Previews

And the Noise Played On: Stochastic Gene Expression and HIV-1 Infection

Stochastic gene expression has been implicated in a variety of cellular processes, including cell differentiation and disease. In this issue of *Cell*, Weinberger et al. (2005) take an integrated computational-experimental approach to study the Tat transactivation feedback loop of HIV-1. They show that fluctuations in a key regulator, Tat, in an isogenic population of infected cells result in two distinct expression states corresponding to latent and productive HIV-1 infection. These findings demonstrate the importance of stochastic gene expression in molecular "decision-making."

The traditional view of cellular processes as being largely deterministic is changing in the wake of an increasing number of studies demonstrating that stochastic gene expression influences the phenotype of cells. Phenotypic heterogeneity has long been observed in populations of genetically identical cells exposed to the same environmental conditions. Recent studies of gene expression in prokaryotes (Ozbudak et al., 2002; Elowitz et al., 2002) and in single-celled eukaryotes (Blake et al., 2003; Raser and O'Shea, 2004) are beginning to detail the physical mechanisms that lead to such variability. In general, stochastic gene expression can be attributed to the random, or "noisy," nature of the biochemical processes involved in transcription and translation (Kaern et al., 2005).

Seminal theoretical work by Harley McAdams and Adam Arkin (McAdams and Arkin, 1997) explored the mechanistic basis for stochastic gene expression and showed that the bifurcation of clonal cells into distinct phenotypic states can result solely from such molecular "noise." Subsequently, Arkin et al. (1998) applied these ideas to a specific biological system—the bacteriophage lambda lysis/lysogeny decision circuit. They demonstrated, through an insightful mathematical model, that fluctuations in the rate of production of key regulators of this switch can be amplified by positive feedback resulting in either a lytic or lysogenic state. Importantly, stochastic gene expression, combined with positive feedback in the lambda circuit, was found to be critical for the behavior of the switch. Further experimental work in *Escherichia coli* (Isaacs et al., 2003), in the budding yeast *Saccharomyces cerevisiae* (Becskei et al., 2001), and in mammalian cells (Kramer and Fussenegger, 2005) confirmed that positive feedback coupled to molecular noise was sufficient to generate distinct phenotypic states in a population of clonal cells.

In this issue of *Cell*, Arkin, together with Leor Weinberger and colleagues, extends these ideas from bacteriophage to retroviruses with an exploration of the role of stochasticity within the Tat transactivation loop of HIV-1. Their aim was to determine whether HIV-1 latency can be explained by stochasticity in the expression of a key regulator, the Tat transactivator. After infection of host cells (usually T lymphocytes) by HIV-1, basal expression of proviral genes, including *Tat*, enables the formation of a protein complex containing Tat that enters the host cell nucleus and facilitates proviral gene expression. This initiates a positive feedback loop, resulting in productive infection, retroviral amplification, and death of the host cell. Alternatively, the HIV-1 provirus can lie dormant in a latent state. The authors hypothesized that the cellular decision to follow one of these two mutually exclusive expression modes is a stochastic one, mediated by random fluctuations in the production of the Tat transactivator. Cells that, by chance, exhibit greater initial bursts of Tat protein production trigger the positive feedback loop resulting in productive infection, whereas cells with lower levels of Tat production enter a latent infection state.

To test this hypothesis, Weinberger et al. (2005) constructed a lentiviral vector system composed only of the Tat transactivation feedback loop. The vector contained the HIV-1 long terminal repeat (LTR) promoter driving the expression of a green fluorescent protein (GFP) and the Tat transactivator separated by an internal ribosome entry sequence (IRES). This LTR-GFP-IRES-Tat (LGIT) vector represents the minimum circuitry involved in the Tat transactivation loop. A control vector was also constructed in which the Tat transactivator was omitted (LTR-GFP, or LG), thus disrupting the feedback mechanism. The authors also developed a computational model of Tat transactivation to determine critical features that would guide their experimental implementation. Key among these was the need to mimic the physiologically low levels of Tat expression upon initial entry of HIV-1 into the host cell. In the model, low levels of Tat protein represent an unstable state that results in transition to either high Tat expression, through activation of the Tat transactivation loop, or to attenuated Tat expression and latency (see Figure 1). This led the researchers to design their model lentiviral vector with the Tat protein expressed as the second cistron in the LGIT vector (leading to reduced Tat expression).

The investigators infected Jurkat T lymphocytes at a low multiplicity of infection with the LGIT or LG lentiviral vectors and used flow cytometric measurements of GFP levels to monitor viral gene expression. LGIT-infected cells exhibited distinct degrees of GFP fluorescence described as *Bright*, *Dim*, *Mid*, and *Off*; control LG-infected cells exhibited low GFP fluorescence (classified as either *Dim* or *Mid*). Cells were sorted from each of the LGIT fluorescence regions to determine whether each of these states of Tat activity was stable. At the two extremes, sorted *Off* cells remained *Off*, whereas sorted *Bright* cells slowly relaxed to the *Off* region. Cells sorted from the *Mid* region transitioned to the *Bright* region, and, most interestingly, cells from the *Dim* region transitioned into the *Bright* or *Off* regions.
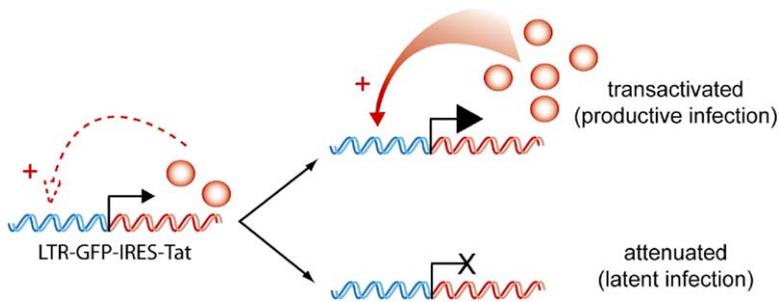
Figure 1. The Tat Transactivation-Positive Feedback Loop

Stochastic fluctuations in the production of the HIV-1 Tat protein within a model positive feedback loop (LTR-GFP-IRES-Tat) result in two distinct phenotypes. Cells that, by chance, exhibit a large burst of Tat protein production initiate the Tat transactivation loop leading to the production of high levels of Tat protein, corresponding to productive HIV-1 infection. Alternatively, in other cells, basal levels of Tat protein production may not be sufficient to initiate the positive feedback loop, corresponding to latent HIV-1 infection.

Because the *Dim* region corresponds to low levels of Tat production, complete loss of *Dim* region cells supports the initial model predictions that low Tat levels are unstable, leading to either high production of viral genes (*Bright*) or latent infection (*Off*). Notably, none of the control LG-infected cells transitioned from a *Dim* or *Mid* fluorescent state into a *Bright* fluorescent state, consistent with the notion that positive feedback is the critical factor mediating the response of LGIT-infected cells.

Clones were subsequently isolated from the *Dim* region of the LGIT-infected cell population to observe the stability of the *Dim* region in a population of cells with similar basal Tat expression rates. Consistent with observations in polyclonal populations, some clones isolated from the *Dim* region transitioned into *Bright* or *Off* fluorescent states, and, strikingly, a large number exhibited two subpopulations representing distinct *Bright* and *Off* phenotypes. The authors termed this phenomenon—that is, the emergence of distinct phenotypes from a single genotype— phenotypic bifurcation or PheB. In contrast to clones isolated from LGIT *Dim* regions, those isolated from LGIT *Bright* regions or LG *Dim* regions did not exhibit PheB. It is important to note that *Off* cells within the clonal PheB population could be induced to the *Bright* state by various means, such as artificially increasing the level of Tat protein, demonstrating that *Off* cells were capable of initiating the positive feedback loop.

The authors performed an extensive list of controls to rule out other possible sources of PheB, including extrinsic noise factors (for example, cell size and cell division) and potential position effects resulting from the chromosomal location of viral integration. Interestingly, there seemed to be preferential integration near human endogenous retroviral long-term repeats, a phenomenon that, as the authors state, warrants further exploration. After analysis of various control data, the authors conclude that the most likely explanation for PheB is the stochastic expression of the Tat protein within the Tat transactivation loop.

These experiments, motivated by the initial in silico model of the Tat transactivation loop, led to further model refinement, with a focus on elucidating molecular properties of the system that directly account for PheB. The authors used stochastic simulations of the biochemical reactions involved in the Tat transactivation loop. A simple representation of the feedback system resulted in a series of models that were then analyzed and compared to experimental findings. Following these model-experiment comparisons, the investigators converged on a single in silico representation with key features including transactivation delays (mediated by cyclical acetylation and deacetylation of Tat) and, as previously mentioned, low basal levels of Tat expression, with the establishment of an initial Tat concentration prior to proviral integration.

The refined model was then used to make a prediction about PheB cell behavior under conditions of a mutant Tat circuit. Specifically, the model predicted that, if the rate of Tat acetylation was decreased (thereby decreasing transactivation strength), the rate at which the *Bright* PheB cells relaxed to the *Off* state would increase. This relaxation is not necessarily relevant physiologically, as *Bright* cells would normally represent productively infected cells that are effectively killed. Nevertheless, it offered the authors the opportunity to test an insightful model prediction. The authors did so by using a mutated version of Tat with attenuated acetylation, and they observed faster relaxation from the *Bright* state to the *Off* state, consistent with their model prediction.

There is growing interest in integrating computational models with experimental data in molecular biology. Much of the prior work in this area has focused on generating a model and showing that it can capture features of existing experimental data. Rarely have the models been used to make a series of predictions that motivate new experiments and provide new insights. An attractive feature of the present study is that in silico modeling was the starting point, motivating the initial set of experiments. Moreover, the authors used modeling throughout the study to interpret the collected data and to generate testable predictions, while also using the experimental data to validate and refine the models.

The work of Weinberger et al. (2005) represents an important step in moving from studies that elucidate the origins of stochasticity in gene expression to those that investigate the consequences of such molecular noise on cellular function. The authors describe their PheB observations in terms of a possible scenario in which HIV-1 can "hedge its bets" by having an inherent ability to proceed to either latency or viral production— analogous to similar arguments made for bacteriophage lambda's lysis/lysogeny decision (Arkin et al., 1998). This intriguing notion still needs to be tested ex-

perimentally, and, more broadly, much work remains to be done to understand the functional role that gene expression noise potentially plays in the progression of disease.

William J. Blake and James J. Collins
Center for BioDynamics and Department
  of Biomedical Engineering
Boston University
Boston, Massachusetts 02215

Selected Reading

Arkin, A., Ross, J., and McAdams, H.H. (1998). Genetics *149*, 1633–1648.

Becskei, A., Seraphin, B., and Serrano, L. (2001). EMBO J. *20*, 2528–2535.

Blake, W.J., Kaern, M., Cantor, C.R., and Collins, J.J. (2003). Nature *422*, 633–637.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Science *297*, 1183–1186.

Isaacs, F.J., Hasty, J., Cantor, C.R., and Collins, J.J. (2003). Proc. Natl. Acad. Sci. USA *100*, 7714–7719.

Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Nat. Rev. Genet. *6*, 451–464.

Kramer, B.P., and Fussenegger, M. (2005). Proc. Natl. Acad. Sci. USA *102*, 9517–9522.

McAdams, H.H., and Arkin, A. (1997). Proc. Natl. Acad. Sci. USA *94*, 814–819.

Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., and van Oudenaarden, A. (2002). Nat. Genet. *31*, 69–73.

Raser, J.M., and O'Shea, E.K. (2004). Science *304*, 1811–1814.

Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D.V. (2005). Cell *121*, this issue, 169–182.