Mathematical Analysis and Quantification of Fluorescent Proteins as Transcriptional Reporters

Xiao Wang,* Beverly Errede,† and Timothy C. Elston‡
*Department of Biomedical Engineering, Center for BioDynamics, and Center for Advanced Biotechnology, Boston University, Boston, Massachusetts 02215; †Department of Biochemistry and Biophysics, and ‡Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

ABSTRACT Fluorescent proteins are often used as reporters of transcriptional activity. Here we present a mathematical characterization of a novel fluorescent reporter that was recently engineered to have a short half-life (~12 min). The advantage of this destabilized protein is that it can track the transient transcriptional response often exhibited by signaling pathways. Our mathematical model takes into account the maturation time and half-life of the fluorescent protein. We demonstrate that our characterization allows transient transcript profiles to be inferred from fluorescence data. We also investigate a stochastic version of the model. Our analysis reveals that fluorescence measurements can both underestimate and overestimate fluctuations in protein levels that arise from the stochastic nature of biochemical reactions.

INTRODUCTION

A common property of signaling pathways is that they often act transiently in the presence of a sustained stimulus. For example, yeast respond to mating pheromone by inducing a transient transcriptional program. Therefore there is great interest in measuring gene expression changes in individual living cells as they respond to stimuli in real time. In principle, this could be accomplished with fluorescent proteins. In a recent study, one of us (Beverly Errede) engineered and experimentally characterized a set of short-lived fluorescent reporters (1). These novel reporters were shown to accurately track the time-dependent behavior of pheromone-induced transcription. Fluorescent proteins also have been used to measure variability, both temporal and intercellular, in protein expression levels (2–14). Determining the origins and magnitude of these fluctuations is of interest because of their implications for cell fate decisions and nongenetic individuality.

Many studies on gene expression in single cells have been motivated by theoretical and computational analyses of mathematical models of the underlying system (15–21). Using mathematical models to interpret fluorescence measurements requires a quantitative characterization of the biochemical properties of fluorescent proteins used as reporters. In particular, knowledge of the fluorescent reporter's half-life and maturation kinetics (i.e., folding and oxidation (22)) is critical for this comparison. Here, we use mathematical modeling to quantitatively characterize the short-lived fluorescent proteins reported in Hackett et al. (1). We show that this characterization allows us to infer the underlying transcriptional response from fluorescent measurements, thereby providing a tool for monitoring transcript levels in single cells. Next we use stochastic modeling to investigate how the fluorescence maturation time and protein half-life influence fluctuations in fluorescence levels. Our analysis reveals that for proteins with short half-lives fluorescence measurements can overestimate fluctuations in protein levels, whereas for long-lived reporters fluorescence measurements typically underestimate these fluctuations.

Submitted September 17, 2007, and accepted for publication November 7, 2007.

Address reprint requests to Timothy C. Elston, Dept. of Pharmacology, University of North Carolina at Chapel Hill, 405 Mary Ellen Jones Building, Campus Box 7260, Chapel Hill, NC 27599-7260. Tel.: 919-843-7670; Fax: 919-966-5640; E-mail: telston@med.unc.edu.

Editor: Jason M. Haugh.

0006-3495/08/03/2017/10 $2.00

METHODS

Experimental characterization of short-lived fluorescent protein reporters

We begin by briefly summarizing recent work carried out in the Errede laboratory to develop and experimentally characterize a novel class of short-lived fluorescent proteins (1). The approach used to generate a family of cyan fluorescent reporter proteins (CFP) with different stabilities was based on the ubiquitin fusion strategy for programmable N-end rule degradation developed by Varshavsky and colleagues (23). None of the proteins involved in the degradation process are regulated by the cell cycle (24). To experimentally characterize the novel short-lived reporters, the galactose-dependent and glucose-repressible *GAL1* promoter was used to drive their expression. Immune blot analysis of protein extracts and fluorescence imaging of individual living cells were used to determine protein half-lives after further transcription was inhibited. Protein accumulation and the emergence of fluorescence were also monitored after shifting cultures from a glucose to a galactose medium. These measurements revealed a long delay between the appearance of newly synthesized protein and the onset of fluorescence (see Hackett et al. (1) for details).

Having experimentally characterized the intrinsic properties of the short-lived reporters, we next tested them for their ability to act as reporters of time-dependent transcriptional activity. Yeast respond to mating pheromone by inducing a transient transcription program. *FUS1* expression is strongly induced by pheromone and serves as a standard indicator for mating specific gene expression. Therefore, the *FUS1* promoter was exploited to compare the performance of destabilized (^{P}*FUS1-UbiY-dkCFP*) versus stable (^{P}*FUS1-UbiM-dkCFP*) fluorescent genes as transcription reporters. The pheromone-induction kinetics measured by fluorescence for both reporters is significantly delayed compared with that measured by messenger RNA (mRNA) abundance (Fig. 1). The speed with which either reports transcription induction is

doi: 10.1529/biophysj.107.122200

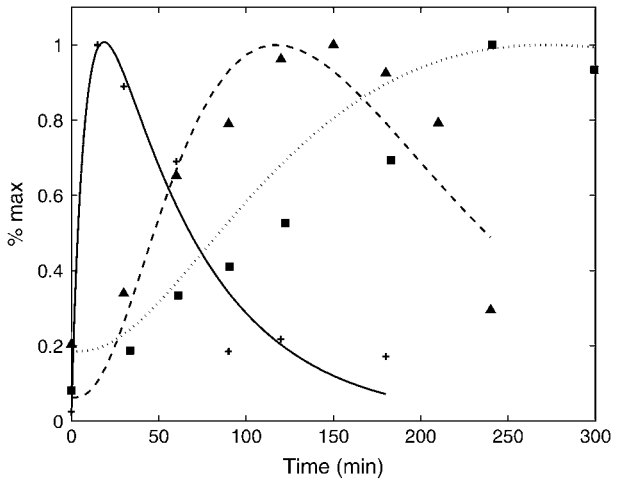

FIGURE 1 Time courses for the transcript level (*crosses*) and fluorescence measurements from a short-lived reporter *YdkCFP* (half-life = 12 min, *triangles*) and long-lived reporter *MdkCFP* (half-life = 76 min, *squares*) (1). The solid curve is the mRNA profile used as input for the model. The dashed and dotted curves are the model output for the short- and long-lived reporters, respectively (see text for details). The values of the parameters estimated from fitting the model to the experimental data are $k_m = 0.0054$ min^{-1}, $k_u = 34$ min^{-1}, and $dk_u = 81.7$ min^{-1}.

constrained by the inherent time required for fluorophore maturation. However, the advantage of the short-lived reporter is evident in that the attenuation phase of the pheromone-induced profile is similar to that for its mRNA. By contrast, accumulation of the stable *MdkCFP* reporter completely masks the transient profile.

RESULTS

Mathematical characterization of fluorescent protein reporters

Our ultimate goal is to use short-lived reporters as experimental readouts that can be quantitatively compared with output from computational models of pathway activity. Therefore it is critical to have a mathematical model that accurately describes the synthesis, maturation, and degradation events associated with these proteins. Here we present a model that reproduces experimental data used to characterize these reporters. In our model, premature (nonfluorescent) protein, P, is synthesized at a rate that is proportional to current mRNA concentration. Once synthesized the premature protein can either mature into a fluorescently competent protein, P_M, or be ubiquitinated, P_U. We assume that the ubiquitination process is reversible and that ubiquitinated protein is subject to degradation. Mature protein can be ubiquitinated, and likewise ubiquitinated protein can mature. Both processes produce the species P_{MU}. These considerations lead to the following four equations for the concentrations of the various protein species:

$$\frac{dP}{dt} = \gamma\, mRNA(t) - k\mathrm{u}\, P - k\mathrm{m}\, P + dk\mathrm{u}\, P_U \quad (1)$$

$$\frac{dP_U}{dt} = k\mathrm{u}\, P - \delta P_U - k\mathrm{m}\, P_U - dk\mathrm{u}\, P_U \quad (2)$$

$$\frac{dP_M}{dt} = k\mathrm{m}\, P - k\mathrm{u}\, P_M + dk\mathrm{u}\, P_{MU} \quad (3)$$

$$\frac{dP_{MU}}{dt} = k\mathrm{u}\, P_M - \delta P_{MU} + k\mathrm{m}\, P_U - dk\mathrm{u}\, P_{MU}. \quad (4)$$

In Eq. 1, $mRNA(t)$ represents the concentration of mRNA at time t, and γ is the translation efficiency. The parameters k_u, dk_u, k_m, and δ are the ubiquitination, deubiquitination, maturation, and degradation rates, respectively.

Fig. 1 shows data for the short-lived reporter *YdkCFP* (half-life = 12 min, *triangles*) and the long-lived reporter *MdkCFP* (half-life = 76 min, *squares*). The half-lives correspond to $\delta = 0.055$ min^{-1} for the short-lived reporter and $\delta = 0.009$ min^{-1} for the long-lived reporter (1). The input for the model is the time-dependent mRNA profile (Fig. 1, *crosses*). These data were fit assuming a functional form that consists of the difference of two exponentials (i.e., $mRNA(t) = a\exp(-\alpha_1 t) - b\exp(-\alpha_2 t)$). This produced the solid curve shown in Fig. 1. This curve then served as input for Eq. 1. The total mature protein concentration $P_M + P_{MU}$ was fit to both sets of fluorescence data using the nonlinear least squares routine in MATLAB (The MathWorks, Natick, MA). The results of this process are shown as the dotted (half-life = 76 min) and dashed (half-life = 12 min) curves in Fig. 1. The estimated parameter values are $k_m = 0.0054$ min^{-1}, $k_u = 34$ min^{-1}, and $dk_u = 81.7$ min^{-1}. Because we do not know the absolute levels of mRNA and protein concentrations, the synthesis rates cannot be directly determined from fitting the data. This is not a problem if we are only trying to determine the shape of the transcript profile from fluorescent measurements. However, to investigate fluctuations in gene expression requires these values (see below).

Note that the estimated ubiquitination and deubiquitination rates are much faster than the other biochemical processes in the model. Therefore, we can utilize a quasi-steady-state approximation that assumes the ubiquitinated and deubiquitinated forms of the protein are in equilibrium to simplify the model. This results in the following two equations:

$$\frac{dP_A}{dt} = \gamma\, mRNA(t) - k\mathrm{m}\, P_A - \delta' P_A \quad (5)$$

$$\frac{dP_{MA}}{dt} = k\mathrm{m}\, P_A - \delta' P_{MA}, \quad (6)$$

where $P_A = P + P_U$, $P_{MA} = P_M + P_{MU}$, and ($\delta' = \delta/(1 + dk_u/k_u)$). Equations 5 and 6 can be written in dimensionless form as follows:

$$\frac{dP'_A}{dt} = (k\mathrm{m} + \delta')\,(mRNA(t) - P'_A) \quad (7)$$

$$\frac{dP'_{MA}}{dt} = \delta'(P'_A - P'_{MA}) \quad (8)$$

where P_A' and P_{MA}' are defined as $P_A(k_m + \delta')/\gamma$ and P_{MA} δ' $(k_m + \delta')/(k_m\ \gamma)$, respectively. For the estimated model