., and C. B. Anfinsen. 1965. Purification, composition, and molecular weight of the β -galactosidase of *Escherichia coli* K12. *J. Biol. Chem.* 240:2468–2477.
57. Ulbrich, M. H., and E. Y. Isacoff. 2007. Subunit counting in membrane-bound proteins. *Nat. Methods*. 4:319–321.
58. Drepper, T., R. Huber, ..., K. E. Jaeger. 2010. Flavin mononucleotide-based fluorescent reporter proteins outperform green fluorescent protein-like proteins as quantitative in vivo real-time reporters. *Appl. Environ. Microbiol.* 76:5990–5994.
59. Elf, J., G. W. Li, and X. S. Xie. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*. 316:1191–1194.
60. Dong, H., L. Nilsson, and C. G. Kurland. 1995. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J. Bacteriol.* 177:1497–1504.
61. Benson, R. C., R. A. Meyer, ..., G. M. McKhann. 1979. Cellular autofluorescence—is it due to flavins? *J. Histochem. Cytochem.* 27: 44–48.
62. Aubin, J. E. 1979. Autofluorescence of viable cultured mammalian cells. *J. Histochem. Cytochem.* 27:36–43.
63. Robinett, C. C., A. Straight, ..., A. S. Belmont. 1996. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.* 135:1685–1700.
64. Straight, A. F., A. S. Belmont, ..., A. W. Murray. 1996. GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Curr. Biol.* 6:1599–1608.
65. Viollier, P. H., M. Thanbichler, ..., L. Shapiro. 2004. Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *Proc. Natl. Acad. Sci. USA*. 101:9257–9262.
66. Wiggins, P. A., K. C. Cheveralls, ..., J. Kondev. 2010. Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament. *Proc. Natl. Acad. Sci. USA*. 107:4991–4995.
67. Garí, E., L. Piedrafita, ..., E. Herrero. 1997. A set of vectors with a tetracycline-regula promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast*. 13:837–848.
68. Qin, J. Y., L. Zhang, ..., B. T. Lahn. 2010. Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS ONE*. 5:e10611.
69. Lu, P., C. Vogel, ..., E. M. Marcotte. 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* 25:117–124.
70. Oehler, S., M. Amouyal, ..., B. Müller-Hill. 1994. Quality and position of the three lac operators of *E. coli* define efficiency of repression. *EMBO J.* 13:3348–3355.