The Stochastic Nature of Gene Expression Revealed at the Single-Molecule Level

Xin Chen and Paul S. Cremer*

Department of Chemistry, Texas A&M University, 3255 TAMU, College Station, Texas 77843

Every child is different. Even identical twins can be readily differentiated through subtle but noticeable differences in appearance and personality. Similarly, phenotypic differences can be observed for individual *Escherichia coli* cells, even if they have identical genomes. This diversity cannot be related back to inherent genetic variation. A major factor in such differences at the cellular level can be, however, related to variability in gene expression (*1, 2*), which is intrinsically stochastic when a low copy number of molecules is involved. Until very recently, most of our knowledge about gene expression has been gleaned through ensemble measurements where the underlying stochastic nature of the process can be easily masked by population averages. To fully understand stochastic events, it is necessary to study them at the single-molecule level. In the case of gene expression, this means following transcription, translation, and the production of proteins at the single-molecule level. In two recent publications, X. Sunney Xie and co-workers have followed single-protein expression events *in vivo* in an elegant set of studies (*3, 4*).

In the first study (*3*), a Tsr–Venus fusion protein was used as both a gene reporter and fluorescent signal. Venus (*5*), which is a type of yellow fluorescent protein, was fused with a membrane protein Tsr (Figure 1, panel a). The gene encoding Tsr–Venus was spliced into *E. coli* to replace the native *LacZ* gene. Fusion to Tsr positioned the fluorescent Venus at the cell surface, which significantly limited its

diffusion and improved signal sensitivity. *In vivo* single-molecule protein detection was achieved in real time by taking epifluorescence measurements every 3 min after applying a short photobleaching pulse (Figure 1, panel b). Each signal burst represented no more than a few Tsr–Venus molecules, and the peak heights were quantized, corresponding to the number of protein molecules that were present. Only nascently inserted proteins generated fluorescence signals, as photobleaching eliminated the response of previously observed molecules.

In the second study (*4*), enzymatic amplification was exploited to detect the *in vivo* production of the enzyme, β-galactosidase (β-gal). β-Gal was used to catalyze the hydrolysis of a synthetic substrate (FDG), which generates a fluorescent product (Figure 2, panel a). Single-molecule detection has long been demonstrated with commercial optical setups *in vitro* (*6, 7*); however, to apply this method *in vivo*, one has to face a practical challenge because fluorescein is continuously and efficiently pumped out of living cells and diffuses away. To circumvent this difficulty, an ingenious lab-on-a-chip method was employed to confine single cells inside enclosed micron-sized chambers in a microfluidic device (Figure 2, panel b). The chambers not only offered microscale confinement, but also parallelism, allowing multiple cells to be monitored simultaneously. The fluorescence from the chamber increased as the hydrolysis reaction proceeded. The detected reaction rate

ABSTRACT Two recent papers have monitored the real-time synthesis of proteins *in vivo* at the single-molecule level. The work was done by two separate methods: fluorescent protein labeling and enzymatic amplification. Statistical analysis of the data reveals the inherent stochastic nature of gene expression.

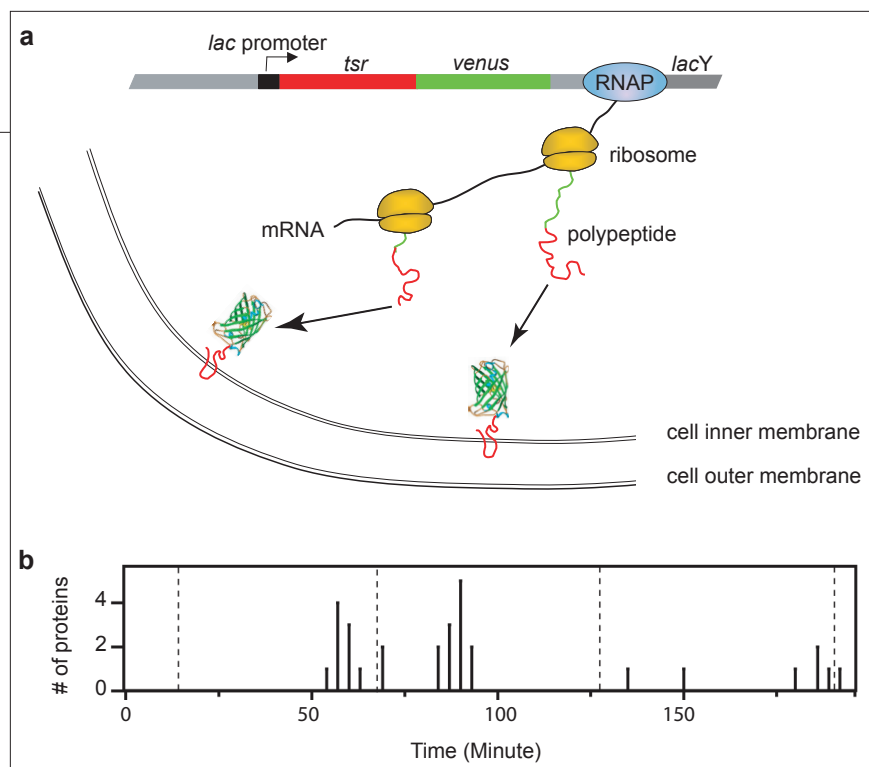*To whom correspondence should be addressed. E-mail: cremer@mail.chem. tamu.edu.

Figure 1. Gene expression of a green fluorescent protein detected at the single-molecule level. a) The gene encoding Tsr–Venus is expressed by a standard transcription and translation process. The nascently formed protein is then inserted into the inner membrane of *E. coli*. Venus was chosen over more traditional species such as green fluorescence protein because of its fast maturation time, ~4 min. Since the maturation is probably a stochastic event itself, the time resolution is inherently limited by the maturation event (*5*). **b)** Mature Tsr–Venus molecules are detected as individual burst events by fluorescence microscopy. The fluorescence signal was obtained every 3 min after photobleaching previously inserted Tsr–Venus molecules. The duration of this experiment was limited by the cells' resistance to photodamage under these conditions. The vertical axis represents the number of proteins synthesized in a 3-min time period, and the vertical dashed lines mark cell division events. Adapted from ref *3*.

was quantized, since the number of β-gal proteins present in the chamber was an integer number. The steps were therefore due to nascently synthesized β-gal, and the height of each step was proportional to the number of proteins that were made (Figure 2, panel c).

The reporter proteins in the two studies, Tsr–Venus and β-gal, were both expressed under highly repressed conditions and were monitored in real time with a resolution on the scale of minutes. The most important result in both experiments was bursts in protein production, which demonstrated that gene expression is an occasional event and that a few proteins are produced nearly simultaneously by such events, consistent with theoretical predictions (*8, 9*). Two key

parameters to describe such behavior, the burst size and frequency, correspond to how productive a single expression event is and how often such events occur. These data can be easily determined using Xie's methods. For the Tsr–Venus assay, each event produced about 4.2 proteins on average and occurred about 1.2 times per cell cycle. For β-gal, the burst sizes were larger (7.8 proteins/event), but less frequent (0.16 events per cell cycle) in the *E. coli* cells employed. Other types of cells, such as *Saccharomyces cerevisiae* (yKT32) and embryonic mouse stem cells, also showed similar characteristics, albeit with different burst sizes and frequencies (*4*).

In addition to the average burst size and frequency, even richer information can be

extracted from a statistical treatment of the temporal evolution profile of burst events. For example, the burst size was not uniform but varied from burst to burst, representing the fluctuation in productivity of a single expression event as well as reflecting its stochastic nature. The distributions measured in the two studies were well-fit by exponential and geometric distributions, respectively. Both distributions are simple statistical functions, which assume total randomness in event occurrence. This, in turn, suggested that the productivity of an expression event fluctuated randomly. Such a finding leads to a very simple, yet fundamental, question. Which step(s) in gene expression can account for the randomness of the process? For the Tsr–Venus project (*3*), the authors addressed this question by measuring the copy number of mRNAs encoding Tsr–Venus. The average number was about 1.0 copy per cell cycle, which matched the burst frequency within experimental error. This strongly suggested that there was only a single mRNA copy for each expression event, at least under the experimental conditions explored. Therefore, the measured fluctuation was not likely coming from transcription. Instead, the authors suggested that it probably can be attributed to the fluctuating number of ribosomal binding events of the mRNA. However, any step after transcription may contribute to the fluctuation. This would also include protein folding, incorporation into the membrane, and maturation of the Venus probe.

Additional information might be gleaned from the temporal evolution profile by looking at the exact timing of a burst event within the cell cycle or the correlation of burst events with one another. Although unlikely to be generally true, a completely random burst distribution would imply that the probability of expression for a particular protein is not affected by extrinsic parameters and that the particular stage of the cell cycle is not the deciding factor. As other protein expression
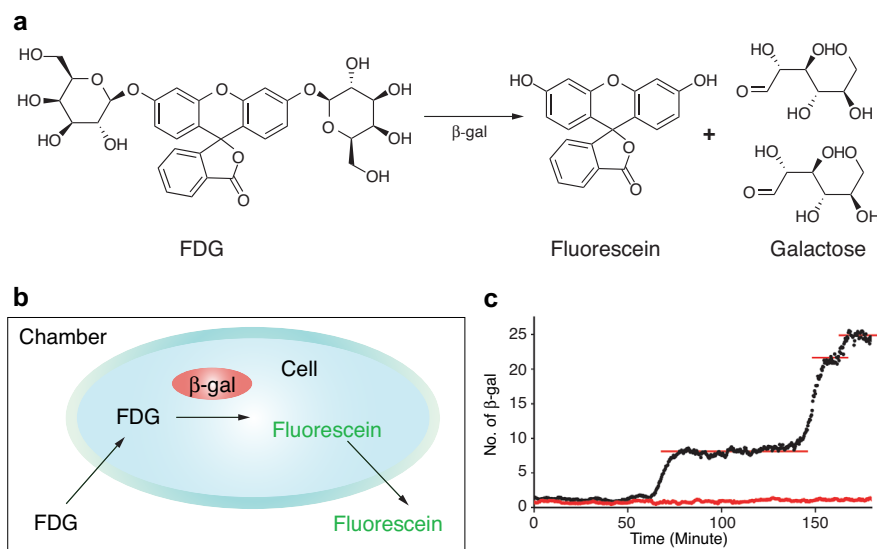
a

FDG → (β-gal) → Fluorescein + Galactose

b

c

Figure 2. Gene expression of β-gal detected at the single-molecule level through enzymatic amplification. a) The enzyme β-gal cleaves FDG to create fluorescein and two galactose molecules. b) Schematic diagram of a single-cell, single-chamber apparatus for β-gal measurements. Each cell was treated with β-lactam antibiotics to increase the permeability of its membrane. This treatment enabled the facile diffusion of FDG into the cell, which would otherwise have been greatly attenuated. c) The reaction rate (number of proteins expressed) vs time plot shows discrete production events. The height of each step corresponds to the number of nascently synthesized β-gal molecules. Adapted from ref 4.

amplification monitors its activity. The difference between these two types of assays, therefore, could potentially differentiate between production and activation. This would be intriguing, as activation processes such as post-translation modification should be stochastic as well.

REFERENCES

1. Novick, A., and Weiner, M. (1957) Enzyme induction as an all-or-none phenomenon. *Proc. Natl. Acad. Sci. U.S.A. 43*, 553–566.
2. Maloney, P. C., and Rotman, B. (1973) Distribution of suboptimally induced beta-d-galactosidase in *Escherichia coli*—enzyme content of individual cells. *J. Mol. Biol. 73*, 77–91.
3. Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X. S. (2006) Stochastic gene expression in live cells, one protein molecule at a time. *Science 311*, 1600–1603.
4. Cai, L., Friedman, N., and Xie, X. S. (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature 440*, 358–362.
5. Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol. 20*, 87–90.
6. Rotman, B. (1961) Measurement of activity of single molecules of beta-d-galactosidase. *Proc. Natl. Acad. Sci. U.S.A. 47*, 1981–1991.
7. Rondelez, Y., Tresset, G., Tabata, K. V., Arata, H., Fujita, H., Takeuchi, S., and Noji, H. (2005) Microfabricated arrays of femtoliter chambers allow single molecule enzymology. *Nat. Biotechnol. 23*, 361–365.
8. Berg, O. G. (1978) Model for statistical fluctuations of protein numbers in a microbial-population. *J. Theor. Biol. 71*, 587–603.
9. Rigney, D. R. (1979) Note on the kinetics and stochastics of induced protein-synthesis as influenced by various models for messenger-RNA degradation. *J. Theor. Biol. 79*, 247–257.
10. Levsky, J. M., Shenoy, S. M., Pezo, R. C., and Singer, R. H. (2002) Single-cell gene expression profiling. *Science 297*, 836–840.
11. Golding, I., Paulsson, J., Zawilski, S. M., and Cox, E. C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell 123*, 1025–1036.
12. Raser, J. M., and O'Shea, E. K. (2004) Control of stochasticity in eukaryotic gene expression. *Science 304*, 1811–1814.
13. Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002) Stochastic gene expression in a single cell. *Science 297*, 1183–1186.

events are explored by this methodology, it will remain to be seen if such correlative behavior can be found or not.

Steady-state population analysis can also be conducted, since many cells/chambers can be monitored simultaneously. The protein distribution over a population of cells depends not only on burst size and frequency, but also on protein partitioning during cell division and the correlation between protein expression and cell division. In the β-gal study (*4*), a gamma-function distribution should be followed assuming equipartitioning between the two daughter cells and no other correlations, as was the case.

The studies reviewed here are among the first single-molecule gene expression experiments (*10, 11*) providing statistical information on stochastic gene expression events, which is very difficult, if not impossible, to obtain through ensemble

averages. They also point to a very promising future in this subfield, as both methods can be extended, in principle, to multireporter systems. For example, one might genetically label one gene with green fluorescent protein and another one with yellow fluorescent protein, in the same cell (*12, 13*). The advantage of two reporters at the single-molecule level is not merely parallelism, but rather the rich information carried in temporal pair-correlations between expression events for different genes. It is reasonable to hypothesize that expression of two genes coding for two proteins with cooperative functions might have a strong positive correlation, while two independent genes might be weakly correlated. Combining the two assays developed by the Xie laboratory may also be uniquely informative. Yellow (or green) fluorescent protein labeling essentially monitors the existence of a protein, and enzymatic