Supplemental Data

Real-Time Kinetics of Gene Activity

in Individual Bacteria

Ido Golding, Johan Paulsson, Scott M. Zawilski, and Edward C. Cox

Observation of RNA Bursting from the P_{RM} Promoter of Bacteriophage lambda

In an attempt to test the generality of transcriptional bursting, we constructed a reporter system using an additional promoter. We chose to look at one of the best characterized genetic switches in all of biology, the P_{RM} promoter of bacteriophage λ, whose role it is to maintain the lysogenic state in a phage-infected *E. coli* (Chen et al., 2005; Ptashne, 1992). P_{RM} controls the expression of a polycistronic transcript containing the *cI*, *rexA* and *rexB* genes. The promoter and genes are delimited by two repressor-binding sequences, O_R and O_L, which bind the lambda repressor cI .

The $P_{lac/ara}$ promoter, used elsewhere in this work, is inducible by external inputs (IPTG and arabinose). In contrast, the P_{RM} system is completely auto-regulatory: the repressor is both the activator and the inhibitor of its own transcription[1]. Because of this feature, the expected RNA kinetics (and population heterogeneity) are radically different from the simple $P_{lac/ara}$ case. This difference—an autoregulatory loop, where cI is both repressor and activator—makes it even more interesting as a subject for RNA kinetic studies.

To construct a reporter system for *cI* transcription, we cloned the immunity region of wild-type λ (location 35.4-38.4 kb, (Hendrix, 1983)) into a promoterless BAC plasmid. Next, we replaced the *rexA* and *rexB* genes (not required for lysogenic maintenance (Hendrix, 1983)) with an array of 48 MS2 binding sites. We used the smaller array because full endogenous-level functioning of the cI switch requires that the two protein binding sites (O_R and O_L) be located at the appropriate distance (~3kb) from each other, enabling DNA looping and cI octamerization at high concentration (Dodd et al., 2001). The smaller array of binding sites, however, leads to a lower fluorescent signal-to-noise ratio, and with it, lower measurement sensitivity. Nonetheless, we have still found it possible to measure RNA levels in the individual cells.

To ascertain that the modified P_{RM} system is fully functional, we used the classical immunity test (Hendrix, 1983) and found that, as expected, the plasmid-born reporter system λ_{imm}(*rexAB*::48bs) confers on the host (DH5α-PRO) immunity against infection of wild-type lambda phage (λ_{WT}), but leaves it susceptible to infection by a virulent strain (λ_{vir}) and by a heteroimmune strain (λ_{434}).

In order to check for the occurrence of transcriptional bursting, cells carrying the reporter plasmid and the MS2-GFP expressing plasmid were grown and induced with aTc as described above. The RNA kinetics in individual cells was tracked as above, with the distinction that no external inducer was required for transcription. In following the cells (12 cells, each tracked for ~2 h) it was clear that RNA bursting indeed occurs. The smaller 48 binding site array limits our ability to characterize transcription events of $\Delta n = 1$ or 2, but the observed histogram of jumps Δn is fully consistent with a geometrical distribution, with mean $\langle \Delta n \rangle \approx 2.8$, and about half the events consisting of more than 2 transcripts (data not shown).

---

[1] In phases other than the maintenance of lysogeny, such as the initial establishment of lysogeny, and the event of switching into lysis (induction), another player - Cro – can bind to O_R and O_L and effect the function of the "genetic switch" (see e.g. Svenningsen, S. L., Costantino, N., Court, D. L., and Adhya, S. (2005). On the role of Cro in lambda prophage induction. Proc Natl Acad Sci U S A *102*, 4465-4469.).

As noted above, the autoregulatory nature of P_{RM} is expected to yield very different, and probably more complicated, microscopic kinetics (for example the statistics of off-times Δt_{off}) and population heterogeneity (e.g. variance) compared to those characterizing the $P_{lac/ara}$ system, and thus these data will have to be expanded and a more comprehensive theoretical analysis will have to be performed (work in progress).

Supplemental Discussion

Comparison with Previous Models and Experiments

Most studies have sought the explanation for protein fluctuations in the inherent randomness of gene expression, distinguishing between truly spontaneous mRNA fluctuations due to small numbers of transcripts, and enslaved mRNA fluctuations due to random changes in gene activity. In the first prokaryotic study, (Ozbudak et al., 2002), using *B. subtilis,* interpreted their data in terms of spontaneous mRNA fluctuations, using a model of Poisson-distributed mRNA copy number, where each one was translated a random number of times. The variance σ_m^2 in the number *m* of protein copies was then predicted to follow $\sigma_m^2/\langle m\rangle \approx 1+b$, where *b* is the average number of proteins translated per transcript. The value of $\sigma_m^2/\langle m\rangle$ could not be determined experimentally because the fluorescence per molecule was not known, but the hypothesis was supported by observations that $\sigma_m^2/\langle m\rangle$ was independent of the rate of transcription and increased linearly with the rate of translation. The same result can also be interpreted by normalizing the protein variance to look at $\sigma_m^2/\langle m\rangle^2$. The burst term then becomes $b/\langle m\rangle = 1/\langle n\rangle \times C$ where $\sigma_n^2/\langle n\rangle^2 = 1/\langle n\rangle$ represents the normalized variance in the mRNA numbers and $C = \tau_n/\tau_m$ is a proportionality constant that depends on the effective half-lives τ of the two species – the current protein concentration is affected by a history of mRNA fluctuations, not just a single fluctuation. A second prokaryotic study, this time in *E. coli* (Elowitz et al., 2002; Swain et al., 2002), also interpreted the inherent randomness of gene expression in terms of the small number of mRNAs per cell, and similarly showed that its contribution to the protein fluctuations followed $\sigma_m^2/\langle m\rangle^2 \approx 1/\langle n\rangle \times C$, though with a more complicated prediction for *C.* The first quantitative eukaryotic study, in *S. cerevisiae*, compared protein distributions with the predictions from detailed computer models (Blake et al., 2003). The data were found to be consistent with chromatin remodeling and quantal transcription bursts, and it was suggested that this may be a general difference between pro- and eukaryotes. A second *S. cerevisiae* study introduced a series of additional experimental controls and carefully compared different ways of affecting the protein variance with a mathematical model that also emphasized transcription bursts, finding good agreement between data and theory (Raser and O'Shea, 2004).

How do these previous results compare with the present findings, where we directly monitored mRNA levels with single-molecule resolution? We showed that even fully induced transcription occurs in quantal bursts, that the burst sizes are geometrically distributed, and that the time intervals between bursts are exponentially distributed, as expected from a simple gene activation-inactivation model. The same bursting parameter was further shown to explain several different statistical properties of the data. This may appear to contradict previous prokaryotic experiments by suggesting that most fluctuations come from quantal transcription bursts (equivalent to gene activation-inactivation). But the results are in fact perfectly consistent. If the quantal transcription demonstrated here also applied to the *B. subtilis* study – and if all their other assumptions held - their modified model can be shown to give $\sigma_m^2/\langle m\rangle \approx 1+b\langle \Delta n\rangle$. The average transcription burst $\langle \Delta n\rangle$ is independent of both translation and transcription frequencies, and would thus be constant in their experiments. Because the GFP measurements are in arbitrary units of fluorescence, this type of transcriptional bursting would thus have been inseparable from a standard Poissonian transcription process ($\langle \Delta n\rangle = 1$) in the absence of the methods developed here. The same is true for the second prokaryotic study (Elowitz et al., 2002). The models devised by the same group (Swain et al., 2002) even allowed for an intricate transcription mechanism that could produce transcription bursts, but GFP experiments could only demonstrate $\sigma_m^2/\langle m\rangle^2 \approx 1/\langle n\rangle \times C$ down to an arbitrary proportionality

constant *C*, again inseparable from $\sigma_m^2 / \langle m \rangle^2 \approx \langle \Delta n \rangle / \langle n \rangle \times C$. The demonstration of transcription bursts thus extends and complements rather than contradicts these two studies.

Intrinsic and Extrinsic Noise

Intrinsic noise is generally defined as the noise that originates within a certain system (however defined), while extrinsic noise comes from the perturbing effects of a changing environment (however defined). An elegant way of estimating these two contributions is to place two independent and identical systems in the same environment and measure correlateions between them (Elowitz et al., 2002). If the fluctuations come from the shared environment, the two systems should be correlated, and if they come from the internal processes, the two systems should be uncorrelated. By using two different fluorescent reporter genes, the intrinsic protein noise may be defined as coming from the spontaneous activation-inactivation of the genes and the spontaneous births and deaths of mRNAs or proteins. Extrinsic noise instead comes from fluctuations in the parameters of these processes, including fluctuations in the concentrations of activators, inhibitors, RNA polymerases, RNases, ribosomes, proteases and the like. Without the use of dual reporters, extrinsic noise can sometimes be estimated by the response of the variance to changes in the average. This is discussed in some detail in (Paulsson, 2004). In short, a constant extrinsic noise *B* can be calculated from $\sigma_n^2 / \langle n \rangle^2 = A / \langle n \rangle + B$. In our system, we observe a direct proportionality between the variance and average (Fig. 2D) over a 100-fold range of averages, so that $\sigma_n^2 \approx A \langle n \rangle$ with no significant extrinsic noise. A changing extrinsic noise should instead make σ_n^2 a more complicated function of $\langle n \rangle$. It is this strict proportionality which makes it possible to interpret our data without considering other substantial sources of noise. The method is not universally applicable, but neither is the dual reporter strategy that relies on true independence between the two reporters, excluding most nonlinear mechanisms.

Several previous studies reported substantial extrinsic fluctuations (Elowitz et al., 2002; Pedraza and van Oudenaarden, 2005; Raser and O'Shea, 2004). Why were they not observed here? A possible explanation is that the different findings reflect the difference between the systems: Extrinsic fluctuations have been shown to be sensitive to the exact genetic context and experimental parameters (Elowitz et al., 2002). We also note that by looking at stabilized RNAs rather than proteins, the present method removes many sources of extrinsic fluctuations, such as fluctuations in ribosomes or nuclease concentrations. Our measurements are also obtained for parameter values where small-number mRNA fluctuations are relatively large, and are consistent with the findings of other studies where extrinsic fluctuations appeared to be small (Ozbudak et al., 2002).

Supplemental References

Blake, W. J., M, K. A., Cantor, C. R., and Collins, J. J. (2003). Noise in eukaryotic gene expression. Nature *422*, 633-637.

Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P. A., Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (2002). A monomeric red fluorescent protein. Proc Natl Acad Sci U S A *99*, 7877-7882.

Chen, Y., Golding, I., Sawai, S., Guo, L., and Cox, E. C. (2005). Population fitness and the regulation of Escherichia coli genes by bacterial viruses. PLoS Biol *3*, e229.

Dodd, I. B., Perkins, A. J., Tsemitsidis, D., and Egan, J. B. (2001). Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. Genes Dev *15*, 3013-3022.

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.

Golding, I., and Cox, E. C. (2004). RNA dynamics in live Escherichia coli cells. Proc Natl Acad Sci U S A *101*, 11310-11315.

Hendrix, R. W. (1983). Lambda II (Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory).

Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res *25*, 1203-1210.

Miller, J. H. (1992). A short course in bacterial genetics : a laboratory manual and handbook for Escherichia coli and related bacteria (Plainview, N.Y., Cold Spring Harbor Laboratory Press).

Neidhardt, F. C., Ingraham, J. L., and Schaechter, M. (1990). Physiology of the bacterial cell : a molecular approach (Sunderland, Mass., Sinauer Associates).

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., and van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. Nat Genet *31*, 69-73.

Pedraza, J. M., and van Oudenaarden, A. (2005). Noise propagation in gene networks. Science *307*, 1965-1969.

Ptashne, M. (1992). A genetic switch : phage lambda and higher organisms, 2nd edn (Cambridge, Mass., Cell Press : Blackwell Scientific Publications).

Raser, J. M., and O'Shea, E. K. (2004). Control of stochasticity in eukaryotic gene expression. Science *304*, 1811-1814.

Svenningsen, S. L., Costantino, N., Court, D. L., and Adhya, S. (2005). On the role of Cro in lambda prophage induction. Proc Natl Acad Sci U S A *102*, 4465-4469.

Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci U S A *99*, 12795-12800.