OPEN ACCESS Freely available online

PLOS ONE

Protein Distributions from a Stochastic Model of the *lac* Operon of *E. coli* with DNA Looping: Analytical solution and comparison with experiments

Krishna Choudhary, Stefan Oehler, Atul Narang*

Department of Biochemical Engineering & Biotechnology, Indian Institute of Technology, Delhi, India

Abstract

Although noisy gene expression is widely accepted, its mechanisms are subjects of debate, stimulated largely by single-molecule experiments. This work is concerned with one such study, in which Choi et al., 2008, obtained real-time data and distributions of Lac permease in *E. coli*. They observed small and large protein bursts in strains with and without auxiliary operators. They also estimated the size and frequency of these bursts, but these were based on a stochastic model of a constitutive promoter. Here, we formulate and solve a stochastic model accounting for the existence of auxiliary operators and DNA loops. We find that DNA loop formation is so fast that small bursts are averaged out, making it impossible to extract their size and frequency from the data. In contrast, we can extract not only the size and frequency of the large bursts, but also the fraction of proteins derived from them. Finally, the proteins follow not the negative binomial distribution, but a mixture of two distributions, which reflect the existence of proteins derived from small and large bursts.

Citation: Choudhary K, Oehler S, Narang A (2014) Protein Distributions from a Stochastic Model of the *lac* Operon of *E. coli* with DNA Looping: Analytical solution and comparison with experiments. PLoS ONE 9(7): e102580. doi:10.1371/journal.pone.0102580

Editor: Mark Isalan, Imperial College London, United Kingdom

Received March 22, 2014; **Accepted** June 20, 2014; **Published** July 23, 2014

Copyright: © 2014 Choudhary et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. Data are from the Choi et al, Science, 322, 442, 2008 study whose authors may be contacted at xie@chemistry.harvard.edu

Funding: AN received grant number SR/SO/BB-79/2010 funded by Department of Science & Technology (www.dst.gov.in). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: anarang@dbeb.iitd.ac.in

Introduction

Data from many independent experiments show that the abundance of any given protein varies among individual cells of isogenic populations growing under identical conditions [1–3]. Early experiments with fluorescent reporters showed that such non-uniformity in protein abundance was due to the inherent stochasticity of gene expression (intrinsic noise) and various forms of cell-to-cell variation (extrinsic noise) [4,5]. The subsequent development of single-molecule techniques has led to deeper insights into the molecular mechanisms generating the noise [6,7]. By measuring the number of mRNAs in single cells, Golding et al. showed that transcription was too bursty to be modeled as a Poisson process [8]. Cai et al. [9] and Yu et al. [10] developed two different methods for measuring the number of proteins in single cells. The real-time data of both studies showed that protein synthesis was bursty, and the burst size was exponentially distributed. Under this condition, the steady state protein distribution follows the Gamma distribution, $p_n = n^{a-1}e^{-n/b}/b^a\Gamma(a)$, where a and b denote the mean burst frequency and burst size [11]. Cai et al. and Yu et al. showed that the Gamma distribution could fit their steady state data, and the values of the mean burst frequency and size derived from the steady state data agreed well with those obtained from real-time measurements.

Armed with these results, Choi et al. [12] attacked a long-standing problem. When non-induced cells of *E. coli* are exposed to small concentrations of the gratuitous inducer TMG, the *lac* operon is induced by stochastic switching of individual cells from the non-induced to the induced state [13]. Choi et al. sought the molecular mechanism of this stochastic switching. To this end, they first quantified the minimum number of LacY molecules required to switch a cell to the induced state, and found this threshold to be 375 molecules. They then suggested a molecular mechanism capable of yielding this threshold by appealing to the known mechanisms of repression and transcription of the *lac* operon. Repression is mediated by the stable DNA loops formed when the Lac repressor is simultaneously bound to the main and auxiliary operators (Fig. 1). Transcription can take place either due to *partial dissociations*, which occur when a repressor trapped in a DNA loop dissociates from the main operator, but not the auxiliary operator; or *complete dissociations*, which occur when the repressor dissociates completely from the DNA. Choi et al. hypothesized that since a partially dissociated repressor remains attached to the DNA, it rapidly rebinds to the main operator, thus limiting the number of transcription events. Although the evidence suggests that no more than one mRNA is made during a partial dissociation, it is conceivable that multiple transcripts are made during a partial dissociation despite its short lifetime, thus leading to a *small transcriptional burst*. In contrast, a completely dissociated repressor takes a relatively long time to find an operator, which results in a *large transcriptional burst*. These large transcriptional bursts can provide enough proteins to cross the threshold for stochastic switching.

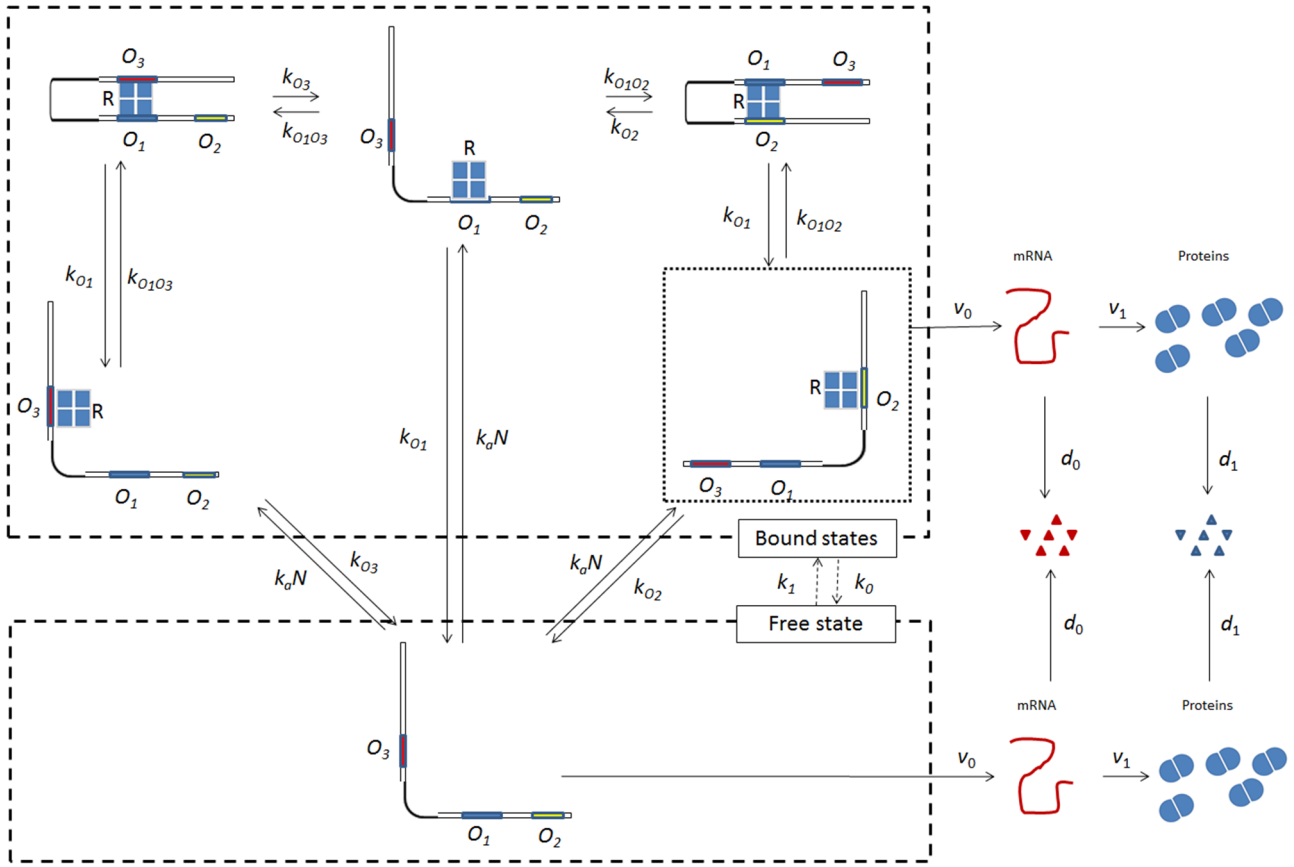

Figure 1. Structure and states of the *lac* operon. The repressor R can bind to any of the three operators, namely the main operator O_1, and the two auxiliary operators O_2, O_3. The repressor-free state is enclosed by the lower dashed box. The repressor-bound states, enclosed by the upper dashed box, consist of the following 5 states (clockwise from the left): the O_3 -bound state $O_3{\cdot}R$, the looped state $O_3{\cdot}R{\cdot}O_1$, the O_1 -bound state $O_1{\cdot}R$, the looped state, $O_1{\cdot}R{\cdot}O_2$, and the O_2-bound state $O_2{\cdot}R$. Transcription occurs only if the operon is in the repressor-free state or the repressor-bound state $O_2{\cdot}R$. Small bursts occur whenever the repressor dissociates from the looped state $O_1{\cdot}R{\cdot}O_2$ to form the O_2 -bound state $O_2{\cdot}R$. Large bursts occur whenever the repressor dissociates from the DNA to form the repressor-free state. Transitions between repressor-free and repressor-bound states occur with propensities k_0 and k_1.
doi:10.1371/journal.pone.0102580.g001

Choi et al. tested the foregoing hypotheses as follows. The statistics of small transcriptional bursts were obtained with strain SX701, a $lacY^-$ strain that exhibits mostly small bursts. To capture the statistics of large bursts, they deleted the auxiliary operators of their $lacY^-$ cells, thus creating strain SX703 which yields only large bursts. The statistics of the small and large bursts were quantified by measuring the steady-state protein distributions for both strains at various inducer concentrations. They then concluded, based on the model of Friedman et al. [11], that if μ,σ^2 denote the mean and variance of a protein distribution obtained with strain SX701, then the Fano factor, $F\equiv\sigma^2/\mu$, and the reciprocal of the noise, $\eta^{-2}\equiv(\mu/\sigma)^2$, represent the size and frequency of the small bursts. Likewise, if $\bar{\mu},\bar{\sigma}^2$ denote the mean and variance for SX703, then $\bar{F}\equiv\bar{\sigma}^2/\bar{\mu}$, $\bar{\eta}^{-2}\equiv(\bar{\mu}/\bar{\sigma})^2$ represent the size and frequency of the large bursts. Analysis of the data for SX703 with this method showed that $\bar{\eta}^{-2}$ did not change with inducer levels, but $\bar{F}$ increased dramatically (Fig. 2a), thus confirming their hypothesis that large bursts can generate enough proteins to trigger stochastic switching. Surprisingly, analysis of the data for SX701 also yielded similar trends (Fig. 2b), but this was attributed to the distortions created by the few cells exhibiting large bursts. Indeed, if the data were filtered by removing the contribution of large bursts, η^{-2} and F did not change much with the inducer concentration (Fig. 2c), leading the authors to conclude that the small burst frequency and size were independent of the inducer level.

Choi et al. also explained these results by appealing to the known states of the *lac* operon (Fig. 1). However, the mathematical model of Friedman et al., which forms the basis of their data analysis, does not account for these complexities — it only considers a constitutive (unregulated) promoter. Consequently, there is no strong support for the assumption that the proteins follow the Gamma distribution; $F,\bar{F}$ represent the size of small and large bursts; and $\eta^{-2},\bar{\eta}^{-2}$ represent the frequency of small and large bursts. The goal of this study is to verify the validity of these assumptions by formulating a stochastic model accounting for the known states of the operon, and deriving analytical expressions for the steady state protein distribution, Fano factor, and noise.

There are stochastic models accounting for the details shown in Fig. 1 [14–16], but these studies do not give analytical expressions for the steady state protein distribution. The literature also contains several stochastic models of gene regulation for which analytical solutions were obtained [11,17–24], but they do not account for the presence of multiple auxiliary operators and DNA

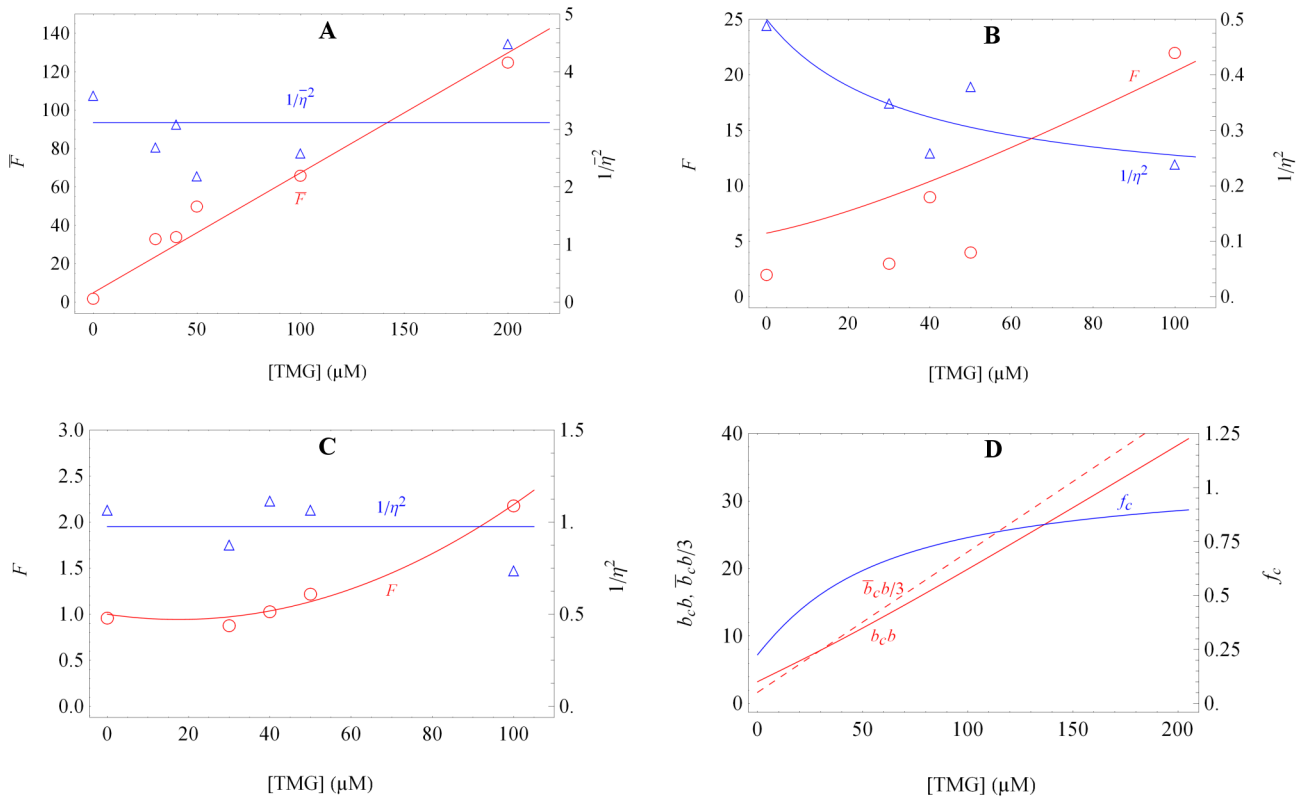

Figure 2. The variation of the Fano factor and the reciprocal of the noise with the inducer level [12]**.** (a) Derived from data for strain SX703, which exhibits only large transcriptional bursts, since it lacks both auxiliary operators. Choi et al. proposed that $\bar{F}$ and $1/\bar{\eta}^2$ represent the size and frequency of large transcriptional bursts. (b) Derived from raw data for strain SX701, which exhibits mostly small transcriptional bursts, since it has both auxiliary operators. Choi et al. did not consider this data on the grounds that the occurrence of large bursts in a few cells distorted the statistics of the small transcriptional bursts. (c) Derived from data for strain SX701 that was filtered by rejecting the data corresponding to the few cells exhibiting large bursts. Choi et al. proposed that this F and $1/\eta^2$ represent the size and frequency of small transcriptional bursts. (d) Mean size of large transcriptional bursts in strain SX701, $b_c b$, (full red curve) and fraction of proteins derived from such bursts, f_c, (full blue curve) estimated from the data in (b). The ordinate of the dashed red line is one-third of the ordinate of the $\bar{F}$ vs. [TMG] line shown in (a), and therefore represents one-third of the (large) transcriptional burst size in strain SX703. The proximity of the full and dashed red lines implies that the mean size of large transcriptional bursts in strain SX701 is approximately one-third of the transcriptional burst size in strain SX703, which is consistent with our model predictions.
doi:10.1371/journal.pone.0102580.g002

looping. Our model fills this gap in the theoretical literature, and its analysis yields deeper insights into the experimental data. Specifically, we show that the size and frequency of small bursts cannot be extracted from the data for strain SX701 because they are averaged out. However, we can extract not only the size and frequency of the large bursts, but also their contribution to total protein synthesis, provided the data is not filtered (Fig. 2d). This result also yields tests for the consistency of the model by providing relationships between the size and frequency of large bursts in strains SX701 and SX703. Finally, we show that neither one of the two strains follow the negative binomial (or Gamma) distribution.

The paper is organized as follows. In the **Analysis** section, we describe the model, derive the master equation, and explain the key approximations used to obtain the steady state protein distribution. In the **Results** section, we perform simulations to check the validity of the analytical expression for the protein distribution, and we derive the expressions for mean and the variance of the distribution. We also show that the mean, variance, and hence, the Fano factor and the reciprocal of the noise, can be expressed in terms of the size and frequency of the transcriptional and translational bursts. In the **Discussion** section, the latter are compared with the assumptions of Choi et al. We also show that negative binomial distributions are obtained only if the size of the large transcriptional bursts is relatively small.

Analysis

The model scheme, shown in Figure 1, is based on the following facts enunciated by Oehler et al. [25,26]. The *lac* operon of *E. coli* contains three operators, namely the main operator O_1, and the two auxiliary operators O_2,O_3, lying downstream and upstream of O_1. The *lac* operon rarely entertains more than one Lac repressor, and this single repressor R can bind to any one of the operators, thus forming the operon states, $O_1{\cdot}R$, $O_2{\cdot}R$, and $O_3{\cdot}R$. Since the tetrameric repressor is a "dimer of dimers," it has a free dimer even after it is bound to one of the operators. This free dimer can bind to one of the remaining two free operators, thus forming a DNA loop. In principle, three looped states are feasible, namely, $O_1{\cdot}R{\cdot}O_2$, $O_1{\cdot}R{\cdot}O_3$, and $O_2{\cdot}R{\cdot}O_3$, but the last one is very unlikely to form. We are therefore led to consider only six feasible states of the operon — the repressor-free state, and the five repressor-bound states, $O_1{\cdot}R$, $O_2{\cdot}R$, $O_3{\cdot}R$, $O_1{\cdot}R{\cdot}O_2$, and $O_1{\cdot}R{\cdot}O_3$. Only three of these six states permit transcriptional activity, namely, the repressor-free state and the repressor-bound states, $O_2{\cdot}R$ and

$O_3 \cdot R$. The first two states permit full transcriptional activity. The last state can be neglected since it permits only 3–5% of the full transcriptional activity.

The model kinetics are based on the following assumptions. All cells have the same number of repressors, N, which is tantamount to neglecting extrinsic noise [4]. Since association of a cytosolic repressor to an operator is diffusion-limited, we assume that a cytosolic repressor has the same propensity, $k_a N$, for association with each of the operators. In contrast, the propensity for dissociation of operator-bound repressor does depend on the identity of the operator, and we denote the propensity for dissociation of O_i -bound repressor by k_{O_i}. Next, we consider the kinetics of looping. The looped state $O_1 \cdot R \cdot O_2$ can be formed from either $O_1 \cdot R$ or $O_2 \cdot R$, but both pathways have the same propensity because they are driven by the same local concentration effect [26]. Thus, we denote the propensity for formation of $O_1 \cdot R \cdot O_2$ from $O_1 \cdot R$ or $O_2 \cdot R$ by the same symbol, $k_{O_1 O_2}$. Similarly, we denote the propensity for formation of $O_1 \cdot R \cdot O_3$ from $O_1 \cdot R$ or $O_3 \cdot R$ by the same symbol, $k_{O_1 O_3}$. Finally, we let v_0,d_0 denote the propensities for mRNA synthesis and degradation, and v_1,d_1 denote the propensities for protein synthesis and dilution.

Equations

We take a master equation approach to describe the system, our state variables being the number of mRNAs, m, the number of proteins, n, and the six states of the operon shown in Figure 1. We let $p^s_{m,n}$ denote the probability of m mRNAs and n proteins when the operon is in state s. Here, $s=f$ when the operon is free, and $s=i$ or $s=ij$ when the operon is repressor-bound, where i,j are integers identifying the operator(s) to which the repressor is bound (e.g., $s=1$ denotes the state $O_1 \cdot R$ and $s=12$ denotes the state $O_1 \cdot R \cdot O_2$). Then the master equations for the kinetic scheme in Figure 1 are

$$\begin{aligned}\frac{dp^f_{m,n}}{dt} &= \left[k_{O_1} p^1_{m,n} + k_{O_2} p^2_{m,n} + k_{O_3} p^3_{m,n} - 3k_a N p^f_{m,n}\right] \\ &+ v_0\left(p^f_{m-1,n} - p^f_{m,n}\right) + v_1 m\left(p^f_{m,n-1} - p^f_{m,n}\right) \\ &+ d_0\left[(m+1)p^f_{m+1,n} - m p^f_{m,n}\right] \\ &+ d_1\left[(n+1)p^f_{m,n+1} - n p^f_{m,n}\right],\end{aligned} \tag{1}$$

$$\begin{aligned}\frac{dp^1_{m,n}}{dt} = &\left[k_{O_2} p^{12}_{m,n} + k_{O_3} p^{13}_{m,n} + k_a N p^f_{m,n} - \left(k_{O_1} + k_{O_1 O_2} + k_{O_1 O_3}\right) p^1_{m,n}\right] \\ &+ v_1 m\left(p^1_{m,n-1} - p^1_{m,n}\right) + d_0\left[(m+1)p^1_{m+1,n} - m p^1_{m,n}\right] \\ &+ d_1\left[(n+1)p^1_{m,n+1} - n p^1_{m,n}\right],\end{aligned} \tag{2}$$

$$\begin{aligned}\frac{dp^2_{m,n}}{dt} &= \left[k_{O_1} p^{12}_{m,n} + k_a N p^f_{m,n} - \left(k_{O_2} + k_{O_1 O_2}\right) p^2_{m,n}\right] \\ &+ v_0\left(p^2_{m-1,n} - p^2_{m,n}\right) + v_1 m\left(p^2_{m,n-1} - p^2_{m,n}\right) \\ &+ d_0\left[(m+1)p^2_{m+1,n} - m p^2_{m,n}\right] \\ &+ d_1\left[(n+1)p^2_{m,n+1} - n p^2_{m,n}\right],\end{aligned} \tag{3}$$

$$\begin{aligned}\frac{dp^3_{m,n}}{dt} &= \left[k_{O_1} p^{13}_{m,n} + k_a N p^f_{m,n} - \left(k_{O_3} + k_{O_1 O_3}\right) p^3_{m,n}\right] \\ &+ v_1 m\left(p^3_{m,n-1} - p^3_{m,n}\right) + d_0\left[(m+1)p^3_{m+1,n} - m p^3_{m,n}\right] \\ &+ d_1\left[(n+1)p^3_{m,n+1} - n p^3_{m,n}\right],\end{aligned} \tag{4}$$

$$\begin{aligned}\frac{dp^{12}_{m,n}}{dt} &= \left[k_{O_1 O_2}\left(p^1_{m,n} + p^2_{m,n}\right) - \left(k_{O_1} + k_{O_2}\right) p^{12}_{m,n}\right] \\ &+ v_1 m\left(p^{12}_{m,n-1} - p^{12}_{m,n}\right) + d_0\left[(m+1)p^{12}_{m+1,n} - m p^{12}_{m,n}\right] \\ &+ d_1\left[(n+1)p^{12}_{m,n+1} - n p^{12}_{m,n}\right],\end{aligned} \tag{5}$$

$$\begin{aligned}\frac{dp^{13}_{m,n}}{dt} &= \left[k_{O_1 O_3}\left(p^1_{m,n} + p^3_{m,n}\right) - (k_{O_1} + k_{O_3}) p^{13}_{m,n}\right] \\ &+ v_1 m\left(p^{13}_{m,n-1} - p^{13}_{m,n}\right) + d_0\left[(m+1)p^{13}_{m+1,n} - m p^{13}_{m,n}\right] \\ &+ d_1\left[(n+1)p^{13}_{m,n+1} - n p^{13}_{m,n}\right].\end{aligned} \tag{6}$$

Our goal is to derive the steady state protein distribution corresponding to these equations.

Parameter values

Table 1 shows the parameter values in the absence of the inducer. The parameters d_0 and d_1 reflect the experimental values measured by Yu et al. [10]. The parameter v_1 was chosen such that the the mean burst size, $b \equiv v_1/d_0$, agreed with the measured value $b=4$, reported by Yu et al. The parameter v_0 was estimated by assuming that the mean burst frequency of fully induced cells, $a \equiv v_0/d_1$, is 600. The rationale for this assumption is as follows. An uninduced cell contains, on average, 0.5 molecules of the tetrameric LacZ [9], and hence, is expected to contain 2 molecules of the monomeric LacY. Since the number of LacY and LacZ molecules increases $\sim$1200-fold in fully induced cells [25], there are 2400 LacY molecules in such cells, i.e., $ab=2400$, which implies that $a=600$. All other parameter values were estimated using the method of Vilar & Leibler [15]. They estimated all the equilibrium constants using the repression data of Oehler et al. [26]. Then, given an experimental estimate of any one parameter,

Table 1. Parameter values in the absence of inducer.

Parameter	Value (in s^{-1})	Parameter	Value (in s^{-1})
d_0	0.011	k_{O_1}	0.0016
d_1	0.0002	k_{O_2}	0.019
v_0	0.12	k_{O_3}	0.73
v_1	0.044	$k_{O_1O_2}$	4
k_aN	0.07	$k_{O_1O_3}$	24

doi:10.1371/journal.pone.0102580.t001

they could find all other parameter values. They took that one parameter to be the dissociation rate constant, k_{O_1}, and assigned to it the value obtained from *in vitro* data [27]. Based on this procedure, the association rate, k_aN, was found to be 0.73 s^{-1}. However, recent *in vivo* measurement show that the association rate for a dimeric repressor is 0.014 s^{-1} [28]. If the dimeric and tetrameric repressor associate at the same rate, and each cell contains 10 repressors [29], the estimated value of k_aN from these measurements is 0.14 s^{-1}. We assumed $k_aN = 0.07$ s^{-1}, and chose k_{O_1}, k_{O_2}, k_{O_3}, $k_{O_1O_2}$, $k_{O_1O_3}$ to ensure consistency with the repression data. As we show later, these parameter values yield good fits of the experimental data.

Since we are also concerned with protein distributions in the presence of the inducer, it is necessary to identify the parameters that change under these conditions. We assume that v_0, d_0, v_1, and d_1 are independent of the inducer level. The propensities for looping, $k_{O_1O_2}$,$k_{O_1O_3}$, are also unlikely to change in the presence of small inducer concentrations because a partially dissociated repressor has too little time to interact with the inducer: In the presence of 10 μM IPTG (considered equivalent to 100 μM TMG), the pseudo-first-order rate constant for repressor-inducer binding is 0.1 s^{-1} [30], which is negligible compared to the looping rate constant of 4 s^{-1}. Thus, the only parameters that can change with the inducer concentration are the association rate, k_aN, and the dissociation rates, k_{O_i}. Based on the analysis of their experimental protein distributions, Choi et al. concluded that the dissociation rates are independent of the inducer concentration, while the association rate decreases with the inducer concentration. We shall also assume that this is the case. This assumption holds only if the concentration of TMG is significantly below 1 mM [31,32], a condition satisfied by all the concentrations used by Choi et al., except possibly the highest concentration of 200 μM.

Model reduction

The determination of the steady state protein distribution corresponding to eqs. (1)–(6) is facilitated by the fact that loop formation and mRNA degradation are relatively fast.

Rapid loop formation. Table 1 shows that in the absence of the inducer, $k_{O_1O_2}$,$k_{O_1O_3}$ are much greater than all other propensities, and as explained above, this persists even in the presence of low inducer concentrations. It follows that the repressor-bound states rapidly equilibrate on the fast time scale $\max\left\{k_{O_1O_2}^{-1},k_{O_1O_3}^{-1}\right\}$, after which there are relatively infrequent transitions between the repressor-free and repressor-bound states. To capture this physical fact, we replace eq. (2) with the equation for the slow variable

$$p_{m,n}^{b} \equiv p_{m,n}^{1}+p_{m,n}^{2}+p_{m,n}^{3}+p_{m,n}^{12}+p_{m,n}^{13}, \tag{7}$$

which represents the probability of m mRNAs and n proteins when the operon is repressor-bound. We then apply the quasi-steady state approximation to the fast variables, $p_{m,n}^{2}$, $p_{m,n}^{3}$, $p_{m,n}^{12}$, $p_{m,n}^{13}$, and find that the probabilities of the equilibrated bound states are given by the relations

$$p_{m,n}^{12} \approx \left(\frac{k_{O_1O_2}/k_{O_2}}{k_{O_1O_2}/k_{O_2}+k_{O_1O_3}/k_{O_3}}\right)p_{m,n}^{b}, \tag{8}$$

$$p_{m,n}^{13} \approx \left(\frac{k_{O_1O_3}/k_{O_3}}{k_{O_1O_2}/k_{O_2}+k_{O_1O_3}/k_{O_3}}\right)p_{m,n}^{b}, \tag{9}$$

$$p_{m,n}^{1} \approx \left(\frac{k_{O_2}}{k_{O_1O_2}}\right)p_{m,n}^{12} \approx \left(\frac{1}{k_{O_1O_2}/k_{O_2}+k_{O_1O_3}/k_{O_3}}\right)p_{m,n}^{b}, \tag{10}$$

$$p_{m,n}^{2} \approx \left(\frac{k_{O_1}}{k_{O_1O_2}}\right)p_{m,n}^{12} \approx \left(\frac{k_{O_1}/k_{O_2}}{k_{O_1O_2}/k_{O_2}+k_{O_1O_3}/k_{O_3}}\right)p_{m,n}^{b}, \tag{11}$$

$$p_{m,n}^{3} \approx \left(\frac{k_{O_1}}{k_{O_1O_3}}\right)p_{m,n}^{13} \approx \left(\frac{k_{O_1}/k_{O_3}}{k_{O_1O_2}/k_{O_2}+k_{O_1O_3}/k_{O_3}}\right)p_{m,n}^{b}, \tag{12}$$

which express the physical fact that after the bound states reach quasi-equilibrium, they obey the principle of detailed balance and are almost always in one of the looped states (Table 2). Moreover, the slow variables follow the equations

Table 2. Magnitudes of important derived parameters in the absence of the inducer.

Parameter	Value	Parameter	Value
$p^{12}_{m,n}/p^{b}_{m,n}$	0.86	k_0	2×10^{-5} s^{-1}
$p^{13}_{m,n}/p^{b}_{m,n}$	0.14	k_1	0.22 s^{-1}
$p^{1}_{m,n}/p^{b}_{m,n}$	4×10^{-3}	λ	3×10^{-4}
$p^{2}_{m,n}/p^{b}_{m,n}$	3×10^{-4}	a	600
$p^{3}_{m,n}/p^{b}_{m,n}$	9×10^{-6}	b	4

doi:10.1371/journal.pone.0102580.t002

$$\begin{aligned}\frac{dp^{f}_{m,n}}{dt} &= \left(k_0p^{b}_{m,n}-k_1p^{f}_{m,n}\right)+v_0\left(p^{f}_{m-1,n}-p^{f}_{m,n}\right)\\ &+v_1m\left(p^{f}_{m,n-1}-p^{f}_{m,n}\right)+d_0\left[(m+1)p^{f}_{m+1,n}-mp^{f}_{m,n}\right]\\ &+d_1\left[(n+1)p^{f}_{m,n+1}-np^{f}_{m,n}\right],\end{aligned} \tag{13}$$

$$\begin{aligned}\frac{dp^{b}_{m,n}}{dt} &= \left(k_1p^{f}_{m,n}-k_0p^{b}_{m,n}\right)+v_0\lambda\left(p^{b}_{m-1,n}-p^{b}_{m,n}\right)\\ &+v_1m\left(p^{b}_{m,n-1}-p^{b}_{m,n}\right)\\ &+d_0\left[(m+1)p^{b}_{m+1,n}-mp^{b}_{m,n}\right]\\ &+d_1\left[(n+1)p^{b}_{m,n+1}-np^{b}_{m,n}\right],\end{aligned} \tag{14}$$

where

$$\lambda\equiv\frac{p^{2}_{m,n}}{p^{b}_{m,n}}\approx\frac{k_{O_1}/k_{O_2}}{k_{O_1O_2}/k_{O_2}+k_{O_1O_3}/k_{O_3}}, \tag{15}$$

$$\begin{aligned}k_0 &\equiv k_{O_1}\left(\frac{p^{1}_{m,n}}{p^{b}_{m,n}}\right)+k_{O_2}\left(\frac{p^{2}_{m,n}}{p^{b}_{m,n}}\right)+k_{O_3}\left(\frac{p^{3}_{m,n}}{p^{b}_{m,n}}\right)\\ &\approx 3k_{O_1}\left(\frac{p^{1}_{m,n}}{p^{b}_{m,n}}\right)\approx 3k_{O_2}\lambda,\end{aligned} \tag{16}$$

$$k_1\equiv 3k_aN. \tag{17}$$

Equations (13)–(14) describe the evolution of the reduced model containing only two operon states — the free and the equilibrated bound states — between which are transitions with propensities, k_0, k_1, which are slow compared to the propensities for looping (Table 2). This is highlighted in Figure 1 by enclosing the free and bound states in dashed boxes, and drawing dashed arrows with labels, k_0 and k_1, to denote the transitions between them. The reduced model is similar to Shahrezaei & Swain's three-stage model for a regulated promoter [22], but there is an important difference. Both operon states are transcriptionally active: The transcription rates in the free and bound states are v_0 and $v_0\lambda$, respectively, where λ is the probability of the $O_2{\cdot}R$ state. Even though $\lambda\ll1$ (Table 2), we cannot neglect the transcription from the bound state, since it captures the effect of the small transcriptional bursts, which can account, as we show later, for almost 80% of the mRNAs synthesized per cell cycle.

Table 2 shows that in the absence of the inducer, $k_0\ll k_1$, so that the free state occurs infrequently and lasts for very short periods of time, i.e., $p^{b}_{m,n}\approx1$. We shall show later that this persists in the presence of the low inducer concentrations ($\leq$ 200 μM TMG) used by Choi et al. Hence, under the experimental conditions of interest, the conditional probabilities in (8)–(12) are essentially equal to the absolute probabilities.

Rapid mRNA degradation. The second approximation appeals to the fact that mRNA degradation is rapid compared to protein dilution, i.e., $d_0\gg d_1$. To apply this approximation, we follow Shahrezaei & Swain [22]. Thus, we begin by rescaling time with respect to the time scale for protein degradation. Letting $\tau=d_1t$ transforms the reduced equations to the form

$$\begin{aligned}\frac{dp^{f}_{m,n}}{d\tau} &= \left(\kappa_0p^{b}_{m,n}-\kappa_1p^{f}_{m,n}\right)+a\left(p^{f}_{m-1,n}-p^{f}_{m,n}\right)\\ &+\gamma bm\left(p^{f}_{m,n-1}-p^{f}_{m,n}\right)+\gamma\left[(m+1)p^{f}_{m+1,n}-mp^{f}_{m,n}\right]\\ &+\left[(n+1)p^{f}_{m,n+1}-np^{f}_{m,n}\right],\end{aligned} \tag{18}$$

$$\begin{aligned}\frac{dp^{b}_{m,n}}{d\tau} &= \left(\kappa_1p^{f}_{m,n}-\kappa_0p^{b}_{m,n}\right)+a\lambda\left(p^{b}_{m-1,n}-p^{b}_{m,n}\right)\\ &+\gamma bm\left(p^{b}_{m,n-1}-p^{b}_{m,n}\right)+\gamma\left[(m+1)p^{b}_{m+1,n}-mp^{b}_{m,n}\right]\\ &+\left[(n+1)p^{b}_{m,n+1}-np^{b}_{m,n}\right],\end{aligned} \tag{19}$$

where $\kappa_0\equiv k_0/d_1$ and $\kappa_1\equiv k_1/d_1$ are the frequencies of transitions between the free and bound operator states, $a\equiv v_0/d_1$ is the frequency of unregulated transcription (in the absence of the repressor), $b\equiv v_1/d_0$ is the translational burst size, i.e., the average number of proteins produced per mRNA, and $\gamma\equiv d_0/d_1\gg1$ is the ratio of protein and mRNA lifetimes. Next, we define the generating functions, $f^{f}(z,z',t)=\sum_{m,n}z'^{m}z^{n}p^{f}_{m,n}$ and

$f^b(z,z',t)=\sum_{m,n} z'^m z^n p^b_{m,n}$, to obtain the partial differential equations

$$\frac{\partial f^f}{\partial \tau}-\gamma[bv(1+u)-u]\frac{\partial f^f}{\partial u}+v\frac{\partial f^f}{\partial v}=\left(\kappa_0 f^b-\kappa_1 f^f\right)+auf^f, \quad (20)$$

$$\frac{\partial f^b}{\partial \tau}-\gamma[bv(1+u)-u]\frac{\partial f^b}{\partial u}+v\frac{\partial f^b}{\partial v}=\left(\kappa_1 f^f-\kappa_0 f^b\right)+a\lambda uf^b, \quad (21)$$

where $u=z'-1$ and $v=z-1$. Since $\gamma\gg 1$, we have the quasi-steady state approximation, $bv(1+u)-u\approx 0$. The steady state protein distribution is therefore given by the equations

$$v\frac{df^f}{dv}=\left(\kappa_0 f^b-\kappa_1 f^f\right)+a\frac{bv}{1-bv}f^f, \quad (22)$$

$$v\frac{df^b}{dv}=\left(\kappa_1 f^f-\kappa_0 f^b\right)+a\lambda\frac{bv}{1-bv}f^b. \quad (23)$$

Since we are interested in the generating function, $f(v)\equiv f^f(v)+f^b(v)$, it is convenient to rewrite these equations as

$$v\frac{df^f}{dv}=\left[a\frac{bv}{1-bv}-(\kappa_0+\kappa_1)\right]f^f-\kappa_0 f, \quad (24)$$

$$v\frac{df}{dv}=a(1-\lambda)\frac{bv}{1-bv}f^f+a\lambda\frac{bv}{1-bv}f, \quad (25)$$

which reduce to the second-order differential equation

$$\frac{d^2f}{dv^2}+\left(\frac{\kappa_0+\kappa_1}{v}+\frac{1+a+a\lambda}{v-1/b}\right)\frac{df}{dv}+\frac{a}{v-1/b}\left(\frac{\kappa_0+\kappa_1\lambda}{v}+\frac{a\lambda}{v-1/b}\right)f=0. \quad (26)$$

We solve this equation with the initial condition, $f(0)=1$, and revert to z as the independent variable, to obtain the following generating function for the steady state protein distribution

$$f(z)=[1-b(z-1)]^{-a\lambda}\cdot{}_2F_1[\alpha,\beta,\kappa_0+\kappa_1;b(z-1)], \quad (27)$$

where ${}_2F_1$ denotes the Gaussian hypergeometric function and

$$\alpha,\beta\equiv\frac{1}{2}\left[a(1-\lambda)+(\kappa_0+\kappa_1)\pm\sqrt{\{a(1-\lambda)+(\kappa_0+\kappa_1)\}^2-4a(1-\lambda)\kappa_0}\right]. \quad (28)$$

As expected, if $\lambda=0$, (27) reduces to the generating function of the negative hypergeometric distribution [22]. In general, however, (27) is the generating function for a mixture of the negative binomial and negative hypergeometric distributions, which reflects, as we show below, the existence of two sub-populations of proteins, namely those derived from small and large transcriptional bursts.

Results

Analytical expressions for the statistics of the protein distributions

Strain with auxiliary operators. The generating function (27) yields the following expressions for the mean, μ, and variance, σ^2, of the protein distribution

$$\mu=a_r b,\ a_r\equiv a\left(\lambda+\frac{\kappa_0}{\kappa_0+\kappa_1}\right), \quad (29)$$

$$\sigma^2=\mu(1+b)+[ab(1-\lambda)]^2\frac{\kappa_0\kappa_1}{(\kappa_0+\kappa_1)^2(\kappa_0+\kappa_1+1)}. \quad (30)$$

Since b represents the mean number of proteins synthesized per mRNA, (29) implies that a_r is the mean frequency of *regulated* transcription. The two terms of a_r also have simple physical interpretations: Since λ and $\kappa_0/(\kappa_0+\kappa_1)$ are the probabilities of the $O_2{\cdot}R$ and free states, $a\lambda$ and $a\kappa_0/(\kappa_0+\kappa_1)$ represent the mean number of mRNAs produced per cell cycle due to small and large transcriptional bursts.

Expanding $f(z)$ about $z=0$ yields the steady state protein distribution

$$p_n=\frac{b^n(1+b)^{-a\lambda}}{n!}\sum_{j=0}^{n}\binom{n}{j}\frac{1}{(1+b)^{n-j}}\frac{\Gamma(a\lambda+n-j)}{\Gamma(a\lambda)}\frac{\Gamma(\alpha+j-1)}{\Gamma(\alpha-1)}\frac{\Gamma(\beta+j-1)}{\Gamma(\beta-1)}\frac{\Gamma(\kappa_0+\kappa_1)}{\Gamma(\kappa_0+\kappa_1+j)}\cdot{}_2F_1(\alpha+j-1,\beta+j-1,\kappa_0+\kappa_1+j;-b). \quad (31)$$

Figure 3 shows that the protein distributions obtained from this expression agree well with those obtained by simulating the full model with the Optimized Direct Method implementation of Gillespie's Stochastic Simulation Algorithm [33] provided in the simulation package StochKit2 [34]. The protein distribution in the absence of the inducer, shown in Fig. 3a, was obtained with the parameter values in Table 1. The distributions in the presence of the inducer were obtained by decreasing the association rate, k_aN, 10-fold (Fig. 3b) and 20-fold (Fig. 3c). Evidently, (31) is a good approximation to the exact solutions in all three cases. We conclude that our approximate solution is accurate down to a 20-fold reduction of the association rate.

Table 2 shows that in the absence of the inducer, $\lambda, k_0/k_1\ll 1$. These relations remain valid at the relatively low inducer levels studied by Choi et al. ($\leq 200\ \mu$M TMG). Indeed, under these conditions, the operon is expressed to no more than 1% of the fully induced level [12], i.e.,

$$\frac{a_r}{a}=\lambda+\frac{\kappa_0}{\kappa_0+\kappa_1}\leq 0.01\Rightarrow\lambda,\ \frac{\kappa_0}{\kappa_0+\kappa_1}\leq 0.01\Rightarrow\lambda,\ \frac{\kappa_0}{\kappa_1}\ll 1, \quad (32)$$

and (29)–(30) can be rewritten as

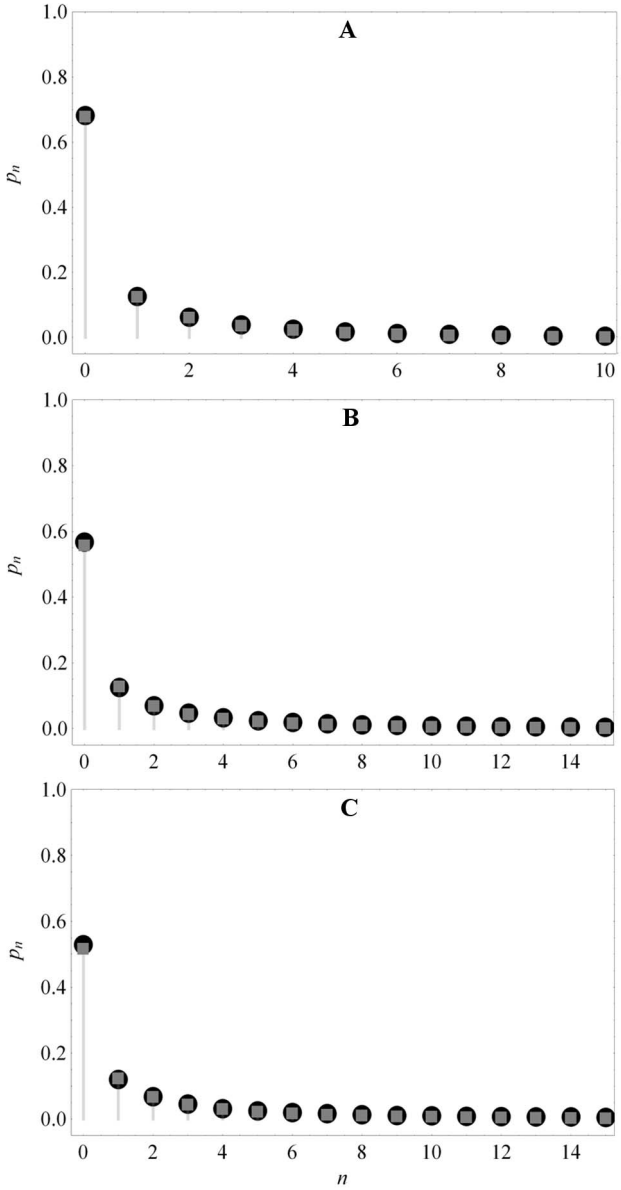

Figure 3. Despite a 20-fold change in the repressor association rate, k_aN, the protein distributions derived from the analytical expression (31) (grey squares) are in good agreement with those obtained from stochastic simulations of the model (black disks). (a) Parameter values in Table 1. (b) k_aN is 1/10th of the value in Table 1; other parameter values as in Table 1. (c) k_aN is 1/20th of the value in Table 1; other parameter values as in Table 1.
doi:10.1371/journal.pone.0102580.g003

$$\mu = a_r b,\ a_r \approx a\left(\lambda + \frac{\kappa_0}{\kappa_1}\right), \tag{33}$$

$$\sigma^2 \approx \mu(1+b) + \frac{a^2 b^2 \kappa_0}{\kappa_1^2}. \tag{34}$$

It is worth noting that due to rapid loop formation, small transcriptional bursts are very bursty (pulsatile). Moreover, under the weakly inducing conditions used in the experiments ($\leq$ 200 μM TMG), k_aN is relatively large, and hence, the large transcriptional bursts are also quite bursty. It follows that under these conditions, (33)–(34) should be expressible in terms of the size and frequency of the small and large transcriptional bursts. We shall show below that this is indeed the case.

Strain without auxiliary operators. In the absence of auxiliary operators, the operon fluctuates between the free and the O_1-bound state, and only the former allows transcription. This is identical to Shahrezaei & Swain's 3-stage model of a regulated promoter [22], and corresponds to the special case, $\lambda = 0$, $k_0 = k_{O_1}$, $k_1 = k_aN$ of our model. It follows that the generating function for the steady state protein distribution is the Gaussian hypergeometric function

$$\bar{f}(z) = {}_2F_1[\bar{\alpha}, \bar{\beta}, \bar{\kappa}_0 + \bar{\kappa}_1; b(z-1)] \tag{35}$$

where

$$\bar{\kappa}_0 \equiv \frac{k_{O_1}}{d_1},\ \bar{\kappa}_1 \equiv \frac{k_aN}{d_1}, \tag{36}$$

and

$$\bar{\alpha}, \bar{\beta} \equiv \frac{1}{2}\left[a + \bar{\kappa}_0 + \bar{\kappa}_1 \pm \sqrt{(a + \bar{\kappa}_0 + \bar{\kappa}_1)^2 - 4a\bar{\kappa}_0}\right]. \tag{37}$$

Moreover, the protein distribution is given by the expression

$$\begin{aligned}\bar{p}_n = {} & \frac{\Gamma(\bar{\alpha}+n)\Gamma(\bar{\beta}+n)\Gamma(\bar{\kappa}_0+\bar{\kappa}_1)}{\Gamma(n+1)\Gamma(\bar{\alpha})\Gamma(\bar{\beta})\Gamma(\bar{\kappa}_0+\bar{\kappa}_1+n)}\left(\frac{b}{1+b}\right)^n\left(1-\frac{b}{1+b}\right)^{\bar{\alpha}} \\ & \times {}_2F_1\left(\bar{\alpha}+n, \bar{\kappa}_0+\bar{\kappa}_1-\bar{\beta}, \bar{\kappa}_0+\bar{\kappa}_1+n; \frac{b}{1+b}\right),\end{aligned} \tag{38}$$

and the mean and variance are

$$\bar{\mu} = \bar{a}_r b,\ \bar{a}_r = a\frac{\bar{\kappa}_0}{\bar{\kappa}_0 + \bar{\kappa}_1}, \tag{39}$$

$$\bar{\sigma}^2 = \mu(1+b) + a^2 b^2 \frac{\bar{\kappa}_0 \bar{\kappa}_1}{(\bar{\kappa}_0 + \bar{\kappa}_1)^2(\bar{\kappa}_0 + \bar{\kappa}_1 + 1)}. \tag{40}$$

At TMG concentrations of $\leq$ 100 μM, which are equivalent to an IPTG concentration of $\leq$ 10 μM, the operon is expressed to no more than 5% of the fully induced level [35]. It follows that under the experimental conditions of interest

$$\frac{\bar{\kappa}_0}{\bar{\kappa}_0 + \bar{\kappa}_1} \leq 0.05 \Rightarrow \frac{\bar{\kappa}_0}{\bar{\kappa}_1} \leq \frac{0.05}{0.95} \approx 0.05, \tag{41}$$

and $\bar{\mu}, \bar{\sigma}^2$ can be approximated by the expressions

$$\bar{\mu} = \bar{a}_r b,\ \bar{a}_r \approx a\frac{\bar{\kappa}_0}{\bar{\kappa}_1}, \tag{42}$$

$$\bar{\sigma}^2 \approx \mu(1+b)+\frac{a^2b^2\bar{\kappa}_0}{\bar{\kappa}_1^2}. \quad (43)$$

Expressing the statistics in terms of the burst size and frequency

Choi et al. assumed that the quantities $F\equiv\sigma^2/\mu$ and $\eta^{-2}\equiv(\sigma/\mu)^{-2}$ represent the size and frequency of small transcriptional bursts, and $\bar{F}\equiv\bar{\sigma}^2/\bar{\mu}$ and $\bar{\eta}^{-2}\equiv(\bar{\sigma}/\bar{\mu})^{-2}$ represent the size and frequency of large transcriptional bursts. To check the validity of these assumptions, we shall express (33)–(34) and (42)–(43) in terms of the size and frequency of the transcriptional bursts. Given these expressions, we can immediately infer the dependence of $F,\eta^{-2},\bar{F},\bar{\eta}^{-2}$ on the size and frequency of the transcriptional bursts, and then compare them to the assumptions made by Choi et al.

Strain with auxiliary operators. To express $\mu=a_rb$ in terms of the size and frequency of the transcriptional bursts, we begin by recalling that a_r consists of two terms, $a\lambda$ and $a\kappa_0/\kappa$, which represent the mean frequency of transcription due to partial and complete dissociations of the repressor, respectively. Since partial dissociations occur when a repressor trapped in the O_1O_2-loop dissociates from O_1, we define the number of the partial dissociations per cell cycle as

$$a_p\equiv\frac{p^{12}_{m,n}k_{O_1}}{d_1}=\frac{p^{2}_{m,n}k_{O_1O_2}}{d_1}\approx\frac{\lambda k_{O_1O_2}}{d_1}, \quad (44)$$

where we have appealed to the detailed balance between the operon states $O_2{\cdot}R$ and $O_1{\cdot}R{\cdot}O_2$. We also define the number of mRNAs synthesized per partial dissociation as

$$b_p\equiv\frac{v_0}{k_{O_1O_2}}, \quad (45)$$

since the time for rebinding of a partially dissociated repressor to O_1 is on the order of $k^{-1}_{O_1O_2}$. It follows from these definitions that

$$a_pb_p\approx a\lambda, \quad (46)$$

i.e., we have successfully expressed the first term of a_r in terms of frequency and mRNA burst size due to partial dissociations. We now proceed to express the second term of a_r in terms of the frequency and mRNA burst size due to complete dissociations. Since complete dissociations occur whenever the operon becomes repressor-free, it is natural to define the number of complete dissociations per cell cycle as

$$a_c\equiv\frac{k_0}{d_1}\approx\frac{3\lambda k_{O_2}}{d_1}. \quad (47)$$

We also define the number of mRNAs synthesized per complete dissociation as

$$b_c\equiv\frac{v_0}{k_1}\approx\frac{v_0}{3k_aN}, \quad (48)$$

because the time for rebinding of a completely dissociated repressor to an operator is on the order of k_1^{-1}. Evidently

$$a_cb_c\approx a\frac{k_0}{k_1}\approx a\lambda\frac{k_{O_2}}{k_aN}, \quad (49)$$

and we conclude that

$$a_r\approx a\left(\lambda+\frac{\kappa_0}{\kappa_1}\right)=a_pb_p+a_cb_c. \quad (50)$$

Hence, (33)–(34) can be rewritten as

$$\mu=a_rb\approx\left(a_pb_p+a_cb_c\right)b, \quad (51)$$

$$\sigma^2\approx\mu(1+b)+a_c(b_cb)^2, \quad (52)$$

which imply that

$$F\equiv\frac{\sigma^2}{\mu}\approx1+b+f_c(b_cb), \quad (53)$$

$$\eta^{-2}\equiv\left(\frac{\mu}{\sigma}\right)^2=\frac{\mu}{F}=\frac{\left(a_pb_p+a_cb_c\right)b}{1+b+f_cb_cb}=\frac{(a_cb_c/f_c)b}{1+b+f_cb_cb}, \quad (54)$$

where

$$f_c\equiv\frac{a_cb_c}{a_pb_p+a_cb_c}=\frac{k_{O_2}/k_aN}{1+k_{O_2}/k_aN}, \quad (55)$$

is the fraction of proteins derived from complete dissociations. It follows from (53) that the total burstiness, F, is entirely due to translational and large transcriptional bursts. Moreover, the burstiness of large transcriptional bursts depends on their intrinsic burstiness, b_cb, suitably weighted by f_c, the fraction of proteins derived from such bursts. Importantly, f_c is completely determined by k_{O_2}/k_aN, the equilibrium constant for dissociation of the repressor from O_2. In the absence of the inducer, this equilibrium constant is 0.25 [25,26], and hence, $f_c=0.2$, i.e., 20% of the proteins are derived from large transcriptional bursts. As the inducer concentration increases, f_c increases because k_aN decreases.

Strain without auxiliary operators. In this case, if we define the number of complete dissociations per cell cycle as

$$\bar{a}_c\equiv\frac{k_{O_1}}{d_1}, \quad (56)$$

and the number of mRNAs synthesized per complete dissociation as

$$\bar{b}_c\equiv\frac{v_0}{k_aN}, \quad (57)$$

the mean frequency of regulated transcription can be rewritten as

$$\bar{a}_r \approx a\frac{\bar{\kappa}_0}{\bar{\kappa}_1} = \bar{a}_c\bar{b}_c. \tag{58}$$

It follows that (42)–(43) can be rewritten as

$$\bar{\mu} = \bar{a}_r b \approx \bar{a}_c\bar{b}_c b, \tag{59}$$

$$\bar{\sigma}^2 = \bar{\mu}(1+b) + \bar{a}_c\bar{b}_c^2 b^2, \tag{60}$$

which imply that

$$\bar{F} \equiv \frac{\bar{\sigma}^2}{\bar{\mu}} = 1 + b + \bar{b}_c b, \tag{61}$$

$$\bar{\eta}^{-2} = \frac{\bar{\mu}}{\bar{F}} = \frac{\bar{a}_r b}{1+b+\bar{b}_c b} \approx \frac{\bar{a}_c\bar{b}_c b}{1+b+\bar{b}_c b}. \tag{62}$$

We are now ready to address questions concerning the physical meaning of the parameters of the distribution and their variation with inducer concentration [12].

Discussion

Interpretation of the protein distribution data

Strain with auxiliary operators. Interpretation of F and η^{-2} derived from filtered data. Choi et al. assumed that F and η^{-2} derived from the filtered data (Fig. 2c) represent the size and frequency of small transcriptional bursts. In terms of our model, these assumptions have the form

$$F \approx b_p b, \tag{63}$$

$$\eta^{-2} \approx a_p. \tag{64}$$

However, (53)–(54) imply that this F and η^{-2}, obtained by eliminating the contribution of the large transcriptional bursts, have a different physical meaning. Indeed, (53) implies that the Fano factor obtained from the filtered data has the form, $F = 1 + b$, which represents the size of the translational, rather than small transcriptional, bursts. Similarly, (54) implies that the reciprocal of the noise derived from the filtered data has the form, $\eta^{-2} = a_p b_p b/(1+b)$, which is proportional to $a_p b_p$, the average number of mRNAs derived from small bursts, rather than the frequency of the small bursts. Since $\eta^{-2} \approx 1$ (Fig. 2c) and $b \approx 4$, our interpretation of the filtered data implies that $a_p b_p \approx 1.25$, which is close to the estimate obtained from the model (Table 3).

Evidently, there is a discrepancy between the assumptions of Choi et al. and the implications of our model. To understand its origin, observe that their assumptions are equivalent to the relations

Table 3. Burst frequency and size in uninduced cells with and without auxiliary operators.

Strain	With auxiliary operators						Without auxiliary operators		
Bursts due to	**Partial dissociations**			**Complete dissociations**			**Complete dissociations**		
Burst properties	a_p	b_p	$a_p b_p$	a_c	b_c	$a_c b_c$	$\bar{a}_c$	$\bar{b}_c$	$\bar{a}_c\bar{b}_c$
Model-based value [1]	6.9	0.03	0.20	0.1	0.55	0.05	8	1.65	13
Data-based value [2]	—	—	1.25	0.2			3		

[1] Calculated from eqs. (44)–(45), (47)–(48), and (56)–(57) with the parameter values in Table 1.
[2] Determined from the data in Figs. 2 a–c by appealing to eqs. (53)–(54) and (61)–(62).
doi:10.1371/journal.pone.0102580.t003

$$\mu = F\eta^{-2} \approx a_p b_p b, \tag{65}$$

$$\sigma^2 = F\mu \approx a_p (b_p b)^2, \tag{66}$$

i.e., they assumed, in effect, that both the mean and the variance are dominated by contributions from small transcriptional bursts. In contrast, (51)–(52) show that small bursts contribute to the mean, but not to the variance. This difference arises because we assumed that looping is so fast that the rapid fluctuations due to partial dissociations are averaged out on the slow time scale of the other processes. This averaging process preserves the contribution of small transcriptional bursts to the mean, but eliminates their contribution to the variance.

The assumption $F \approx b_p b$ appears to be implausible. Indeed, (53) implies that translational bursts contribute the term b to the Fano factor. For the small bursts to make a significant, let alone dominant, contribution to the Fano factor, it is clear that $b_p \sim 1$, i.e., on average, approximately one mRNA must be synthesized per partial dissociation. However, looping is so fast compared to transcription that $b_p \equiv v_0/k_{O_1O_2} \approx 0.03$ in the absence of the inducer (Table 3). Moreover, b_p is unlikely to change even in the presence of the inducer since v_0 and $k_{O_1O_2}$ are constant over the range of inducer concentrations used in the experiments. We conclude that the bursts due to partial dissociations are so small that they cannot be the dominant source of burstiness.

Interpretation of F and η^{-2} derived from raw data. Choi et al. rejected the raw data shown in Fig. 2b since the occurrence of large bursts in a few cells distorted the statistics of the small bursts. We show below that these data are a valuable source of information about the statistics of *large* bursts. Specifically, (53)–(54) predict the observed variation of F and η^{-2} derived from the raw data, and thus provide a method for estimating not only the size and frequency of the large transcriptional bursts, but also the fraction of proteins derived from them. This method is particularly useful because, as we show below, there are simple relationships between the size and frequency of the large bursts in strains SX701 and SX703, but they are *not* identical.

The analysis of the raw data shows that the total burstiness, F, increases with inducer concentration (Fig. 2b). Eq. (53) implies that this is due to the growing burstiness of the large transcriptional bursts: Since both b_c and f_c increase with inducer level, so does $f_c b_c b$. This increase occurs so rapidly that at 100 μM TMG, large trancriptional bursts become the dominant source of burstiness, i.e, $F \approx f_c b_c b$. Indeed, assuming $b \approx 4$, (53) implies $F \approx f_c b_c b$ whenever $F \gg 5$. Inspection of Fig. 2b shows that at 100 μM TMG, $F \approx 25$, and hence, $F \approx f_c b_c b$. We shall show below that at such inducer levels, $f_c \approx 1$ and $F \approx b_c b$.

In contrast to the total burstiness, F, the reciprocal of the total noise, $\eta^{-2} = \mu/F$, decreases with inducer concentration until it reaches a constant value (Fig. 2b). The model suggests that this is because both μ and F increase with inducer level, but F increases faster than μ: Indeed, both b_c and f_c increase with inducer level, and Eq. (54) shows that μ is proportional to the ratio b_c/f_c, whereas F increases with the product $f_c b_c$. The decreasing trend of η^{-2} continues until the inducer levels become so high that large bursts account for all the proteins ($f_c \approx 1$) and burstiness ($F \approx b_c b$). Under these conditions η^{-2} approaches a_c, the frequency of large bursts, which is independent of inducer concentration. Comparison with the data in Fig. 2b then implies that $a_c \approx 0.2$.

Given $a_c \approx 0.2$ and $b \approx 4$, (53)–(54) provide a method for estimating the variation of $b_c b$ and f_c with inducer levels from the raw data for SX701. To see this, it is convenient to rewrite (53)–(54) in the form

$$f_c(b_c b) = F - (1+b), \tag{67}$$

$$\frac{b_c b}{f_c} = \frac{1}{a_c} F\eta^{-2}. \tag{68}$$

Since the variation of F and η^{-2} with the inducer concentration is known (Fig. 2b), we can solve the above equations to obtain $b_c b$ and f_c as a function of the inducer concentration. These calculated profiles, shown in Fig. 2d, agree with the claims above: Both $b_c b$ and f_c increase with the inducer level, and the latter approaches 1 at 100 μM TMG.

Strain without auxiliary operators. Interpretation of $\bar{F}$ and $\bar{\eta}^{-2}$. Choi et al. assumed that the $\bar{F}$ and $\bar{\eta}^{-2}$ shown in Fig. 2a represent the size and frequency of large transcriptional bursts, i.e.,

$$\bar{F} \approx \bar{b}_c b, \tag{69}$$

$$\bar{\eta}^{-2} \approx \bar{a}_c. \tag{70}$$

Our model implies that these relations are valid at all non-zero inducer concentrations used in the experiments. Indeed, since $b \approx 4$, (61)–(62) imply that the above relations are valid whenever $\bar{F} \gg 5$, which is satisfied ($\bar{F} \gtrsim 25$) at all the non-zero inducer concentrations used in the experiments (Fig. 2a). In particular, comparison with the data in Fig. 2a implies that $\bar{a}_c \approx 3$.

Relationships between the statistics of large bursts in the strains with and without auxiliary operators. The model predicts simple relationships between the size and frequency of the large transcriptional bursts in strains SX701 and SX703, which provide tests for checking the consistency of the model. Indeed, it follows from (48) and (57) that $b_c/\bar{b}_c = 1/3$, a relationship that is also mirrored by the data (compare full and dashed lines in Fig. 2d). Similarly, (47) and (56) imply that

$$\frac{a_c}{\bar{a}_c} = 3p^1_{m,n} \approx \frac{3}{k_{O_1O_2}/k_{O_2} + k_{O_1O_3}/k_{O_3}}, \tag{71}$$

a ratio estimated to be 1/80 based on the values in Table 1, which is of the same order of magnitude as the value 1/15, obtained from the experimentally determined values of $a_c \approx 0.2$ and $\bar{a}_c \approx 3$.

Condition for the negative binomial distribution

Choi et al. assumed that the protein distributions of both strains follow the Gamma distribution, the continuous analog of the negative binomial distribution. We have shown above that neither one of the strains follows the negative binomial distribution. Here,

we demonstrate that the distributions can reduce to the negative binomial distribution, but only only if the large burst size is negligibly small, i.e., the association rate k_aN, is much larger than the transcription rate v_0. Under this condition, even the large bursts are averaged out, and they contribute to the mean, but not the variance or the burstiness.

We begin by considering the strain without auxiliary operators. Under the weakly induced conditions used in the experiments, $\bar{\kappa}_0 \ll \bar{\kappa}_1$, and the generating function for the protein distribution is the negative hypergeometric function

$$\bar{f}(z) \approx {}_2F_1\left[\bar{\alpha},\bar{\beta},\bar{\kappa}_1,b(z-1)\right] \equiv \sum_{k=0}^{\infty} \frac{(\bar{\alpha})_k (\bar{\beta})_k}{(\bar{\kappa}_1)_k} \{b(z-1)\}^k, \quad (72)$$

which reduces to the generating function for the negative binomial distribution precisely when $\bar{\alpha} = \bar{\kappa}_1$ or $\bar{\beta} = \bar{\kappa}_1$. Now (37) implies that

$$\bar{\alpha} \approx a + \bar{\kappa}_1 \approx \bar{\kappa}_1 \left(\bar{b}_c + 1\right), \quad (73)$$

$$\bar{\beta} \approx \frac{a\bar{\kappa}_0}{a+\bar{\kappa}_1} \approx \frac{\bar{a}_r}{\bar{b}_c+1}, \; \bar{a}_r = a\frac{\bar{\kappa}_0}{\bar{\kappa}_1} = \bar{a}_c\bar{b}_c. \quad (74)$$

The condition $\bar{\beta} = \bar{\kappa}_1$ can never be satisfied since $\bar{\kappa}_1/\bar{\kappa}_0 \gg 1$. However, $\bar{\alpha} \approx \bar{\kappa}_1$ precisely when $\bar{b}_c \ll 1$, in which case $\bar{\beta} \approx \bar{a}_r$ and

$$\bar{f}(z) \approx \sum_{k=0}^{\infty} (\bar{a}_r)_k \{b(z-1)\}^k = [1-b(z-1)]^{-\bar{a}_r}, \quad (75)$$

which is the generating function for the negative binomial distribution

$$\bar{p}_n = \frac{\Gamma(\bar{a}_r+n)}{\Gamma(n+1)\Gamma(\bar{a}_r)} \left(\frac{b}{1+b}\right)^n \left(1-\frac{b}{1+b}\right)^{\bar{a}_r}. \quad (76)$$

It is worth noting that under this condition

$$\bar{\mu} = \bar{a}_r b, \; \bar{\sigma}^2 = \bar{\mu}(1+b) \Rightarrow \bar{F} = 1+b, \; \eta^{-2} = \bar{a}_r \frac{b}{1+b}, \quad (77)$$

i.e., large transcriptional bursts make no contribution to the burstiness.

A similar argument shows that the generating function for the strain with auxiliary operators reduces to

$$f(z) = [1-b(z-1)]^{-a_r}, \; a_r = a\left(\lambda + \frac{\kappa_0}{\kappa_1}\right), \quad (78)$$

precisely when $b_c \ll 1$. Under this condition, the proteins follow the negative binomial distribution

$$p_n = \frac{\Gamma(a_r+n)}{\Gamma(n+1)\Gamma(a_r)} \left(\frac{b}{1+b}\right)^n \left(1-\frac{b}{1+b}\right)^{a_r}, \quad (79)$$

and

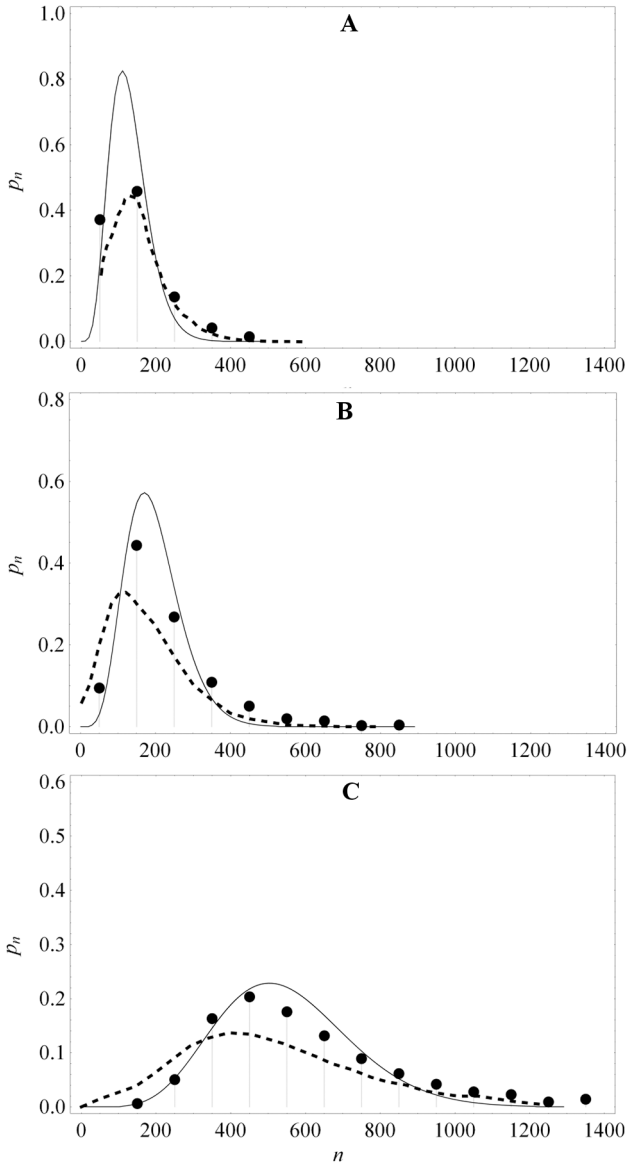

Figure 4. Protein distribution data for strain SX703 (full circles) at various TMG concentrations fitted with the Gamma distribution by Choi et al. (dashed curve) and the negative hypergeometric distribution (full curve). The negative hypergeometric distribution was fitted with the parameter values in Table 1, except k_aN, which was decreased with increasing inducer concentration. (a) Data obtained at 50 μM TMG fitted with $k_aN = 0.028\ \mathrm{s}^{-1}$. (b) Data obtained at 100 μM TMG fitted with $k_aN = 0.018\ \mathrm{s}^{-1}$. (c) Data obtained at 200 μM TMG fitted with $k_aN = 0.0054\ \mathrm{s}^{-1}$.
doi:10.1371/journal.pone.0102580.g004

$$\mu = a_r b, \; \sigma^2 = \mu(1+b) \Rightarrow F = 1+b \approx b, \; \eta^{-2} = a_r \frac{b}{1+b}, \quad (80)$$

i.e., even the large transcriptional bursts do not contribute to the burstiness.

We have shown above that the proteins follow the negative binomial distribution only if the large bursts are, in fact, rather small, and hence, do not contribute to the burstiness. But it follows from the data in Figs. 2a,b that these bursts do contribute significantly to the burstiness of strains SX701 and SX703 — if

this was not true, (77) and (80) imply that the burstiness would be independent of inducer concentration, which contradicts the data. The negative binomial distribution is therefore unlikely to provide good fits to the raw data for both strains, but will fit the filtered data well, since the contribution of large bursts has been eliminated from it. The fits in Choi et al. are consistent with this conclusion. The Gamma distribution fits the filtered data for strain SX701 rather well. However, this is less so for the protein distributions obtained with strain SX703, which exhibits only large bursts. Figure 4 shows that better fits are obtained with the negative hypergeometric distribution (38).

Conclusions

We formulated and solved a stochastic model of *lac* expression accounting for auxiliary operators and DNA looping. Based on a comparison of our expressions for the Fano factor, noise, and protein distribution of strains SX701 (with auxiliary operators) and SX703 (without auxiliary operators) with those proposed by Choi et al., we arrive at the following conclusions:

1. The physical interpretations of the Fano factor $\bar{F}$ and reciprocal noise $\bar{\eta}^{-2}$ for strain SX703 are identical to those proposed by Choi et al., namely $\bar{F}$ and $\bar{\eta}^{-2}$ represent the size and frequency of (large) transcriptional bursts.
2. The physical interpretations of the Fano factor F and reciprocal noise η^{-2} derived from the *filtered* data for SX701 differ from those given by Choi et al., namely F and η^{-2} represent the size and frequency of small transcriptional bursts. Instead, we find that F represents the size of translational bursts, and η^{-2} is proportional to the mean number of mRNAs derived from small transcriptional bursts. Our interpretation is different because we assume that looping is so fast that fluctuations due to small transcriptional bursts are averaged out — small bursts therefore contribute to the mean, but not the burstiness, of the protein distribution. This has two consequences:

(a) The information lost due to the averaging implies that the small burst size and frequency cannot be separately extracted from the data. At best, we can only determine the product of the small burst size and frequency, which represents the mean number of mRNAs derived from small bursts.

(b) The burstiness is entirely due to translational and large transcriptional bursts. In particular, the burst size derived from the filtered data for strain SX701, from which the contribution of the large bursts has been deliberately eliminated, yields the size of translational, rather than small transcriptional, bursts.

3. Choi et al. did not consider the *raw* data for SX701 because large bursts, although rare, contributed significantly to protein synthesis. This is consistent with our model: Even in uninduced cells, 20% of the proteins are derived from large bursts. We find that the raw data contains valuable information about the statistics of large bursts. By analyzing this data with our model, we isolate not only the size and frequency of large bursts, but also the fraction of proteins derived from them. The large burst size obtained in this manner is consistent with another prediction of the model, namely, it is one-third of the (large) burst size in strain SX703. The model also predicts that the fraction of proteins derived from large bursts is completely determined by a measurable quantity, namely the dissociation constant for binding of the repressor to the auxiliary operator O_2.
4. The protein distributions for both strains are not negative binomial: SX703 follows a negative hypergeometric distribution, and SX701 follows a mixture of the negative binomial and negative hypergeometric distributions that reflects the existence of two sub-populations of proteins, namely, those derived from small and large bursts. Negative binomial distributions are attained only if large bursts are insignificant, a condition that holds only if the data are filtered by eliminating the contribution of such bursts.

These results imply that interpretation of the steady state protein distributions depends crucially on the details of the regulatory mechanisms.

Acknowledgments

We are grateful to Sayantari Ghosh for critical comments and help with the fits of the experimental data.

Author Contributions

Conceived and designed the experiments: KC AN. Performed the experiments: KC. Analyzed the data: KC SO AN. Contributed reagents/materials/analysis tools: KC AN. Contributed to the writing of the manuscript: KC SO AN.

References

1. Balzsi G, van Oudenaarden A, Collins JJ (2011) Cellular decision making and biological noise: from microbes to mammals. Cell 144: 910–925.
2. Li GW, Xie XS (2011) Central dogma at the single-molecule level in living cells. Nature 475: 308–315.
3. Snijder B, Pelkmans L (2011) Origins of regulated cell-to-cell variability. Nat Rev Mol Cell Biol 12: 119–125.
4. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297: 1183–1186.
5. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nat Genet 31: 69–73.
6. Raj A, van Oudenaarden A (2009) Single-molecule approaches to stochastic gene expression. Annu Rev Biophys 38: 255–270.
7. Xie XS, Choi PJ, Li GW, Lee NK, Lia G (2008) Single-molecule approach to molecular biology in living bacterial cells. Annu Rev Biophys 37: 417–44.
8. Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. Cell 123: 1025–36.
9. Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. Nature 440: 358–62.
10. Yu J, Xiao J, Ren X, Lao K, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. Science (New York, NY) 311: 1600–3.
11. Friedman N, Cai L, Xie X (2006) Linking stochastic dynamics to population distribution: An analytical framework of gene expression. Phys Rev Lett 97: 168302.
12. Choi PJ, Cai L, Frieda K, Xie XS (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science (New York, NY) 322: 442–6.
13. Novick A, Weiner M (1957) Enzyme induction as an all-or-none phenomenon. Proc Nat Acad Sci USA 43: 553–566.
14. Earnest TM, Roberts E, Assaf M, Dahmen K, Luthey-Schulten Z (2013) DNA looping increases the range of bistability in a stochastic model of the *lac* genetic switch. Phys Biol 10: 026002.
15. Vilar JMG, Leibler S (2003) DNA looping and physical constraints on transcription regulation. J Mol Biol 331: 981–989.
16. Stamatakis M, Mantzaris NV (2009) Comparison of deterministic and stochastic models of the *lac* operon genetic network. Biophys J 96: 887–906.
17. Berg OG (1978) A model for the statistical fluctuations of protein numbers in a microbial population. J Theor Biol 71: 587–603.
18. Kepler TB, Elston TC (2001) Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J 81: 3116–36.

19. Peccoud J, Ycart B (1995) Markovian modeling of gene product synthesis. Theor Popul Biol 48: 222–234.
20. Rigney DR (1979) Stochastic model of constitutive protein levels in growing and dividing bacterial cells. J Theor Biol 76: 453–80.
21. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. PLoS Biol 4: e309.
22. Shahrezaei V, Swain PS (2008) Analytical distributions for stochastic gene expression. Proc Natl Acad Sci U S A 105: 17256–61.
23. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci U S A 99: 12795–800.
24. Thattai M, van Oudenaarden a (2001) Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A 98: 8614–9.
25. Oehler S, Eismann ER, Krmer H, Mller-Hill B (1990) The three operators of the *lac* operon cooperate in repression. EMBO J 9: 973–979.
26. Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, Mller-Hill B (1994) Quality and position of the three *lac* operators of *E. coli* define efficiency of repression. EMBO J 13: 3348–3355.
27. Goeddel DV, Yansura DG, Caruthers MH (1978) How *lac* repressor recognizes *lac* operator. Proc Natl Acad Sci U S A 75: 3578–82.
28. Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, et al. (2012) The Lac repressor displays facilitated diffusion in living cells. Science (New York, NY) 336: 1595–8.
29. Gilbert W, Mller-Hill B (1966) Isolation of the *lac* repressor. Proc Natl Acad Sci U S A 56: 1891–1898.
30. Dunaway M, Manly SP, Matthews KS (1980) Model for lactose repressor protein and its interaction with ligands. Proc Natl Acad Sci U S A 77: 7181–7185.
31. Barkley MD, Riggs AD, Jobe A, Bourgeois S (1975) Interaction of effecting ligands with *lac* repressor and repressor-operator complex. Biochemistry 14: 1700–1712.
32. Dunaway M, Olson JS, Rosenberg JM, Kallai OB, Dickerson RE, et al. (1980) Kinetic studies of inducer binding to *lac* repressor operator complex. J Biol Chem 255: 10115–10119.
33. Cao Y, Li H, Petzold L (2004) Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. J Chem Phys 121: 4059–4067.
34. Sanft KR, Wu S, Roh M, Fu J, Lim RK, et al. (2011) StochKit2: software for discrete stochastic simulation of biochemical systems with events. Bioinformatics (Oxford, England) 27: 2457–8.
35. Oehler S, Alberti S, Mller-Hill B (2006) Induction of the *lac* promoter in the absence of DNA loops and the stoichiometry of induction. Nucleic Acids Res 34: 606–612.