Effects of protein maturation on the noise in gene expression

Guang Qiang Dong and David R. McMillen*
Institute for Optical Sciences and Department of Chemical and Physical Sciences, University of Toronto Mississauga, 3359 Mississauga Rd N, Mississauga, ON Canada L5L 1C6

(Received 15 August 2007; revised manuscript received 29 November 2007; published 13 February 2008)

Fluorescent proteins are frequently used as reporters for gene expression in living cells, either by being expressed in tandem with a protein of interest or through the creation of fusion proteins. The data yielded by the fluorescence output are of considerable interest in efforts to formulate quantitative models of cellular behavior underway in fields such as systems biology and synthetic biology. An often neglected aspect of these proteins, however, is their maturation: Before a fluorescent protein can generate a fluorescent signal, it must mature through a series of steps (folding, cyclization, and oxidation) that may take from many minutes to over a day. The presence of these maturation steps creates a distinction between the observed gene expression profile and the actual profile. We examine this effect through a simplified gene expression model and conclude that fluorescent protein maturation can have significant effects on estimates of both the mean protein levels and the variability in gene expression. The model shows that in many regimes, the observed variability will be increased by the maturation process, but indicates the existence of regimes in which the observed variability will actually be less than the true variability of the target protein. The latter effect arises from a low-pass filtering effect introduced by the chain of maturation reactions. The results suggest that the maturation of fluorescent proteins should be taken into account when using such proteins as quantitative indicators of gene expression levels.

DOI: 10.1103/PhysRevE.77.021908 PACS number(s): 87.15.−v, 87.15.A−, 82.39.−k, 87.15.Ya

I. INTRODUCTION

The recent advent and refinement of sophisticated techniques for probing and manipulating the internal workings of living cells has led to the emergence of two new fields: systems biology [1–5], which aims to assemble the new data sets into models for the behavior of sets of coupled genes, and synthetic biology [6–10], which generally aims to use experimental molecular biology to alter cellular behavior in specific ways, thus exerting control over a cell's internal dynamics. The fields share a focus on quantitative, mathematical modeling of biological processes, and both require correspondingly quantitative experimental data able to provide accurate reporting of these processes as they proceed inside the cell.

One widely studied aspect of gene expression dynamics is the stochasticity or "noise" associated with the process. Processes inside the cell are driven by the sets of highly coupled biochemical reactions, and the resulting dynamics may be approximated by ordinary differential equations, using standard chemical kinetics approaches [11,12]. There is considerable evidence, however, that biological processes, and gene expression in particular, are subject to significant fluctuations and are not purely deterministic [13–18]: A genetically identical population of bacteria growing under identical conditions, for example, will exhibit a substantial range of gene expression levels even from simplified synthetic constructs [19–23]. This variability arises partly from non-negligible fluctuations in the underlying biochemical reactions, called the "intrinsic" noise in the expression of a given gene of interest, and partly from cell-to-cell differences in the background against the gene of interest is expressed, incorporating effects such as variations in numbers of available enzymes, called the "extrinsic" noise [12–14,16–18,21–27].

Green fluorescent proteins (GFPs), a natural fluorophore derived from the jellyfish *Aequorea victoria* and related fluorescent proteins, have found wide application in systems and synthetic biology [1–10,28,29]. By providing a fluorescent signal from a protein that may be incorporated into a living, healthy cell and detected from outside without causing cell death, these proteins offer a key advantage over other protein monitoring such as Western blotting or microarray technologies, wherein cells must be destroyed in the process of the assay, making real-time observation of living cells impossible. Both systems and synthetic biology have made extensive use of plasmid-borne systems expressing fluorescent proteins as reporters in single-celled organisms, either as deliberately simplified testing grounds for their models [12,13,17,20–22,26,27,30] or as the basis for controllers able to function inside growing cells and alter their behavior [6,7,9,19,31–36].

Models formulated of these gene expression systems connect to experimental results by using the fluorescence intensity at a specified wavelength as an observable presumed to be proportional to the number of proteins being expressed in the cell. While there is a reasonable basis for presuming that such a proportionality exists, it is not exact because of the internal dynamics of fluorescent proteins themselves, which can perturb the observed fluorescence away from being a true representation of the state of the proteins in the cell. In particular, the proteins must undergo a process of "maturation" before they become fluorescent; the process involves the folding of the protein, cyclization of a tripeptide motif, and oxidation of the cyclized motif [37–39]. These steps can take from a few minutes (for proteins such as the yellow-fluorescing Venus), to hours (GFP), up to over a day (DsRed,

*mcmillen@utm.utoronto.ca

a redshifted variant of GFP) [28,29,40]. Until the proteins have finished their maturation, they are invisible to fluorescent detection techniques, implying that in any given cell there will be a population of unobserved proteins in addition to those that are detected. This maturation process, while known and acknowledged, is not generally incorporated explicitly into gene expression models, and this has the potential to impact both estimates of the numbers of proteins present [41–43] and estimates of the variability in gene expression [13,17,20–22,26,27,30]. Here, we examine the effect of fluorescent protein maturation on the observed level and variability of gene expression by comparing the number and variability estimates in simple models with and without the maturation steps included, the latter being equivalent to instant maturation. We find that the observed number and variability can in fact differ substantially from the true variability when maturation is included. At low protein production rates, the effect is dominated by the size of fluctuations in the reaction rates and the increased observed variability is attributable to the decreased rate of production of the matured proteins as compared to the true number of proteins being produced. At higher production rates, however, the maturation steps act as a low-pass filter, removing some of the noise in the protein production process and yielding observed variabilities that are lower than the true values.

We begin by formulating a model of the biochemical processes underlying gene expression, neglecting explicit representation of intermediate steps such as the elongation of transcript and peptide chains and the binding and open complex formation of enzymes such as RNA polymerase, and leaving out explicit consideration of the process of cell growth and division. This simplified system is amenable to analytical treatment, and we present the results of solving the master equation for the stochastic processes representing the chemical reactions. We then examine the noise power spectra for this case, again obtaining analytic solutions that illustrate the low-pass filtering effect of the maturation steps. In summary, our results indicate that while fluorescent proteins are an invaluable tool for characterizing and quantifying gene expression, their output must be interpreted with caution, in particular with regard to the observed levels of variability.

II. GENE EXPRESSION MODEL

A. Analysis of steady-state statistics

The molecular basis of gene expression is well established [44,45]; particularly accessible overviews have been presented by Ptashne and Gann [46,47]. Here, we consider a generalized gene expression model as shown in Fig. 1. The process begins with RNA polymerase (RNAP) binding to a DNA strand, forming an open complex, and initiating the process of transcription, wherein a messenger RNA (mRNA) strand is produced, carrying the genetic information. The mRNA is translated into a protein by binding with a ribosome, which decodes the triplet nucleotide codons of the gene into a sequence of amino acids. This model is highly simplified, but it can generate the same statistical behavior as a more elaborate model [48,49]. If the protein in question is a fluorescent reporter, it must undergo three more steps before it can generate fluorescence: folding, cyclization, and oxidation. Each species in the chain is also subject to degradation, represented in Fig. 1 and Eqs. (1)–(10) by processes terminating in the null set symbol.

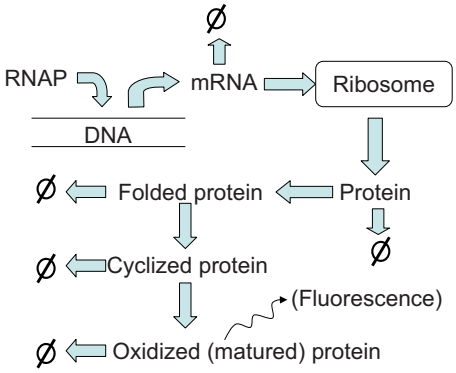

FIG. 1. (Color online) Schematic depiction of the simplified gene expression model considered here. The enzyme RNA polymerase (RNAP) binds to a DNA strand and produces a messenger RNA strand (mRNA), which is then translated by a ribosome into a protein. Fluorescent proteins must then fold, cyclize, and oxidize before they become "mature" and thus able to fluoresce and be detected experimentally. This schematic is converted into the set of biochemical reactions (1)–(10).

The gene expression model in Fig. 1 may be approximated as a system of elementary biochemical reactions, as follows:

$$\mathrm{RNAP} + \mathrm{DNA} \xrightarrow{c_1'} \mathrm{RNAP} + \mathrm{DNA} + \mathrm{mRNA}, \quad (1)$$

$$\mathrm{mRNA} \xrightarrow{c_2} \varnothing, \quad (2)$$

$$\mathrm{mRNA} + \mathrm{ribo} \xrightarrow{c_3'} \mathrm{mRNA} + \mathrm{ribo} + \mathrm{protein}, \quad (3)$$

$$\mathrm{protein} \xrightarrow{c_4} \varnothing, \quad (4)$$

$$\mathrm{protein} \xrightarrow{c_5} \mathrm{protein\text{-}folded}, \quad (5)$$

$$\mathrm{protein\text{-}folded} \xrightarrow{c_4} \varnothing, \quad (6)$$

$$\mathrm{protein\text{-}folded} \xrightarrow{c_6} \mathrm{protein\text{-}cyclized}, \quad (7)$$

$$\mathrm{protein\text{-}cyclized} \xrightarrow{c_4} \varnothing, \quad (8)$$

$$\mathrm{protein\text{-}cyclized} \xrightarrow{c_7} \mathrm{protein\text{-}matured}, \quad (9)$$

$$\mathrm{protein\text{-}matured} \xrightarrow{c_4} \varnothing. \quad (10)$$

Using the common simplifying assumption that our biochemical reactions occur in a well-stirred volume and ne-

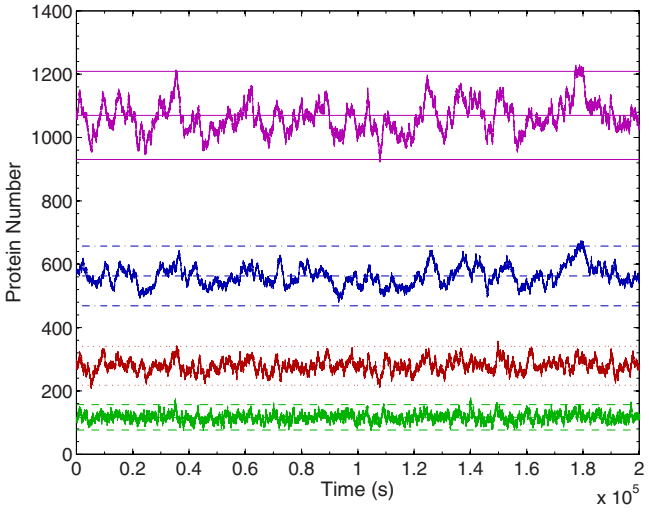

FIG. 2. (Color online) Time series of one realization of the stochastic process defined by biochemical reactions (1)–(10), generated by a Monte Carlo method [50,51], at transcription rate: 0.46 s^{-1}, translation rate: 0.0047 s^{-1}, and maturation time 28 min. From top (greatest number of molecules) to bottom (least number), the traces represent the following species: protein, protein-folded, protein-cyclized, and protein-matured (the fluorescing species visible experimentally); mRNA numbers are not shown. The horizontal lines represent the mean and ± 3 standard deviations for each species, calculated analytically by solving Eqs. (19) and (20).

glecting macromolecular crowding effects [52,53], the spatial information about individual molecules may be ignored. In this approximation, the dynamics of the system is describable using ordinary differential equations when the reaction rates are high, but in lower-rate (generally corresponding with lower number) regimes we must consider fluctuations, treating each reaction as a Poisson process and describing the resulting stochastic process using the chemical master equation [17,50,51,54–56]. Making the further approximation that the numbers of RNAP, DNA, and ribosome (ribo) molecules are constant, the model becomes linear and the master equations may be solved analytically. The state of the system is represented by a five-dimensional vector giving the number of molecules of each species not being kept constant: $\mathbf{n}=(n_1,n_2,n_3,n_4,n_5)$, where the elements are the numbers of molecules of species mRNA, protein, protein-folded, protein-cyclized, and protein-matured, respectively (see Fig. 2). Here, we adopt a symbol system employed by Gadgil *et al.* [57] to represent the four categories of reactions occurring in the system: production from a source $\varnothing \xrightarrow{c_i^s} n_i$, degradation $n_i \xrightarrow{c_i^d} \varnothing$, conversion $n_j \xrightarrow{c_{ij}^{\mathrm{con}}} n_i$, and catalytic production from a source, $n_j \xrightarrow{c_{ij}^{\mathrm{cat}}} n_i + n_j$, where c_i or c_{ij} is the stochastic reaction constant [50]. Using this symbol system, the corresponding reaction-rate equations may be written as

$$\frac{d\mathbf{n}}{dt} = \mathbf{C}^{s\prime} \times \mathbf{1} - \mathbf{C}^{d} \times \mathbf{n} + \mathbf{C}^{\mathrm{cat}} \times \mathbf{n} + \mathbf{C}^{\mathrm{con}} \times \mathbf{n}, \quad (11)$$

where $\mathbf{1}=(1,1,1,1,1)^T$, $\mathbf{C}^s=\mathrm{diag}\{c_i^s\}$, $\mathbf{C}^d=\mathrm{diag}\{c_i^d\}$, $\mathbf{C}_{ij}^{\mathrm{cat}}=c_{ij}^{\mathrm{cat}}$, and

$$\mathbf{C}_{ij}^{\mathrm{con}} = \begin{cases} c_{ij}^{\mathrm{con}}, & i \neq j, \\ -\sum_k c_{kj}^{\mathrm{con}}, & i = j. \end{cases} \quad (12)$$

For our model, $\mathbf{C}^{s\prime}$, $\mathbf{C}^d$, $\mathbf{C}^{\mathrm{cat}}$, and $\mathbf{C}^{\mathrm{con}}$ are

$$\mathbf{C}^{s\prime} = \begin{pmatrix} c_1' n_{\mathrm{RNAP}} n_{\mathrm{DNA}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \equiv \begin{pmatrix} c_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad (13)$$

$$\mathbf{C}^{d} = \begin{pmatrix} c_2 & 0 & 0 & 0 & 0 \\ 0 & c_4 & 0 & 0 & 0 \\ 0 & 0 & c_4 & 0 & 0 \\ 0 & 0 & 0 & c_4 & 0 \\ 0 & 0 & 0 & 0 & c_4 \end{pmatrix}, \quad (14)$$

$$\mathbf{C}^{\mathrm{cat}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ c_3' n_{\mathrm{ribo}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \equiv \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ c_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad (15)$$

and

$$\mathbf{C}^{\mathrm{con}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -c_5 & 0 & 0 & 0 \\ 0 & c_5 & -c_6 & 0 & 0 \\ 0 & 0 & c_6 & -c_7 & 0 \\ 0 & 0 & 0 & c_7 & 0 \end{pmatrix}, \quad (16)$$

where n_{DNA}, n_{RNAP}, and n_{ribo} represent the number of DNA, RNAP, and ribosomes, respectively, and the rate constants are chosen to reflect reasonable values for gene expression in the bacterium *Escherichia coli* [49,58]. Throughout the following, we use maturation time rather than rate to characterize the maturation steps, where the maturation time is the average time required to mature a fraction $(1-1/e)$ of the protein. Fluorescent reporter proteins often have long degradation half-lives, on the order of tens of hours [28,59,60], and we thus make the simplifying assumption that the degradation of the proteins considered here is dominated by dilution due to cell growth, in which the protein number is halved after each cell division (a 24-min cycle in our model *E. coli*); our results are qualitatively unchanged for faster protein degradation rates, but we will not present this case in

detail, here. See the Appendix for a complete list of stochastic reaction constants and other parameters.

The deterministic reaction-rate equations do not capture the stochastic behavior of the system. For this, we represent the probability density associated with each state of the number vector as $P(\mathbf{n})$ and write the master equations for the time evolution of this probability density:

$$\frac{dP(\mathbf{n},t)}{dt} = \sum_{i=1}^{5} \Bigg[\mathbf{C}_{ii}^{s} P(\ldots, n_i - 1, \ldots, t) + \sum_{j=1}^{5} [\mathbf{C}_{ij}^{\mathrm{con}}(n_j+1)P(\ldots, n_i - 1, \ldots, n_j + 1, \ldots, t) + \mathbf{C}_{ij}^{\mathrm{cat}}(n_j P(\ldots, n_i - 1, \ldots, t) - n_j P(\mathbf{n},t)) - \mathbf{C}_{ij}^{d}(n_j P(\ldots, n_i + 1, \ldots, t) - n_j P(\mathbf{n},t))] \Bigg]. \quad (17)$$

Equation (17) is not directly solvable. Exact solutions for the chemical master equation have been found for the case of monomolecular reactions [56], but the presence of the catalytic reaction (3) means that our system does not fall into this category. We obtain a partial solution of Eq. (17) using moment-generating functions. In this approach, we multiply both sides of the master equations by dummy variables $z_1^{n_1}, \ldots, z_5^{n_5}$ and sum over all equations to obtain the generating function $G(\mathbf{z},t) = \Sigma_{n_i=0}^{\infty} z_1^{n_1}, \ldots, z_5^{n_5} P(\mathbf{n})$, where $\mathbf{z} = (z_1, z_2, z_3, z_4, z_5)$ [55]. The time evolution of this function is given by

$$\frac{\partial G(\mathbf{z},t)}{\partial t} = \sum_{i=1}^{5} (z_i - 1) \Bigg(\mathbf{C}_{ii}^{s} \times G(\mathbf{z},t) + \sum_{i=1}^{5} (\mathbf{C}_{ij}^{\mathrm{con}} + \mathbf{C}_{ij}^{\mathrm{cat}} \times z_j - \mathbf{C}_{ij}^{d}) \times \frac{\partial G(\mathbf{z},t)}{\partial z_j} \Bigg). \quad (18)$$

Here we focus on the first and second moments of the distribution, taking the first and second derivatives of the generating function,

$$\mathbf{M}_k(t) = \left. \frac{\partial G(\mathbf{z},t)}{\partial z_j} \right|_{\mathbf{z}=1} = E[n_k(t)] \quad (19)$$

and

$$\mathbf{V}_{ik}(t) = \left. \frac{\partial G(\mathbf{z},t)}{\partial z_i z_k} \right|_{\mathbf{z}=1} = \begin{cases} E[n_i(t)n_k(t)], & i \neq k, \\ E[n_k^2(t)] - E[n_k(t)], & i = k. \end{cases} \quad (20)$$

Equations (19) and (20) allow us to determine the coefficient of variation (η) for each species, defined as the standard deviation divided by the mean. The coefficient of variation of the ith species is

$$\eta_i(t) = \frac{\sqrt{\mathbf{V}_{ii}(t) + \mathbf{M}_i(t) - \mathbf{M}_i^2(t)}}{\mathbf{M}_i(t)}. \quad (21)$$

Taking first and second order derivatives of Eq. (18) and using the quantities defined by Eqs. (19) and (20), we find equations for the time evolution of the first and second moments of the system,

$$\frac{d\mathbf{M}(t)}{dt} = \mathbf{C} \times \mathbf{M}(t) + \mathbf{C}^{s} \quad (22)$$

and

$$\frac{d\mathbf{V}(t)}{dt} = \mathbf{C} \times \mathbf{V}(t) + [\mathbf{C} \times \mathbf{V}(t)]^T + \mathbf{\Psi}(t) + \mathbf{\Psi}^T(t), \quad (23)$$

where $\mathbf{C} \equiv \mathbf{C}^{\mathrm{con}} + \mathbf{C}^{\mathrm{cat}} - \mathbf{C}^{d}$, $\mathbf{M}(t) = (E[n_1(t)], \ldots, E[n_5(t)])$, and $\mathbf{\Psi}_{ij}(t) \equiv (\mathbf{C}_{ii}^{\mathrm{cat}} + \mathbf{C}_{ii}^{s})\mathbf{M}_j(t)$.

Solving for the steady-state solution of the above differential equations yields exact solutions for the first and second moments of the distribution of each species. We first consider an "instant maturation" case in which the protein matures infinitely quickly, so that species protein and protein-matured become identical. In this case, the system can be described by reactions (1)–(4), and the state of the system is fully characterized by the number of mRNA and protein molecules, whose mean (m) and coefficient of variation (η) are

$$\langle n_{\mathrm{mRNA}} \rangle \equiv m_{\mathrm{mRNA}} = \frac{c_1}{c_2}, \quad (24)$$

$$\eta_{\mathrm{mRNA}} = \sqrt{\frac{c_2}{c_1}}, \quad (25)$$

$$\langle n_{\mathrm{protein}} \rangle \equiv m_{\mathrm{protein}} = \frac{c_1 c_3}{c_2 c_4}, \quad (26)$$

and

$$\eta_{\mathrm{protein}} = \sqrt{\left(\frac{1}{m_{\mathrm{protein}}} \right) \left(1 + \frac{c_3}{c_2 + c_4} \right)}. \quad (27)$$

As expected, the mean number of proteins expressed varies directly with the transcription (c_1) and translation (c_3) stochastic reaction constants and varies inversely with the mRNA degradation (c_2) and protein degradation (c_4) constants.

Analytic results also exist for models with finite maturation steps (noninstant maturation), including the subsequent species in the chain of maturation steps: protein-folded, protein-oxidized, and protein-matured. However, the number of terms in the expressions grows rapidly for these species, and they are too cumbersome to reproduce here. The full solutions were obtained using the MAPLE symbolic manipulation package (Maplesoft, Waterloo, Ontario) [61]. To verify the analytical results, we have carried out Monte Carlo realizations of the stochastic process corresponding to the set of biochemical reactions [50], using BIONETS [51], a piece of software specialized for that purpose.

We are mainly interested in the mean and variability of the fully matured protein species, protein-matured, since this is the form whose fluorescence is observable in biological experiments. We also wish to examine the impact of neglect-

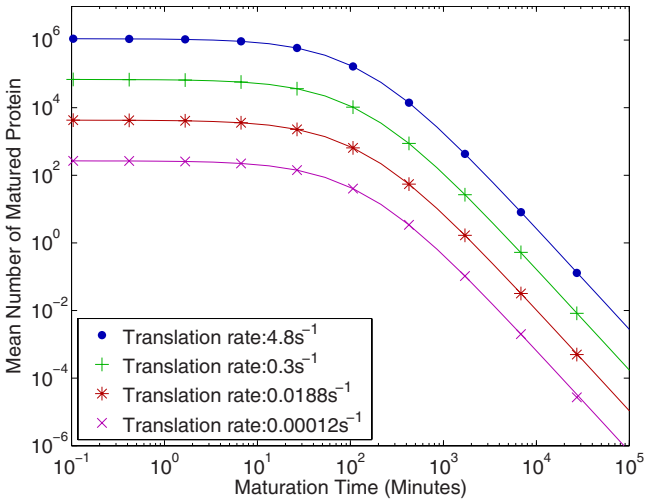

FIG. 3. (Color online) Mean number of matured proteins versus protein maturation time. The translation rate (constant c_3) is varied as shown, for a fixed transcription rate $c_1=0.46$ s^{-1}. The fastest maturation time has $c_5=c_6=4$ s^{-1} and $c_7=4^{-1}$ s^{-1}, and longer maturation times (slower maturation rates) are achieved by successively reducing all three constants by the same factor. The lines represent analytical results from solving the master equations, while the points represent computational simulations run for approximately 5×10^9 reaction update steps.

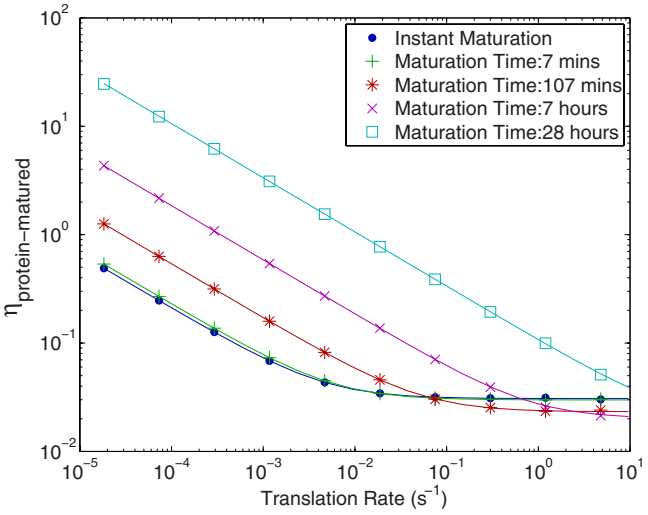

FIG. 4. (Color online) Coefficient of variation (η, the standard deviation over the mean) of the fully matured protein species versus translation rate (c_3), for varying fluorescent protein maturation times, at a fixed transcription rate $c_1=0.46$ s^{-1}. The lines represent analytical solutions from the master equations, and the points are from computational simulations run for approximately 5×10^9 reaction update steps. The "instant maturation" case shows the result when the proteins are taken to fluoresce instantly upon being produced. Note the high-rate regime wherein we see a qualitative change in the behavior, and the instantly maturing protein is more variable than the versions with finite maturation times. Stochastic reaction constants for the maturation processes are of the form $c_5=c_6=4^{2-k}$ s^{-1} and $c_7=4^{1-k}$ s^{-1}, with k taking on the following values: 7 min maturation time, $k=5$; 107 min, $k=7$; 7 h, $k=8$; and 28 h, $k=9$.

ing the maturation effects on estimates of the variability of the protein species, and thus for each case we compare the full model with the instant-maturation model described above. In the following, we vary three sets of stochastic reaction constants: transcription rate (c_1), translation rate (c_3), and the maturation time (set by the reaction constants for the folding, cyclization, and oxidation processes, c_5, c_6, and c_7). To alter the total maturation time, all three maturation constants are varied by the same factor, and as noted above, the maturation time is given as the average time to mature a fraction $(1-1/e)$ of the protein.

Figure 3 shows the average number of matured proteins as a function of maturation time (set by c_5, c_6, and c_7) and translation rate (c_3), keeping the transcription rate (c_1) fixed. As expected, the average number of mature (and thus experimentally observable) proteins decreases with maturation time and increases with translation rate. Figure 4 shows the coefficient of variation of the matured protein species, again as a function of translation rate (c_3) and with varying maturation times. The coefficient of variation decreases as the translation rate increases, and over much of the range the observed variability $\eta_{\text{protein-matured}}$ is increased by longer maturation times. The increase in the coefficient of variation from the instant-maturation to the finite-maturation cases varies strongly with maturation time: at the lowest translation rate shown in Fig. 3, the difference is 8% for a maturation time of 7 min, 160% for a 107 min maturation time, a factor of 9 times for a 7-h maturation, and approximately a factor of 50 times for a 28-h maturation time. Note that the difference becomes less pronounced as the translation rate increases, until in the high-translation-rate regime the behavior is reversed and proteins that mature quickly (or instantly) show higher variability than those that mature slowly. The trends seen in Fig. 4 are the result of a combination of two effects induced by maturation steps: The smaller number of matured proteins than the total number produced, leading to an increase in observed variability, and a noise-filtering effect arising from the maturation steps, leading to a decrease in observed variability (described in Sec. II B, below). Which effect dominates depends on the regime we consider.

To see the effect of changing the total number of matured proteins on the coefficient of variation at a given maturation time, we vary the transcription or translation constants, equivalent to changing the promoter strength or the translational efficiency (e.g., by altering ribosome binding sites), respectively. Figure 5 plots the coefficient of variation against the mean number of matured proteins when the translation constant c_3 is fixed, and we vary the transcription constant c_1, for several different maturation times. Figure 6 shows the coefficient of variation as a function of mean number of matured proteins present when we fix the transcription constant c_1 and vary the translation constant c_3, for the same set of maturation times.

The behavior of the curve of the "instant-maturation" case in Figs. 5 and 6 may be understood by considering Eqs. (26) and (27), giving the mean and coefficient of variation for the instant-maturation case. Taking logarithms and combining Eqs. (26) and (27), we find that a log-log plot of coefficient

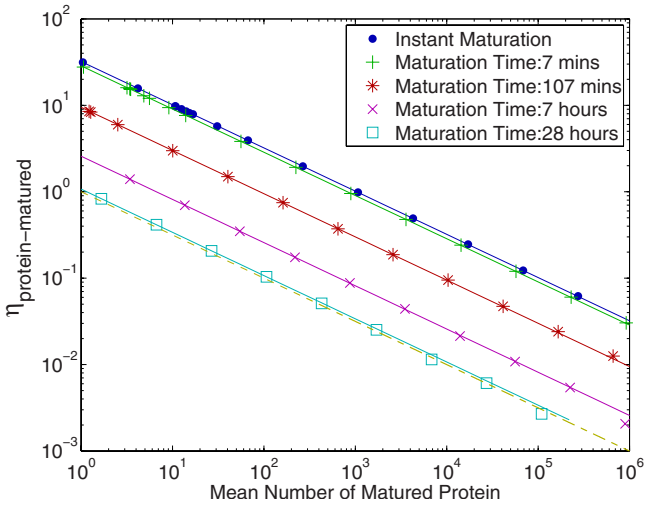

FIG. 5. (Color online) Coefficient of variation (η) of the fully matured protein species versus mean protein number, for varying fluorescent protein maturation times, at a fixed translation rate. Mean protein numbers are varied by altering the transcription rate (c_1), keeping the translation rate fixed at $c_3=4.8\ \text{s}^{-1}$. The dashed line shows $\eta=1/\sqrt{m_{\text{protein-matured}}}$. Reaction constants for the varying maturation times are as given in Fig. 4.

of variation against mean protein number, in the instant-maturation case, has the form

$$\log_{10}\eta_{\text{protein}} = -\tfrac{1}{2}\log_{10} m_{\text{protein}} + f(c_2,c_3,c_4), \qquad (28)$$

where $f(c_2,c_3,c_4)\equiv\frac{1}{2}\log[1+c_3/(c_2+c_4)]$. Thus our log-log plot of the coefficient of variation for the instantly maturing

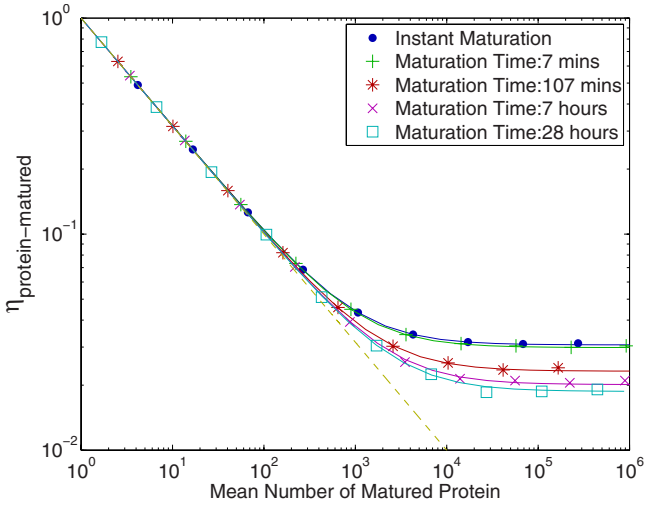

FIG. 6. (Color online) Coefficient of variation (η) of the fully matured protein species versus mean protein number, for varying fluorescent protein maturation times, at a fixed transcription rate. For each protein maturation time, the translation rate (c_3) is varied while keeping the transcription rate fixed at $c_1=0.46\ \text{s}^{-1}$. The dashed line shows $\eta=1/\sqrt{m_{\text{protein-matured}}}$. Note that in the high-number regime, the coefficient of variation deviates from $\eta=1/\sqrt{m_{\text{protein-matured}}}$, and the instantly maturing version shows less variability than the finite-maturation-time versions. Reaction constants for the varying maturation times are as given in Fig. 4.

protein versus its mean has a constant slope of negative one-half, offset by a constant depending only on the translation rate (c_3) and the mRNA and protein degradation rates (c_2 and c_4); note that this slope is the same as that from the classical result that the coefficient of variation varies inversely with the square root of the mean for a Poisson process [55].

In Fig. 5, Eq. (28) yields a perfectly straight line on the plot, since we are varying only the transcription rate c_1 and thus the offset $f(c_2,c_3,c_4)$ is fixed across the plot. The analytic solution for the finite maturation time case does not reduce to a simple form such as that of Eq. (27), but as Fig. 5 illustrates, the effect of the additional maturation steps in the large translation rate case shown is to reduce the variability of slow-maturing proteins compared to fast-maturing ones at all different transcription rates, with the lower limit being $\eta=1/\sqrt{m_{\text{protein-matured}}}$ (appearing as a dashed line in Figs. 5 and 6).

In Fig. 6, we are varying the translation rate c_3, and thus the offset $f(c_2,c_3,c_4)$ changes across the plot. For small translation rates, corresponding to low mean numbers of matured proteins, $f(c_2,c_3,c_4)\to\log_{10}1=0$, and thus the coefficient of variation approaches the $\eta=1/\sqrt{m_{\text{protein-matured}}}$ limit (shown as a dashed line on the plot). To address the behavior at large translation rates (and thus large mean protein numbers), we note that Eq. (27) may be rearranged into the form shown in Eq. (29), as has been previously derived [62]:

$$\eta_{\text{protein}} = \sqrt{\left(\frac{1}{m_{\text{protein}}}\right) + \eta_{\text{mRNA}}^2\frac{c_4}{c_2+c_4}}. \qquad (29)$$

As c_3 increases in Eq. (29), the $1/m_{\text{protein}}$ term vanishes, yielding $\eta_{\text{protein}}\approx\eta_{\text{mRNA}}\sqrt{c_4/(c_2+c_4)}$, a function only of c_1, c_2, and c_4: for high translation rates, the variability of the protein is dominated by that of the mRNA. Since we are only varying c_3 in Fig. 6, this term will be a constant, yielding the flat line seen in the high-number regime for the instant-maturation case. The analytic solution for the finite-maturation case once again does not have a compact form, but as seen in Fig. 6, the filtering effect of the maturation steps is observed only in the high-number (high-translation-rate) regime, where the maturation process reduces the coefficient of variation of proteins with slow maturation compared to those with fast or instant maturation. The difference in behavior in the high-number regime in this case arises from a low-pass filtering effect created by the maturation steps, as discussed in Sec. II B, below.

The results in Figs. 5 and 6 are not significantly affected in the case where the fluorescent proteins are actively degraded by proteolysis rather than the degradation being dominated by dilution due to cell growth. Taking the proteins to degrade as rapidly as the mRNA by setting $c_4=c_2$ yields a revised Fig. 6 (not shown here) with the same features discussed above: the coefficient of variability is dominated by the $1/m_{\text{protein}}$ term when the mean is small, converges to the constant $\eta_{\text{mRNA}}\sqrt{c_4/(c_2+c_4)}$ for large protein numbers, and shows a reduction in variability with increasing maturation time in the large-number regime.

The results in this section clearly indicate that neglecting protein-maturation effects will lead to significant underesti-

mates of the number of proteins of interest being produced in the cell (Fig. 3), since the number of observable (matured and fluorescing) proteins decreases as the maturation time increases. The observed fluorescence also provides an inaccurate measure of the variability in the total production of the target protein (Figs. 5 and 6). When the translation rate is small, the small-number effects dominate and the increased variability introduced by the process of maturation closely matches that expected from the reduced mean numbers of matured proteins present, whereas at higher translation rates the trend is reversed, with maturation acting to reduce the observed variability. This latter effect arises from a low-pass filtering introduced by the maturation steps, as we will show by considering noise power spectra, below.

B. Noise power spectra

To further investigate how the fluctuations in species numbers are affected by the addition of the maturation steps, we calculate noise power spectra using a method developed by Warren *et al.* [63]. The noise power spectrum is the Fourier transform of the autocorrelation function and is defined as

$$P_i(\omega) = 2\int_0^\infty \cos(\omega t) R_{ii}(t). \tag{30}$$

The correlation function is defined as

$$R_{ii}(t) = \langle \Delta n_i(0) \Delta n_i(t) \rangle, \tag{31}$$

where $\Delta n_i(t) \equiv n_i(t) - E[n_i(t)]$.

The noise power spectra follow the sum rule

$$\pi^{-1}\int_0^\infty P_i(\omega) d\omega = \sigma_i^2, \tag{32}$$

where σ_i^2 is the variance of the ith species. In the following analysis, we redefine the noise power spectrum to be $P_i(\omega)/\langle n_i \rangle^2$, so that the integral of the noise power spectrum will simply equal $\pi(\eta)^2$, a convenient form given our use of the coefficient of variation as an indicator of the degree of variability in a species.

The equations for the noise power spectra can be obtained by taking the Laplace transform of a set of differential equations giving the time evolution of the correlation coefficients [63]. Solving those equations in MAPLE, we get the analytical results for the noise power spectra; again, the detailed results are too unwieldy to reproduce here [61].

Figure 7 shows the noise power spectra for different species for a parameter set where the noise is suppressed by the maturation steps—that is, where the coefficient of variation of the matured species is lower for finite maturation times than when the maturation is taken to happen instantly (see the high-number regime in Fig. 6 and the whole range of Fig. 5). The noise power spectrum for a given species is obtained by neglecting all reactions (except degradation) downstream of the species of interest.

Equations (25) and (27) demonstrate that for certain parameter values [when $c_3 > c_4(1+c_4/c_2)$], the coefficient of variation of the protein can be lower than that of the mRNA.

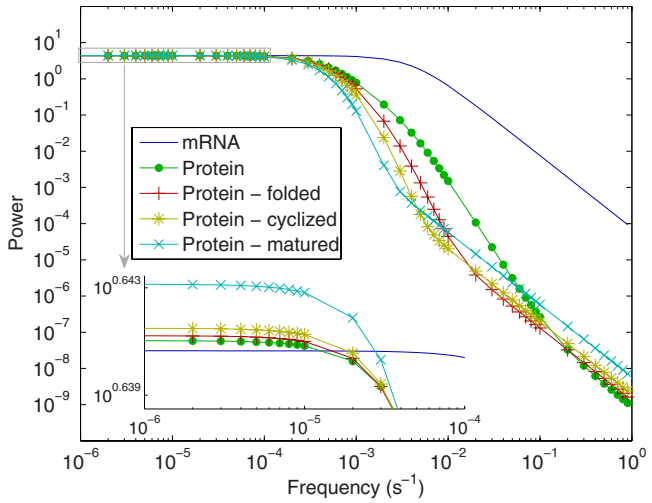

FIG. 7. (Color online) Noise power spectra for different species in the model, calculated as described in the text, at transcription rate c_1=0.46 s^{-1}, translation rate c_3=4.8 s^{-1}, and a maturation time of 107 min ($c_5=c_6=4^{-5}$ s^{-1}, $c_7=4^{-6}$ s^{-1}). Parameter settings have been chosen to yield lower noise in the matured protein species than in the instant-maturation case (not shown). The curves shown are analytic solutions, with points added only to distinguish between the lines.

This may also be seen in the noise power spectra shown in Fig. 7. Except in the very-low-frequency range, the noise power of the protein species is smaller than that of the mRNA species. The fluctuations in the mRNA level are partly suppressed by the translation reaction, which acts like a low-pass filter.

Figure 7 also shows the noise power spectra for the species protein-folded, protein-cyclized, and protein-matured. Each of the three maturation steps decreases the noise in the middle frequency range, increases the noise at high frequencies, and very slightly increases the noise at the low-frequency end of the spectrum. Samoilov *et al.* [64] conclude using deterministic analysis that linear feedforward networks of chemical reactions act as low-pass filters, while Warren *et al.* [63] examine stochastic reaction behavior, noting that such reactions combine low-pass filtering with an injection of noise from their own noisy reaction rates, creating a noisy low-pass filter. In Fig. 7 the model has been placed in a regime wherein the low-pass filtering effects dominate and the coefficient of variation of the final matured protein species is lower than in the instant-maturation case. Figure 8 shows a different parameter set, where the coefficient of variation of the matured protein is higher than for the instant-maturation case. Here, the low-pass filtering behavior is still visible as a reduction in noise power in the higher-frequency ranges, but this reduction in noise power is dominated by an increase in noise power in the low-frequency range, leading to an overall increase in variability as a result of the maturation steps.

Consider just the final step in the maturation process, the oxidation of protein-cyclized that yields protein-matured. The noise in protein-matured is affected by three factors: The intrinsic noise introduced by fluctuations in the rate of the

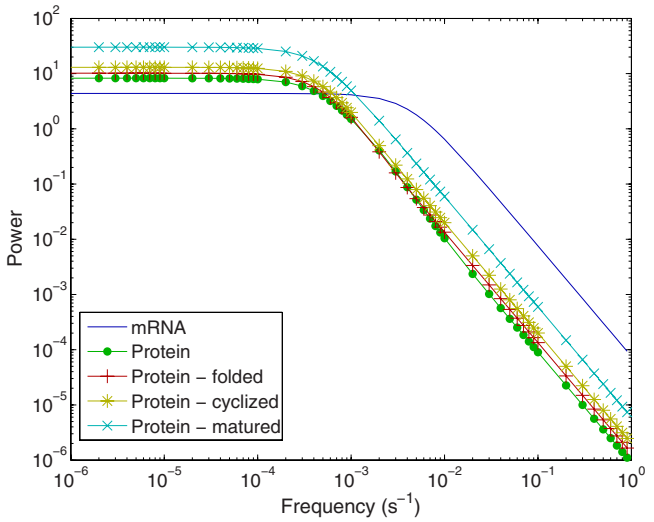

FIG. 8. (Color online) Noise power spectra for different species in the model, at transcription rate $c_1=0.46$ s^{-1}, translation rate $c_3=0.0047$ s^{-1}, and maturation time 107 min ($c_5=c_6=4^{-5}$ s^{-1}, $c_7=4^{-6}$ s^{-1}). Parameter settings have been chosen to yield higher noise in the matured protein species than in the instant-maturation case (not shown). The curves shown are analytic solutions, with points added only to distinguish between the lines.

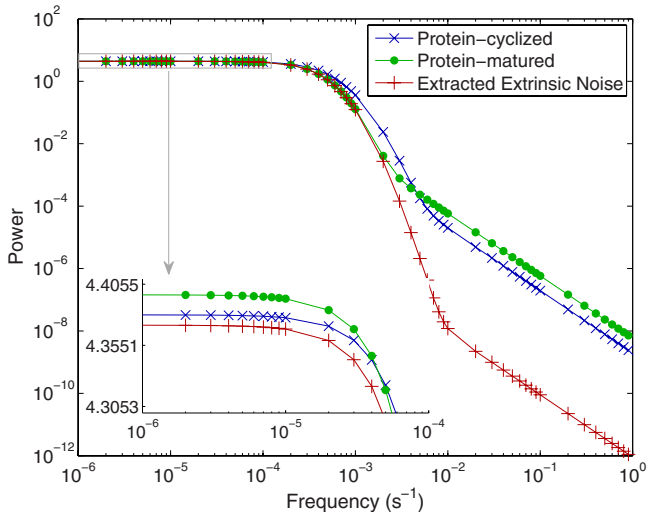

FIG. 9. (Color online) Noise power spectra for species protein-cyclized, protein-matured, and protein-matured minus the intrinsic noise power spectrum of the conversion reaction, which gives the level of upstream extrinsic noise power (from protein-cyclized) still present in species protein-matured, at transcription rate $c_1=0.46$ s^{-1}, translation rate $c_3=4.8$ s^{-1}, and maturation time 107 min ($c_5=c_6=4^{-5}$ s^{-1}, $c_7=4^{-6}$ s^{-1}). The curves shown are analytic solutions, with points added only to distinguish between the lines.

oxidation reaction itself, the extrinsic noise from the previous step in the process (that is, the noise in protein-cyclized), and the suppression of the extrinsic noise by the action of the oxidation reaction. The presence of the oxidation step has two effects: It acts also as a low-pass filter on the noise from upstream (that is, the noise in the species protein-cyclized) and it adds the noise from its own fluctuations [63]. The intrinsic noise introduced by the oxidation step may be calculated by fixing the number of protein-cyclized at its mean value and generating the noise power spectrum for species protein-matured under these conditions. To examine the filtering effect, we subtract this intrinsic noise spectrum from the full noise power spectrum of protein-matured and denote the remaining noise power spectrum as the "extracted extrinsic noise" of the matured species. This noise spectrum, along with the original spectra for the cyclized and matured species, is plotted in Fig. 9; note the large degree of suppression of the higher frequencies evident in the extracted extrinsic noise. We then compute the "passing ratio" for the process by dividing the extracted extrinsic noise in protein-matured by the total noise present in protein-cyclized. The result is plotted in Fig. 10, which clearly shows the noisy low-pass filter behavior of this single-step maturation process. Increasing the maturation time has the effect of increasing the range of frequencies over which the low-pass filtering effect operates, and thus for equal mean numbers of matured protein, longer maturation times will yield a lower observed coefficient of variation, as seen in Fig. 6. Each step in the full maturation process has the same structure as the final step considered here, so the noise suppression is repeated and augmented through the folding, cyclization, and oxidation steps leading to the fully matured protein.

Figure 11 shows the noise power spectra for different maturation times at a fixed transcription and translation rate. At the same mean level of protein expression, longer maturation times lead to greater noise suppression (see Fig. 6), but with fixed transcription and translation rates, longer maturation times yield lower numbers of mature proteins. Figure 11 thus illustrates the competition between the injection of extra noise from the additional reaction steps and the low-pass filtering effect of these reactions: At the parameters shown, for short maturation times the filtering effect dominates, reducing the noise power by filtering away higher-

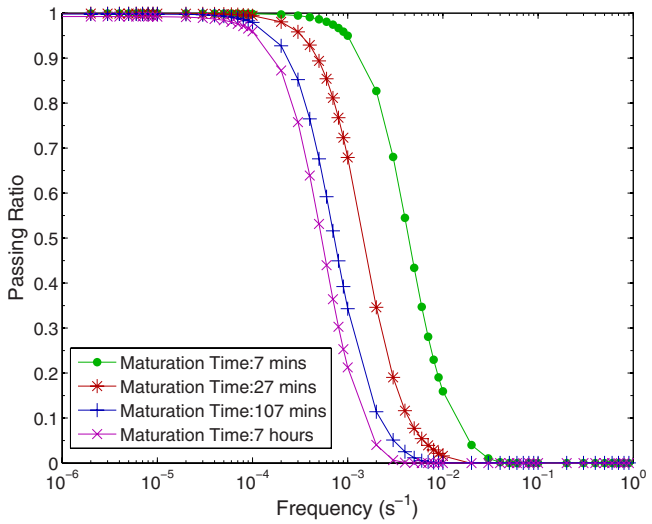

FIG. 10. (Color online) Passing ratio of the upstream extrinsic noise through the final maturation reaction step (oxidation), for different maturation times, at transcription rate $c_1=0.46$ s^{-1} and translation rate $c_3=4.8$ s^{-1}. The curves shown are analytic solutions, with points added only to distinguish between the lines.

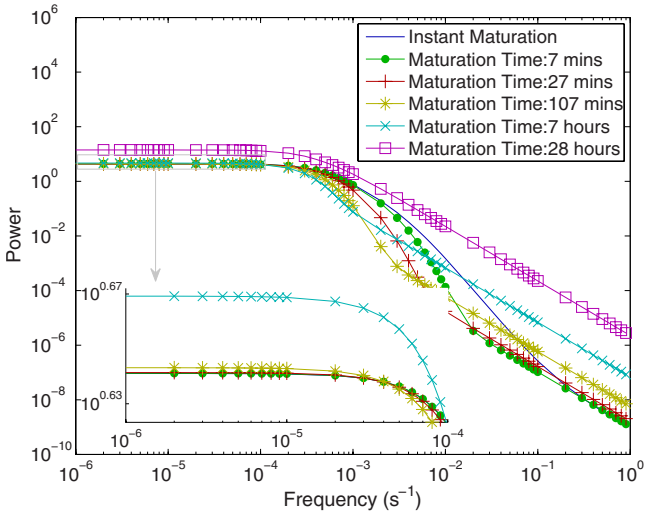

FIG. 11. (Color online) Noise power spectrum for different maturation times while trancription rate and translation rate are kept constant, at transcription rate $c_1=0.46\ s^{-1}$ and translation rate $c_3=4.8\ s^{-1}$. The curves shown are analytic solutions, with points added only to distinguish between the lines.

frequency components; at longer maturation times, this effect is swamped by the injection of fluctuations caused by lowering the molecule numbers, and the 27-h maturation time yields higher noise power at all frequencies, with a corresponding increase in the observed coefficient of variation.

III. CONCLUSION

We have examined a model of the production of fluorescent reporter proteins incorporating significant simplifications: we have omitted a number of intermediate steps in gene expression, and we have neglected nonlinear effects arising from processes such as regulatory feedback. The parameter settings have been selected to reflect typical values for *Escherichia coli* and thus may not be appropriate for all organisms.

Keeping these caveats in mind, we may use the model to extract information about the effects of protein maturation on what is seen experimentally when using fluorescent proteins as reporters. As we have seen, the observed level and variability of gene expression can be significantly affected by the maturation steps that must be completed before a fluorescent protein can generate fluorescence output. Interestingly, the effect is not uniformly in the direction of increasing the observed variability: In some regimes, the observed variability may in fact decrease compared to the actual variability of the total protein production. The maturation process has two effects: It introduces additional intrinsic noise from fluctuations in the rates of the maturation reactions, and it acts as a low-pass filter, removing some of the noise from earlier steps in the gene expression process. Longer maturation times lead to a smaller pass band in the low-pass filter, which increases the frequency range of upstream noise filtered out; at the same time, longer maturation times introduce larger intrinsic noise. Combining these two effects for a given translation rate, there is a range of maturation times where the noise in the output (matured) protein is suppressed rather than enhanced, with the suppression manifesting itself as a reduction in noise power in the intermediate frequency ranges. The low-pass filtering acts to smooth out bursts of protein production from the translation process. Thus the noise-reduction effect occurs most strongly at large translation rates, which will have more high-frequency components in the immature protein numbers, and at longer maturation times, which will have smaller passing bands for the low-pass filter effect. In general, we observe that the larger the translation rate, the larger will be the set of maturation times able to suppress the variability in the matured protein, and the longer the maturation time, the stronger the noise-suppression effect will be. Though we have analyzed the case of maturation of a fluorescent protein, the model can represent any linear chain of transformation reactions, and thus our results might be of use in other cases as well.

Based on the model, the process of fluorescent protein maturation affects the observed number of proteins present, when fluorescent reporters are used as reporters in living cells; as expected, the longer the maturation time, the fewer matured proteins will be visible relative to the total protein population. It is perhaps less intuitively clear that this maturation process also affects the observed variability in gene expression and that the effect can either increase or decrease the observable coefficients of variation relative to the true variability of the protein population. In either case, it is clear that the maturation process can potentially introduce a significant perturbation on the observed data and should be taken into account in quantitative studies of gene expression. Our model suggests a method of correcting experimental data to yield more accurate measurements of levels of protein expression and variability, and future work could extend these results to more complex systems.

ACKNOWLEDGMENTS

This work was supported by the Natural Science and Engineering Research Council (NSERC) of Canada and by the Canada Foundation for Innovation (CFI).

APPENDIX

The connection between the stochastic reaction constants (c) and the deterministic reaction-rate constants (k) [50] is $c_1'=k_1/v$, $c_2=k_2$, $c_3'=k_3/v$, $c_4=k_4$, $c_5=k_5$, $c_6=k_6$, and $c_7=k_7$, where v is the volume of the cell.

The reaction rates used for the parameter sweep are listed below: Transcription rate: $c_1=c_1'n_{\mathrm{RNAP}}n_{\mathrm{DNA}}=(4^{1-m_1}\times 2\times 10^{-19}\ \mathrm{l\ s^{-1}})/v\times 7600\times 154=4^{1-m_1}\times 117\ \mathrm{s^{-1}}$, where $m_1=1,\ldots,7$, we take the number of RNAP and DNA to be 7600 and 154, respectively (modeling a medium-copy plasmid bearing many copies of the fluorescent protein of interest) [49,58], and we assume a typical *E. coli* cell volume of $v=2\times 10^{-15}$ l [65]. Degradation rate of mRNA: $c_2=4.2\times 10^{-3}\ \mathrm{s^{-1}}$, corresponding to a typical bacterial mRNA half-life of 4.8 min [45]. Translation rate: $c_3=c_3'n_{\mathrm{ribo}}=(4^{1-m_2}\times 2\times 10^{-19}\ \mathrm{l\ s^{-1}})/v\times 48\,000=4^{1-m_2}\times 4.8\ \mathrm{s^{-1}}$, where m_2

$=1,\ldots,10$, and we take the number of ribosomes to be 48 000 [49,58]. Degradation rate of protein: $c_4=\ln 2/(1440\ \mathrm{s})=4.8\times10^{-4}\ \mathrm{s}^{-1}$, representing the effective degradation of proteins through dilution by cell growth during a 1440-s (24-min) cell division cycle. Folding rate: $c_5=4^{2-m_3}\ \mathrm{s}^{-1}$, $m_3=1,\ldots,20$. Cyclization rate: $c_6=4^{2-m_3}\ \mathrm{s}^{-1}$, $m_3=1,\ldots,20$. Oxidation rate: $c_7=4^{1-m_3}\ \mathrm{s}^{-1}$, $m_3=1,\ldots,20$ [37–39].

The maturation time is calculated for the case where the protein is subject to no processes other than maturation; that is, we neglect degradation and consider the following three reactions: protein→protein-folded, protein-folded →protein-cyclized, and protein-cyclized→protein-matured. The average protein-matured number as a function of time, given an initial protein number p_0, is

$$m_{\text{protein-matured}}(t)=\frac{c_5p_0}{(c_5-c_6)}\left(\frac{c_7}{c_7-c_5}(e^{-c_5t}-1)+\frac{c_7}{c_6-c_7}(e^{-c_6t}-1)+\frac{c_6(c_6-c_7)}{(c_7-c_5)}\times(e^{-c_7t}-1)\right).$$

We solve the above equation numerically for the time at which the number of the species protein-matured will be $(1-1/e)p_0$, or equivalently when the immature protein number will be p_0/e.

[1] T. Ideker, T. Galitski, and L. Hood, Annu. Rev. Genom. Hum. G. **2**, 343 (2001).
[2] A. P. Arkin, Curr. Opin. Biotechnol. **12**, 638 (2001).
[3] H. Kitano, Science **295**, 1662 (2002).
[4] M. E. Csete and J. C. Doyle, Science **295**, 1664 (2002).
[5] F. J. Doyle and J. Stelling, J. R. Soc., Interface **3**, 603 (2006).
[6] J. Hasty, D. McMillen, and J. J. Collins, Nature (London) **420**, 224 (2002).
[7] M. Kaern, W. Blake, and J. J. Collins, Annu. Rev. Biomed. Eng. **5**, 179 (2003).
[8] T. Pawson and R. Linding, FEBS Lett. **579**, 1808 (2005).
[9] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, Mol. Syst. Biol. **2**, 2006.0028 (2006).
[10] K. E. Tyo, H. S. Alper, and G. N. Stephanopoulos, Trends Biotechnol. **25**, 132 (2007).
[11] P. Smolen, D. Baxter, and J. Byrne, Neuron **26**, 567 (2000).
[12] J. Hasty, D. McMillen, F. Isaacs, and J. J. Collins, Nat. Rev. Genet. **2**, 268 (2001).
[13] M. Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. U.S.A. **98**, 8614 (2001).
[14] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proc. Natl. Acad. Sci. U.S.A. **99**, 12795 (2002).
[15] J. Paulsson, Nature (London) **427**, 415 (2004).
[16] J. M. Raser and E. K. O'Shea, Science **309**, 2010 (2005).
[17] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Nat. Rev. Genet. **6**, 451 (2005).
[18] P. S. Swain and A. Longtin, Chaos **16**, 026101 (2006).
[19] A. Becskei and L. Serrano, Nature (London) **405**, 590 (2000).
[20] W. J. Blake, M. Kaern, C. R. Cantor, and J. J. Collins, Nature (London) **422**, 633 (2002).
[21] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).
[22] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[23] D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty, Nature (London) **439**, 861 (2006).
[24] J. Paulsson and M. Ehrenberg, Q. Rev. Biophys. **34**, 1 (2001).
[25] H. B. Fraser, A. E. Hirsh, G. Giaever, J. Kumm, and M. B. Eisen, PLoS Biol. **2**, e137 (2004).
[26] J. M. Pedraza and A. van Oudenaarden, Science **307**, 1965 (2005).
[27] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, Science **307**, 1962 (2005).
[28] R. Y. Tsien, Annu. Rev. Biochem. **67**, 509 (1998).
[29] T. Nagai, K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba, and A. Miyawaki, Nat. Biotechnol. **20**, 87 (2002).
[30] W. J. Blake, G. Balazsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, and J. J. Collins, Mol. Cell **24**, 853 (2007).
[31] M. B. Elowitz and S. Leibler, Nature (London) **403**, 335 (2000).
[32] T. Gardner, C. R. Cantor, and J. J. Collins, Nature (London) **403**, 339 (2000).
[33] R. Weiss, S. Basu, S. Hooshangi, A. Kalmbach, D. Karig, R. Mehreja, and I. Netravali, Nat. Comput. **2**, 47 (2003).
[34] Y. Yokobayashi, C. H. Collins, J. R. Leadbetter, R. Weiss, and F. H. Arnold, Adv. Complex Syst. **6**, 1 (2003).
[35] L. You, R. S. Cox, R. Weiss, and F. H. Arnold, Nature (London) **428**, 868 (2004).
[36] C. A. Voigt, Curr. Opin. Biotechnol. **17**, 548 (2006).
[37] D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormler, Gene **111**, 229 (1992).
[38] C. W. Cody, D. C. Prasher, W. M. Westler, F. G. Prendergast, and W. W. Ward, Biochemistry **32**, 1212 (1993).
[39] B. G. Reid and G. C. Flynn, Biochemistry **36**, 6786 (1997).
[40] V. V. Verkhusha, N. A. Akovbian, E. N. Efremenko, S. D. Varfolomeyev, and P. V. Vrzheshch, Biochemistry **66**, 1342 (2001).
[41] Y. Gerena-Lopez, J. Nolan, L. Wang, A. Gaigalas, A. Schwartz, and E. Fernandez-Repollet, Cytometry Pt. A **60**, 21 (2004).
[42] N. Rosenfeld, T. J. Perkins, U. Alon, M. B. Elowitz, and P. S. Swain, Biophys. J. **91**, 759 (2006).
[43] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, Nat. Genet. **38**, 636 (2006).
[44] B. Lewin, *Genes IX* (Jones & Bartlett, Sudbury, MA, 2007).
[45] H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, M. Scott, A. Bretscher, H. Ploegh, and P. Matsudaira, *Molecular Cell Biology* (Freeman, New York, 2007).
[46] M. Ptashne, *A Genetic Switch: Phage λ and Higher Organisms*

(Cell Press and Blackwell Scientific, Cambridge, MA, 1986).
[47] M. Ptashne and A. Gann, *Genes & Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2002).
[48] C. G. Dong, L. Jakobowski, and D. R. McMillen, J. Biol. Phys. **32**, 173 (2006).
[49] M. A. J. Iafolla and D. R. McMillen, J. Phys. Chem. B **110**, 22019 (2006).
[50] D. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[51] D. Adalsteinsson, D. McMillen, and T. C. Elston, BMC Bioinf. **5**, 24 (2004).
[52] R. J. Ellis, Trends Biochem. Sci. **26**, 597 (2001).
[53] S. Schnell and T. E. Turner, Prog. Biophys. Mol. Biol. **85**, 235 (2004).
[54] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001).
[55] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007).
[56] T. Jahnke and W. Huisinga J. Math. Biol. **54**, 1 (2007).
[57] C. Gadgil, H. L. Chang, and H. G. Othmer, Bull. Math. Biol. **67**, 901 (2005).
[58] H. Bremer and P. P. Dennis, in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, 2nd ed., edited by F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (American Society for Microbiology, Washington, D.C., 1996), pp. 1553–1569.
[59] P. Corish and C. Tyler-Smith, Protein Eng. **12**, 1035 (1999).
[60] A. Varshavsky, Genes Cells **2**, 13 (1997).
[61] See EPAPS Document No. E-PLEEE8-77-048802 for the MAPLE worksheet yeilding the full results, as a script and as a PDF file. For more information on EPAPS see http://www.aip.org/pubserv/epaps/html.
[62] J. Paulsson, Phys. Life Rev. **2**, 157 (2005).
[63] P. Warren, S. Tânase-Nicola, and P. R. ten Wolde, J. Chem. Phys. **125**, 144904 (2006).
[64] M. Samoilov, A. Arkin, and J. Ross, J. Phys. Chem. A **106**, 10205 (2002).
[65] F. J. Trueba and L. J. H. Koppes, Arch. Microbiol. **169**, 491 (1998).