Biophysical Journal, Volume 98

Supporting Material

Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression

Abhyudai Singh, Brandon Razooky, Michael Simpson, Chris Cox, and Leor S. Weinberger

Construction of clonal cell lines and flow cytometry analysis

An HIV-1 derived lentiviral vector encoding a 2.5 hour half-life d2EGFP reporter gene (Clontech, San Jose, CA), was constructed via standard cloning techniques, packaged in 293T Human Embryonic Kidney cells, and virus harvested as previously described (Dull et al., 1998). This lentiviral vector is referred to as LTR-GFP. Jurkat human T lymphocytes were infected with harvested LTR-GFP virus at a low multiplicity of infection (MOI ≈ 0.05) to minimize multiple integration events. Jurkat cells were grown in RPMI 1640 (supplemented with L-glutamine, 10% fetal bovine serum, and 1% penicillin-streptomycin) in a humidified environment at 37 degrees Celsius and maintained by passage between 2×10^5 and 2×10^6 cells/mL. GFP positive cells were single-cell sorted into 96-well plates by FACS and allowed to expand over 3 weeks into clonal populations creating a final library of 30 distinct LTR-GFP clones. After expansion, GFP fluorescence of individual clones was measured by flow cytometry on a FACSCalibur$^{\text{TM}}$(BD Biosciences, San Jose, CA) and statistically analyzed using FlowJo$^{\text{TM}}$((Treestar Inc., Ashland, Oregon). For induction with TNF-α, clones were treated at a final concentration of 10 ng/ul human TNF-α (Sigma-Aldrich Corporation, St. Louis, MO), and fluorescence measured at 16 hours post TNF-α addition.

EGFP Calibration and stochastic analysis of the two-state model

Fluorescence intensity was converted to Molecular Equivalent of Soluble Fluorophores (MESF), which is a measure of GFP abundance, using EGFP Calibration Beads (BD Biosciences, Clontech, San Jose, CA). GFP abundance of a given clone was calculated by subtracting the mean background Jurkat auto-fluorescence and then multiplying by the measured MESF scaling-factor of 3000 molecules per unit fluorescence intensity (data not shown).

To explain the high variability in HIV-1 gene expression we considered a two-state promoter model where the LTR fluctuates between an inactive and active state with rates kon, koff and transcriptional elongation only occurs from the active state at a rate T. In such two-state models (Raj et al., 2006, Simpson et al. 2004, Nakanishi et al. 2008) mRNAs are created in bursts during promoter transitions from inactive to active state, with kon and T/koff denoting the frequency and the average size of the transcriptional bursts, respectively. The two-state promoter model is given by the following set of chemical reactions:

$$
\begin{aligned}
&G_{off} \xrightarrow{k_{on}} G_{on}, \quad G_{on} \xrightarrow{k_{off}} G_{off} \\
&G_{on} \xrightarrow{T} G_{on} + mRNA, \quad mRNA \xrightarrow{dm} \emptyset \\
&M \xrightarrow{L} M + GFP \\
&GFP \xrightarrow{dp} \emptyset.
\end{aligned}
$$

In the stochastic formulation of this model each reaction is a probabilistic

event, and random firings of individual reactions change population counts based on the stoichiometry of that reaction (Gillespie, 1976). Let $\mathbf{x}(t)$ and $\mathbf{z}(t)$, denote the population count of GFP and its mRNA at time t, respectively. The state of the promoter is represented by $\mathbf{g}(t)$, with $\mathbf{g}(t) = 1$ and $\mathbf{g}(t) = 0$ denoting that the promoter is active or inactive, respectively. A convenient way to represent the time evolution of the vector $\mathbf{y} = [\mathbf{g}, \mathbf{z}, \mathbf{x}]^T$ is through a set of reset maps with corresponding propensity functions (Singh et al., 2009). This representation of the stochastic process $\mathbf{y}(t)$ is equivalent to the Chemical Master equation, but from which it is easier to derive the differential equations for the statistical moments (Davis, 1993). The above model can be represented by the following six reset maps:

$$\mathbf{y} \mapsto \phi_1(\mathbf{y}) = \begin{bmatrix} \mathbf{g}+1 \\ \mathbf{z} \\ \mathbf{x} \end{bmatrix}, \quad \mathbf{y} \mapsto \phi_2(\mathbf{y}) = \begin{bmatrix} \mathbf{g}-1 \\ \mathbf{z} \\ \mathbf{x} \end{bmatrix}, \quad \mathbf{y} \mapsto \phi_3(\mathbf{y}) = \begin{bmatrix} \mathbf{g} \\ \mathbf{z}+1 \\ \mathbf{x} \end{bmatrix} \tag{1a}$$

$$\mathbf{y} \mapsto \phi_4(\mathbf{y}) = \begin{bmatrix} \mathbf{g} \\ \mathbf{z}-1 \\ \mathbf{x} \end{bmatrix}, \quad \mathbf{y} \mapsto \phi_5(\mathbf{y}) = \begin{bmatrix} \mathbf{g} \\ \mathbf{z} \\ \mathbf{x}+1 \end{bmatrix}, \quad \mathbf{y} \mapsto \phi_6(\mathbf{y}) = \begin{bmatrix} \mathbf{g} \\ \mathbf{z} \\ \mathbf{x}-1 \end{bmatrix} \tag{1b}$$

with corresponding propensity functions given by

$$\lambda_1(\mathbf{y}) = k_{on}(1-\mathbf{g}),\ \lambda_2(\mathbf{y}) = k_{off}\ \mathbf{g},\ \lambda_3(\mathbf{y}) = T\ \mathbf{g},\ \lambda_4(\mathbf{y}) = dm\ \mathbf{z}, \tag{2a}$$

$$\lambda_5(\mathbf{y}) = L\ \mathbf{z}, \lambda_6(\mathbf{y}) = dp\ \mathbf{x}. \tag{2b}$$

In particular, whenever a reaction takes place its reset map is "activated" and the population count is reset accordingly, furthermore, the probability of the activation taking place in an "infinitesimal" time interval $(t, t + dt]$ is given by $\lambda_i(\mathbf{y})dt$. To derive the temporal dynamics of the statistical moments of $\mathbf{y}$ the following equation is used (Singh et al., 2009):

$$\frac{d\mathbf{E}[\psi(\mathbf{y})]}{dt} = \mathbf{E}\left[\sum_{i=1}^{6}\left(\psi(\phi_i(\mathbf{y})) - \psi(\mathbf{y})\right)\lambda_i(\mathbf{y})\right] \tag{3}$$

where $\psi(\mathbf{y})$ is an arbitrary differentiable function and $\mathbf{E}$ refers to the expected value. Using (3), the time derivative of vector μ defined by

$$\mu = [\mathbf{E}[\mathbf{g}], \mathbf{E}[\mathbf{z}], \mathbf{E}[\mathbf{x}], \mathbf{E}[\mathbf{g}^2], \mathbf{E}[\mathbf{z}^2], \mathbf{E}[\mathbf{x}^2], \mathbf{E}[\mathbf{gz}], \mathbf{E}[\mathbf{zx}], \mathbf{E}[\mathbf{gx}]]^T \tag{4}$$

is given by

$$\frac{d\mu}{dt} = \bar{\mathbf{a}} + \mathbf{A}\mu \tag{5}$$

for an appropriate constant vector $\bar{\mathbf{a}}$ and a constant matrix $\mathbf{A}$. Steady-state analysis of (5) using Mathematica, yields

$$CV^2 = \frac{dm\ dp\left[\frac{k_{off}(k_{off}+k_{on}+dm+dp)}{(k_{off}+k_{on}+dm)(k_{off}+k_{on}+dp)} + \frac{(k_{off}+k_{on})(L+dm+dp)}{L\ T}\right]}{(dm+dp)k_{on}}, \tag{6a}$$

$$< GFP >= \frac{k_{on}\ L\ T}{(k_{off}+k_{on})\ dm\ dp}, \tag{6b}$$

where CV and $< GFP >$ refer to the steady-state coefficient of variation and mean protein count, respectively. Under the assumptions $k_{off} \gg k_{on}$,

$k_{off} \gg dm$, $k_{off} \gg dp$ and $L \gg dm + dp$ the above equation reduces to

$$CV^2 = \frac{\frac{T}{koff} + 1}{\mathbf{E}[\mathbf{x}(\infty)]} \frac{L}{dm + dp}, \quad < GFP >= \frac{k_{on}\ L\ T}{k_{off}\ dm\ dp}. \tag{7}$$

For an appropriate fixed average transcriptional burst size $T/koff$, increasing the burst frequency kon will give a similar scaling of noise with mean protein levels as observed in Eq. 1 and 2 of the paper.

Monte Carlo simulations in Fig 1c

To obtain GFP histograms corresponding to the constitutive gene expression model, Monte Carlo simulations of the model were done using the Stochastic Simulation Algorithm (Gillespie, 1976) with the following parameters: 2.5-hour GFP half-life; 3 hours GFP mRNA half-life (Raj et al., 2006), translation rate of $L \approx 2500\ hr^{-1}$ and the transcription rate was chosen as

$$T = \frac{< GFP > dm\ dp}{L} \tag{8}$$

where $< GFP >$ is the mean GFP population count of the Jurkat clone and dp and dm are the GFP and GFP mRNA half-life's, respectively. For all Monte Carlo simulations an initial condition of 10 mRNA transcripts and 30,000 GFP molecules were used. GFP histograms were constructed by performing 30,000 different runs of the Stochastic Simulation Algorithm with a simulation time of 25 hours.

To obtain GFP histograms corresponding to the two-state promoter model, Monte Carlo simulations of the model were done using the fol-

lowing parameters: 2.5-hour GFP half-life, 3 hours GFP mRNA half-life, translation rate of $L \approx 2500\ hr^{-1}$ and a transcriptional elongation rate of $T \approx 60\ hr^{-1}$. For a clone with a given $< GFP >$ and coefficient of variation CV, the rates kon and $koff$ for the two-state promoter model were determined by solving Eq. 6. This resulted in an average transcriptional burst size ($T/koff$) of 2 for both clones considered in Fig 1c, and a burst frequency (kon) of 0.3 hr^{-1} and 0.7 hr^{-1} for the clone on the left and right in Fig 1c, respectively.

References

Davis MHA. Markov models and Optimization. Chapman and Hall, 1993.

Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, Trono D, Naldini L. A third-generation lentivirus vector with a conditional packaging system. J Virol. 1998, 72:8463-71.

Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 1976, 22:403434.

Nakanishi, H, Mitarai N, Sneppen, K. Dynamical analysis on gene activity in the presence of repressors and an interfering promoter. Biophysical Journal 2008. 95:4228-4240.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S: Stochastic mRNA synthesis in mammalian cells. PLoS Biol 2006, 4:e309.

Singh A, Hespanha JP. Optimal feedback strength for noise suppression in auto-regulatory gene networks. Biophysical Journal 2009, 96:4013-4023.

Simpson ML, Cox CD, Sayler GS. Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. J Theor Biol. 2004, 229:383-94.