PNAS
www.pnas.org

Supplementary Information for

Probabilistic Control of HIV Latency and Transactivation by the Tat Gene Circuit

Youfang Cao[1,2,*,#], Xue Lei[3], Ruy M. Ribeiro[1,4], Alan S. Perelson[1], Jie Liang[3,*]

[1]Theoretical Biology and Biophysics (T-6), Los Alamos National Laboratory, Los Alamos, NM 87545
[2]Center for Nonlinear Studies (CNLS), Los Alamos National Laboratory, Los Alamos, NM 87545
[3]Department of Bioengineering, University of Illinois at Chicago, Chicago IL, 60607
[4]Laboratório de Biomatemática, Faculdade de Medicina, Universidade de Lisboa. Av. Professor Egas Moniz, 1649–028 Lisboa, Portugal
[#] Present address: Quantitative Pharmacology and Pharmacometrics (QP2), Department of Pharmacokinetics, Pharmacodynamics and Drug Metabolism (PPDM), Merck & Co., Kenilworth, NJ. This work was completed before YC joined Merck.

Corresponding authors:
Youfang Cao, Email: youfang.cao@merck.com
Jie Liang, Email: jliang@uic.edu

This PDF file includes:

Supplementary text
Tables S1 to S4
Figs. S1 to S9
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1811195115

Supplementary Information Text

The Accurate Chemical Master Equation (ACME) method

By taking advantage of the strong coupling among reactions, the ACME method can dramatically reduce the size of the required state space by a factor of $O(n!)$, where n is the number of distinct molecular species in the network (1). This dramatic state space reduction allows exact computation of the probability landscapes of many stochastic networks that were otherwise infeasible (1-4). The ACME method can be used to study broad issues in stochastic biological networks and provides a powerful modeling framework for identifying effective therapeutic targets from the underlying gene regulatory networks. The open source software of ACME is available at http://BioACME.org/. We use the ACME method to study the full stochastic behavior of the HIV Tat circuit. The Tat circuit is an open reaction network with active production and decay of molecular species. We first partition the network into two molecular equivalence groups (MEGs), one contains the Tat mRNA and the other contains the Tat protein (4). We then assign two finite buffers to the two MEGs, respectively, to construct the state space (1, 4). We determine the buffer sizes for the mRNA MEG is 50, and the buffer size for Tat protein MEG is 2000. These buffer sizes ensure a state space truncation error smaller than 10^{-10}, *i.e.* the loss of probability mass in the steady state is less than 10^{-10}. We construct the discrete finite state space for the HIV Tat circuit from the reaction network (Table S1). The state space of the Tat circuit contains 306,051 individual states. We assume each HIV infected cell has only one copy of the HIV genome (5, 6). For each different set of parameter values we construct the transition rate

matrix for the Tat circuit from the state space, and each transition rate matrix contains 1,819,949 nonzero elements.

Computing the switching rates between cell phenotypes based on the first passage time distribution

To obtain the latency transactivation rate, we compute the first passage time distribution from latency to the transactivated state using the ACME method. We first define the latency and transactivated state of a cell according to its Tat copy number. If the Tat protein copy number is smaller than or equal to 34 and the Tat mRNA copy number is not larger than one, *i.e.* the cell state is smaller than the wild-type circuit separatrix state, then the cell is considered to be in the latent state, whereas if the Tat copy number is larger than 34 and the Tat mRNA copy number is larger than one, the cell is considered to be in the transactivated state. Note that in the wild-type Tat circuit, the total probability mass contain in the latent and the transactivated states in the steady state probability landscape is 0.9988985, while the probability mass contained in all other states is less than 1.1×10^{-3}. During latency transactivation, the cell can visit states that are neither in the latent states nor in the transactivated state, *i.e.* with > 34 Tat and 0 mRNA or < 34 Tat and ≥ 1 mRNA. However, since these states have small probabilities, the visit to these states will be transient, and the cell eventually goes to either the latent states or the transactivated states.

Computing the latency transactivation rate can be formalized as a first passage time problem in which the probability mass diffuses from latent states to the absorbed transactivated states in the multidimensional state space (7). The total probability mass that is absorbed in the transactivated states before time t is the cumulative first passage

probability at time t, which gives the cumulative first passage time distribution. To compute the cumulative probability distribution, we first aggregate all transactivated states into an absorbing state and construct a state space with the aggregated absorbing state. We then construct a transition rate matrix based on the state space with the absorbing transactivated state. The new rate matrix naturally retains all the transitions between the latent and transactivated boundary states. We assume that latently infected cells can have different copy numbers of Tat mRNA and protein according to the distribution of the latency peak in the steady state probability landscape. Therefore, we use the renormalized probability peak of latency in the steady state probability landscape of the wild-type Tat circuit as the initial distribution of probability mass over the state space. We then compute the cumulative probability absorbed in the transactivated states for any given time t. Previous studies suggested that the first passage time follows an exponential distribution in simpler systems (7, 8). Results from our computation has shown that the probability of the first passage time from latency to transactivation also follows an exponential distribution (1, 4). The cumulative probability of transactivation p exponentially increase from 0 to 1 over time (Figure S9A), and the probability mass that is not absorbed into the transactivated states $(1 - p)$ follows an exponential decay and is evidenced by a declining straight line in $\log(1 - p)$ (Figure S9B). We can obtain the latency transactivation rate by computing the slope of the straight line in Figure S9B. All above described computations are performed using the ACME method (1, 4). The same procedure can be used similarly to compute the rate of returning to latency from the activated state. The only different is that we need to aggregate all latent states (*i.e.* all states with Tat mRNA copy number < 1 and Tat protein copy number < 34) into an

absorbing state, instead of transactivated states. The rates of back-transitions in the perturbed circuits are completely negligible (Table S4).

Model and parameter values

The processes of HIV transcription and translation are controlled by multiple regulatory steps involving multiple host factors (9). In this study, we focused on the probabilistic control of HIV latency and activation through the stochastic Tat circuit. Specifically, Tat mRNAs are transcribed from the LTR at a basal transcription rate k_{b0}, Tat proteins are produced from the mRNA at translation rate k_t and they naturally decay at rates d_R and d_T. Tat proteins reversibly bind the TAR element at binding rate k_b and an unbinding rate k_u. Tat protein in the TTP complex is acetylated at rate k_a and deacetylated at rate k_d. Tat mRNAs can be produced from the acetylated complex at a transactivated rate k_f. We do not explicitly model phosphorylation and de-phosphorylation steps in this study. We assume that kinases are readily available and phosphorylation is not the rate-limiting step. Values of parameters for the Tat circuit (Table S1) are obtained from the experimental literature. For those parameters with no accurate experimental measurements, we adopt their values from previous modeling studies (10-13) and complement it with detailed exploration of the parameter space.

The default parameter values for the Tat circuit reaction network are included in Table S1. To accurately study the Tat circuit behavior, we first determined parameter values from literature in which experimental measurements of these parameters were made. For those parameters with no accurate experimental measurements, such as the basal transcription rate k_{b0}, and Tat acetylation and deacetylation rates k_a and k_d, we adopt the values from previous modeling studies (10, 11) and comprehensively explore

their parameter space to study the circuit behavior. The transactivated transcription rate of Tat was estimated to be 25-100 fold higher than the basal transcription rate (14, 15), but it was based on measurements from cell populations. In single cells, the value might vary.

Efthymiadis *et al.* estimated the rate of Tat protein import from the cytoplasm to the nucleus to be $k_i = 5.1 \times 10^{-3}\ s^{-1}$ (16). The mRNA export rate from the nucleus to the cytoplasm was estimated as 2.6 per hours, or $k_e = 7.2 \times 10^{-4}\ s^{-1}$ (14). The half-life of the unspliced HIV mRNA was estimated >4 hours in presence of Rev, and the decay rate is accordingly $d_R \leq 4.8 \times 10^{-5}\ s^{-1}$. Marcello and colleagues measured the Tat-induced elongation rate from a single integrated HIV virus using fluorescence recovery after photobleaching (FRAP) as about 50 kb $minute^{-1}$ or 833 bp s^{-1} (17, 18). The density of RNApII is estimated about one polymerase in every 270 bp (17). The fraction of *tat* mRNA in singly- and fully spliced HIV mRNA is about 0.025 (14). Considering the full length of the HIV genome is 9749 bp, we can then calculate the transcription rate of Tat as $0.025 \times \frac{833\ \text{bp}\ s^{-1}}{9749\ \text{bp}} \times \frac{9749\ \text{bp}}{270\ \text{bp}} \approx 0.1\ s^{-1}$. According to Wang *et al*., the Tat-TAR complex follows a first-order dissociation with a half-life of 41 seconds (19), which corresponds to a dissociation rate $k_u = \frac{\ln(2)}{41} = 0.017\ s^{-1}$.

We also modeled the Tat mRNA export rate k_e from the nucleus to the cytoplasm and the Tat protein import rate k_i from the cytoplasm to the nucleus by lumping the two rates with the translation rate k_t' and the Tat-TAR binding rate k_b', respectively. The measured mRNA export rate k_e is over an order of magnitude faster than the mRNA decay rate d_R, and the Tat protein import rate k_i is about two orders of magnitude faster than the Tat decay rate d_T (Table S1). Based on the fast transport between nucleus and

cytoplasm, we assume a quasi-steady state for the mRNA and Tat protein species, in which newly transcribed mRNAs are immediately available in the cytoplasm for translation, and the new Tat proteins are also immediately available in the nucleus for binding with TAR. We derive the new Tat translation rate modified by the mRNA export rate k_e and the mRNA decay rate d_R as: $k_t = k_t' \frac{k_e}{k_e + d_R}$, and the new Tat-TAR binding rate modified by the Tat protein import rate k_i and the Tat decay rate d_T: $k_b = k_b' \frac{k_i}{k_i + d_T}$. The simplified model allows us to more efficiently explore the parameter space. All our computations are based on the simplified model (Table S1).

SI Tables

Table S1. HIV Tat gene circuit reaction scheme and parameters.

Reactions	Parameters	Descriptions	Values	References
$LTR \xrightarrow{k_{b0}} LTR + mRNA$	k_{b0}	Basal transcription rate	1e-8 /s	(10, 11)
$mRNA \xrightarrow{k_t} Tat + mRNA$	k_t	Lumped translation rate	$k_t' \frac{k_e}{k_e + d_R}$	
$LTR + Tat \xrightarrow{k_b} TTP$	k_b	Lumped Tat-TAR binding rate	$k_b' \frac{k_i}{k_i + d_T}$	
	k_t'	Tat translation	1.32e-3 /s	(10)
	k_b'	Binding of Tat-TAR	1.5e-4 /molecule /s	(11)
	k_e	Tat mRNA export to cytoplasm	7.2e-4 /s	(14)
	k_i	Tat protein import to nucleus	5.1e-3 /s	(15, 16)
$TTP \xrightarrow{k_u} LTR + Tat$	k_u	Unbinding of Tat-TAR	1.7e-2 /s	(19)
$TTP \xrightarrow{k_a} acTTP$	k_a	Acetylation of elongation factor	1e-3 /s	(10)
$acTTP \xrightarrow{k_d} TTP$	k_d	Deacetylation of elongation factor	1.3e-1 /s	(10)
$acTTP \xrightarrow{k_f} LTR + Tat + mRNA$	k_f	Transactivated transcription rate	0.1 /s	(14, 15, 17)
$Tat \xrightarrow{d_T} \emptyset$	d_T	Decay rate of Tat protein	4.3e-5 /s	(12)
$mRNA \xrightarrow{d_R} \emptyset$	d_R	Decay rate of Tat mRNA	4.8e-5 /s	(20)

Table S2. Transition rates between the latent and transactivated cell phenotypes, and Tat copy number amplitudes. *The number of cells that are activated per day is computed assuming a latent reservoir size of 10^6 cells. For different latent reservoir size, the number of activation events needs to be scaled accordingly. **The 95-percentile of Tat copy number amplitude is computed from the renormalized distribution of the transactivation peak, which is the Tat copy number that covers 95% of the probability mass.

	Transactivation rate*	Rate of returning to latency	95-percentile of Tat copy number **
WT	$4.5 \times 10^{-4}\ d^{-1}$, $t_{\frac{1}{2}} =$ 51 mo, or 450 $cells\ d^{-1}$	0.065 d^{-1}, $t_{\frac{1}{2}} = 11\ d$	237 copies
No Tat-TAR binding	$3.4 \times 10^{-5}\ d^{-1}$, $t_{\frac{1}{2}} =$ 57 y, or 34 $cells\ d^{-1}$	0.53 d^{-1}, $t_{\frac{1}{2}} = 1.3\ d$	32 copies
No Tat acetylation	$3.4 \times 10^{-5}\ d^{-1}$, $t_{\frac{1}{2}} =$ 57 y, or 34 $cells\ d^{-1}$	4.1 d^{-1}, $t_{\frac{1}{2}} = 0.17\ d$	32 copies

Table S3. Summarized results for parameter perturbations.

	Can eliminate latency peak	Can eliminate activation peak	Can shift activation peak
Basal transcription rate	Yes, when increased	No	No
Tat mRNA export rate	No	Yes, when decreased	No
Tat protein import rate	No	Yes, when decreased	No
Tat acetylation rate	Yes, when increased	Yes, when decreased	Large increase
Tat deacetylation rate	No	Yes, when increased	Medium increase
Tat-TAR binding rate	Yes, when increased	Yes, when decreased	Small increase
Tat-TAR unbinding rate	Yes, when decreased	Yes, when increased	Small increase

Table S4. Comparisons of mean copy number of Tat protein in transactivated cells, the latency transactivation rate and the rate of returning to latency in four reaction perturbations. To effectively reverse latency and deplete the HIV latent reservoir, fast transactivation is desired.

	Mean Tat copy number	Latency transactivation rate	Rate of returning to latency
WT	134	$4.5 \times 10^{-4}\ d^{-1}$, $t_{\frac{1}{2}} = 51\ mo$	$0.065\ d^{-1}$
$k_a \times 100$, acetylation rate	2871	$7.7 \times 10^{-4}\ d^{-1}$, $t_{\frac{1}{2}} = 30\ mo$	$\sim 0\ d^{-1}$
$k_d \times 0.01$, deacetylation rate	463	$6.5 \times 10^{-4}\ d^{-1}$, $t_{\frac{1}{2}} = 35\ mo$	$6.1 \times 10^{-9}\ d^{-1}$
$k_b \times 100$, Tat-TAR binding rate	258	$8.3 \times 10^{-4}\ d^{-1}$, $t_{\frac{1}{2}} = 28\ mo$	$6.2 \times 10^{-7}\ d^{-1}$
$k_u \times 0.01$, Tat-TAR unbinding rate	256	$7.4 \times 10^{-4}\ d^{-1}$, $t_{\frac{1}{2}} = 31\ mo$	$8.4 \times 10^{-7}\ d^{-1}$

SI Figures

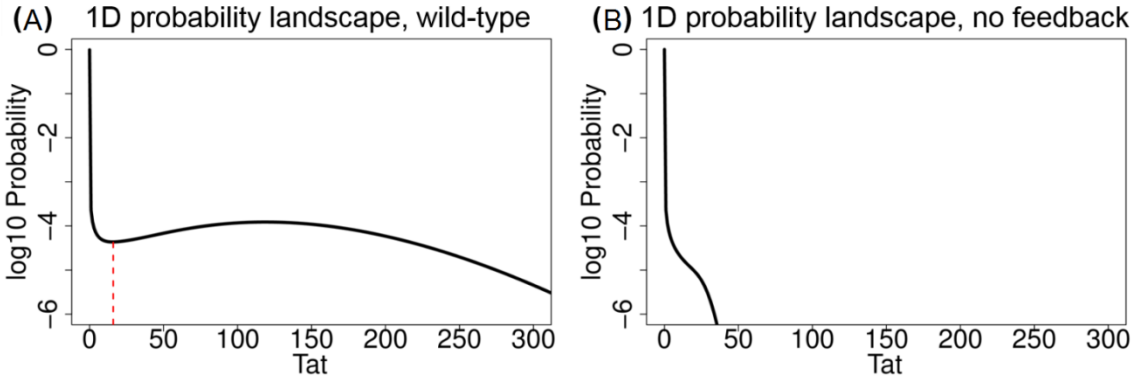

Figure S1. One-dimensional probability landscape for the wild-type Tat circuit and the circuits with no feedback. (A) 1D steady state probability landscape of the wild-type Tat circuit, with the separatrix indicated by the red dashed line. (B) 1D steady state probability landscape of the Tat circuit with no positive feedback (no Tat-TAR binding).

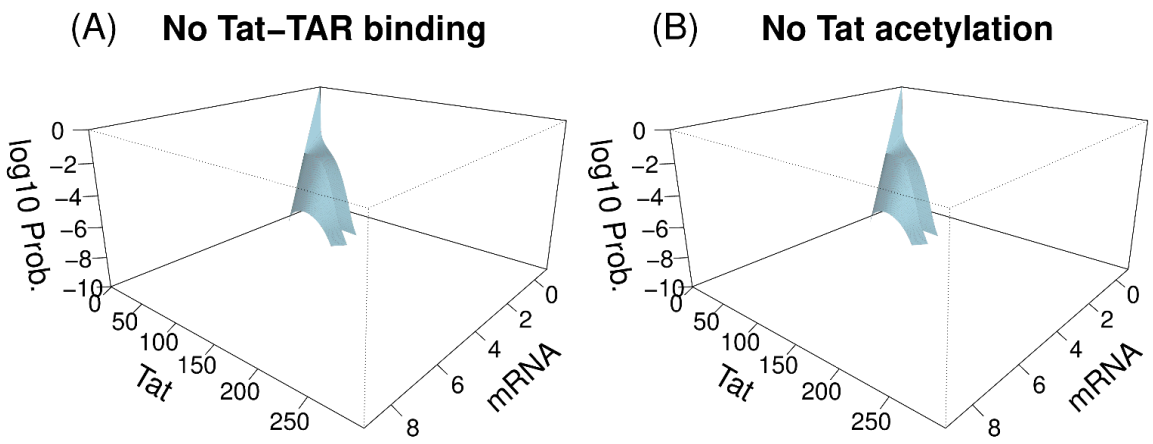

Figure S2. Steady state probability landscape of the HIV Tat circuit with disrupted feedbacks. (A) The steady state probability landscape with no Tat-TAR binding. There is only one probability peak at (0,0). (B) The steady state probability landscape with no Tat acetylation. There is also only one probability peak at (0,0).

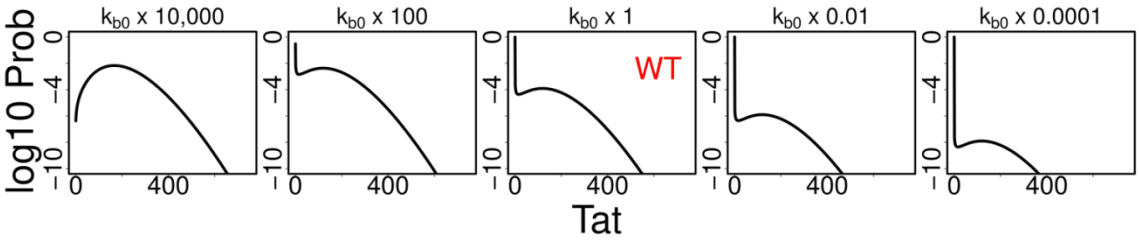

Figure S3. Effects of perturbing the basal transcription rate k_{b0} on the steady state probability landscape.

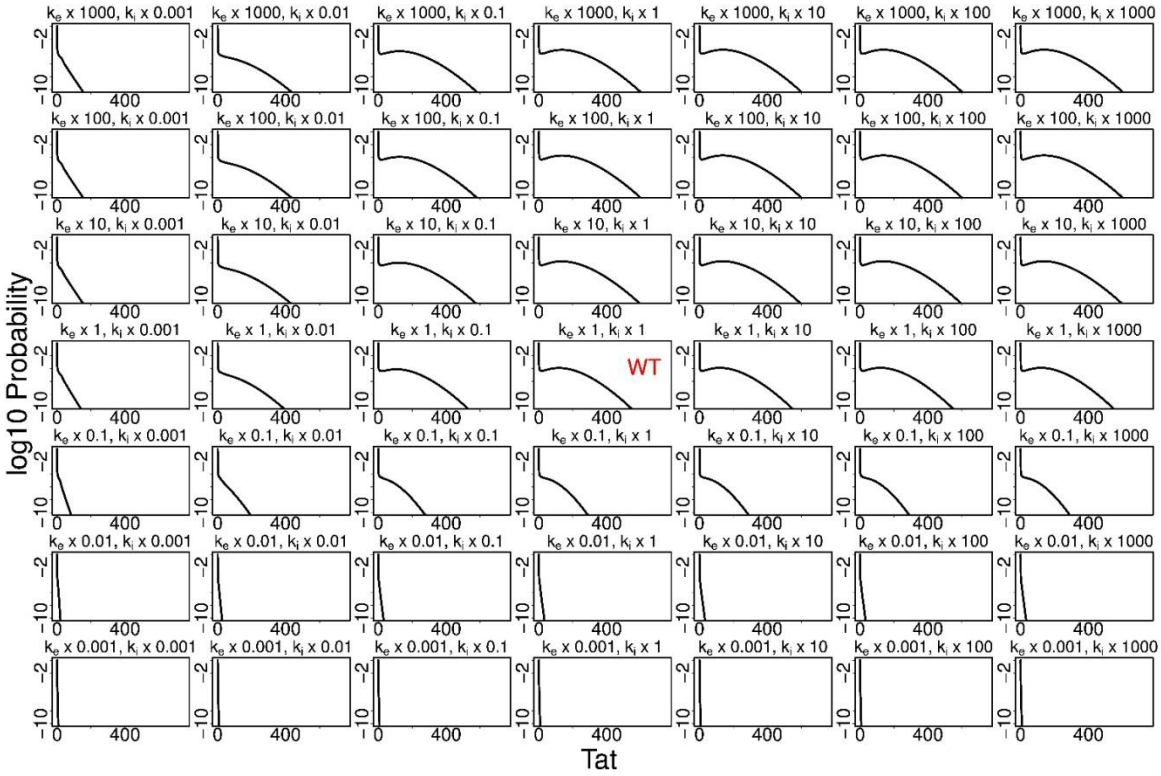

Figure S4. Effects of perturbing the nuclear export rate k_e of mRNA and import rate k_i of Tat protein on the steady state probability landscape.

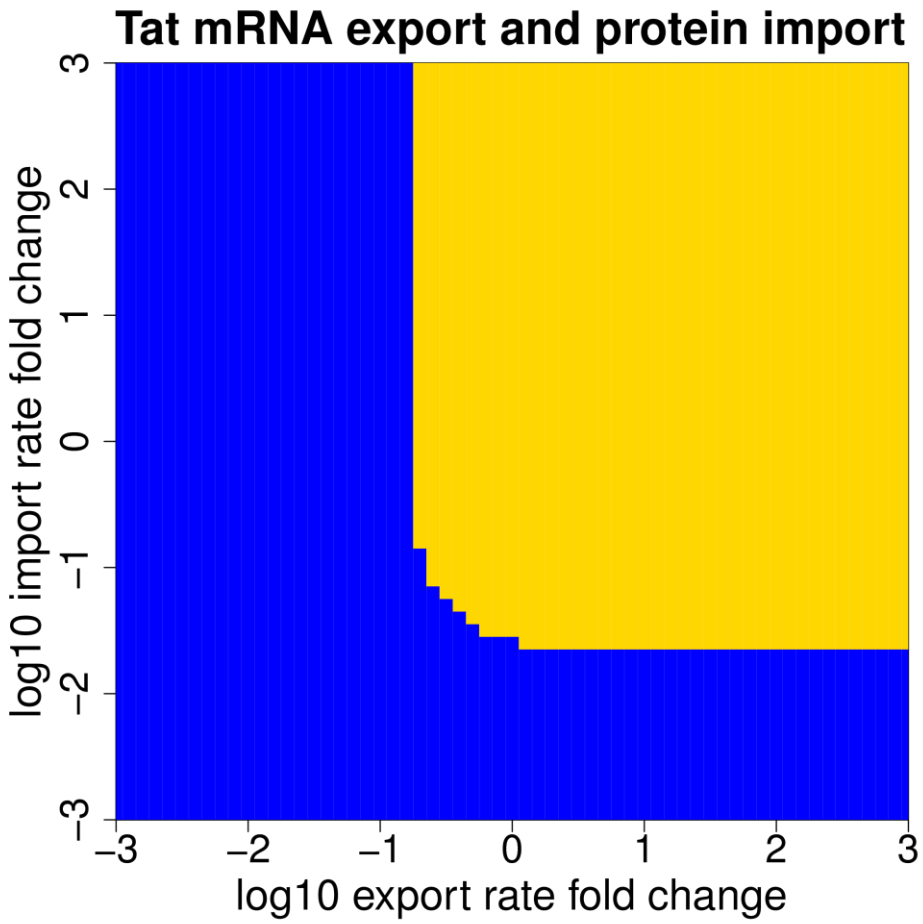

Figure S5. Phase diagram of phenotypes when perturbing the Tat mRNA and protein export and import rates. Blue region: unimodality at latency; yellow region: bimodality.

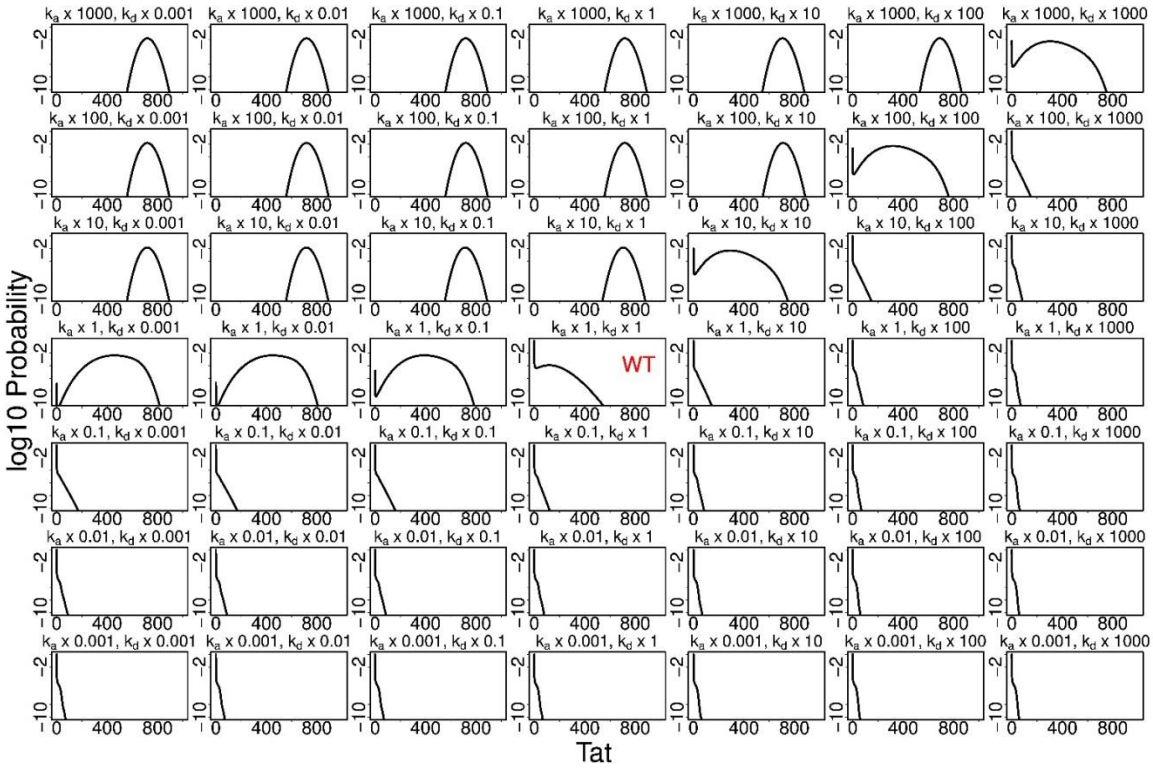

Figure S6. Effects of perturbing Tat acetylation and deacetylation rates k_a and k_d on the steady state probability landscape.

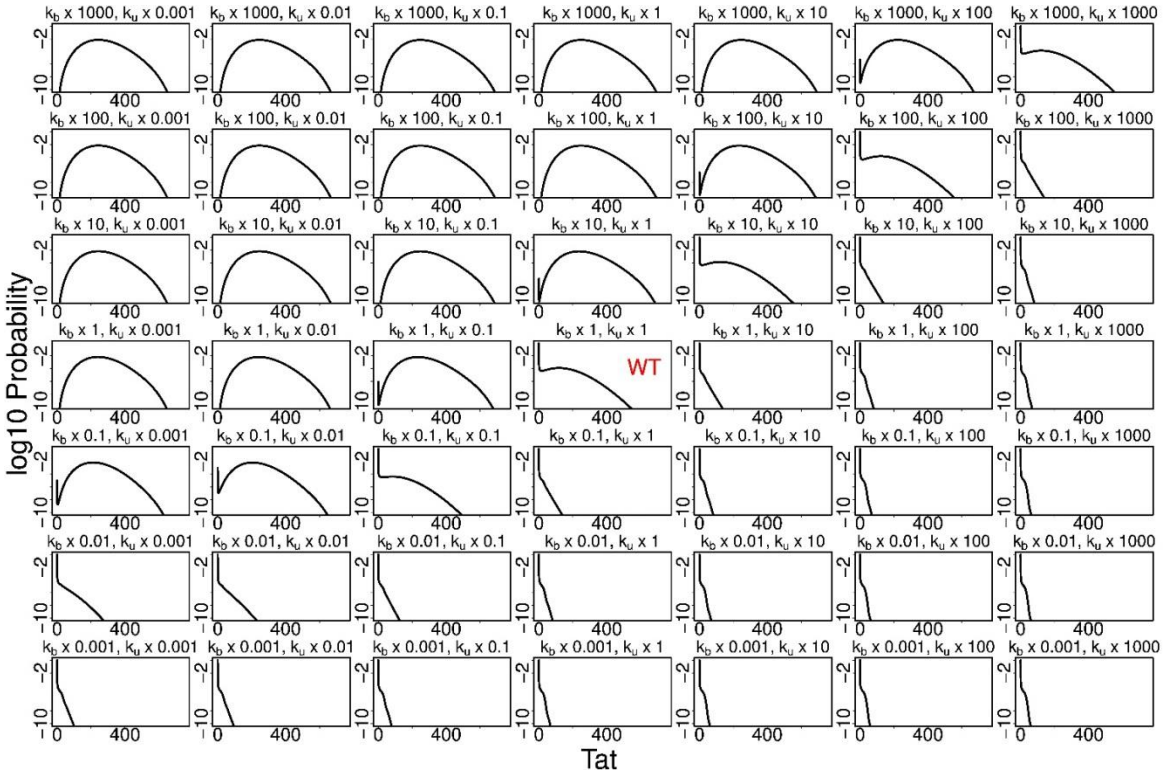

Figure S7. Effects of perturbing Tat-TAR binding and unbinding rates k_b and k_u on the steady state probability landscape.

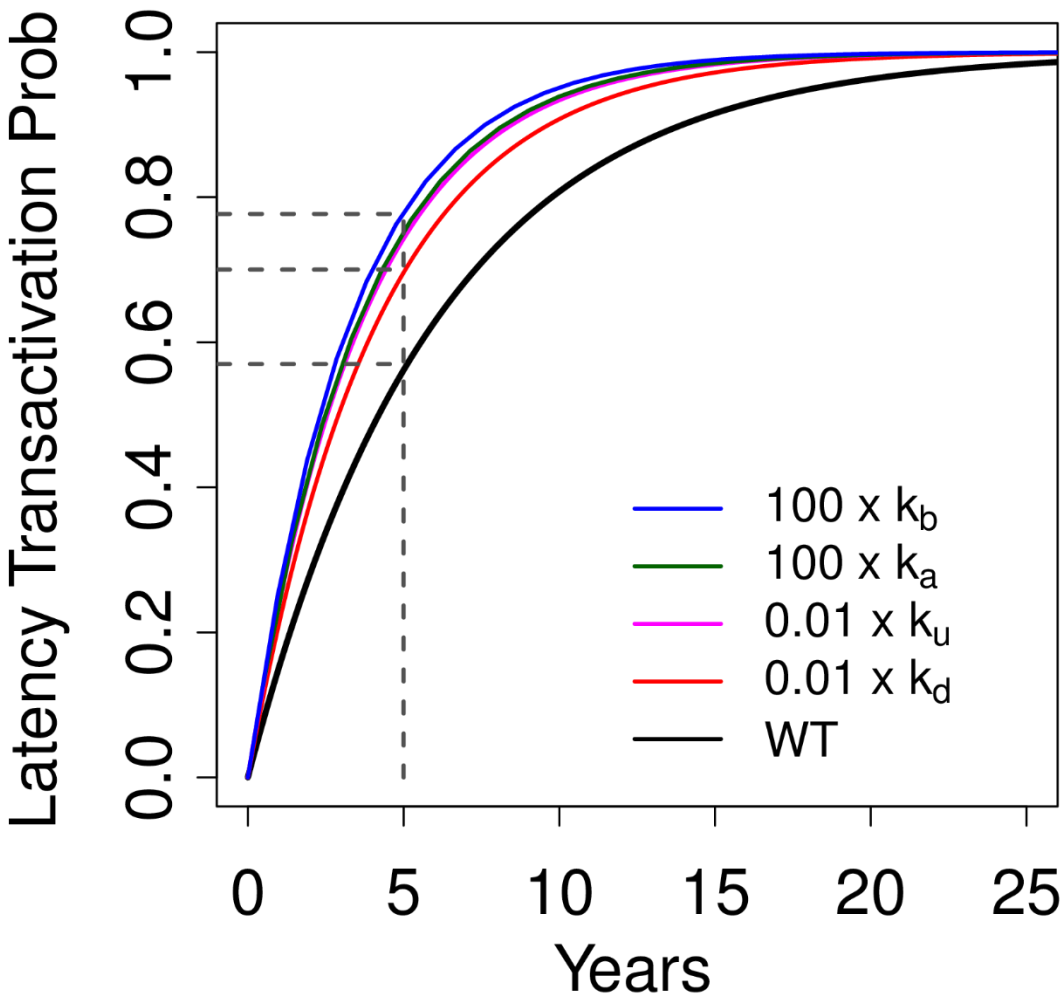

Figure S8. Cumulative latency transactivation probability distributions and latency transactivation rates with perturbations in each case. Black line: the wild-type Tat circuit. Red line: perturbed Tat deacetylation rate. Purple line: perturbed Tat-TAR unbinding rate. Green line: perturbed Tat acetylation rate. Blue line: perturbed Tat-TAR binding rate.

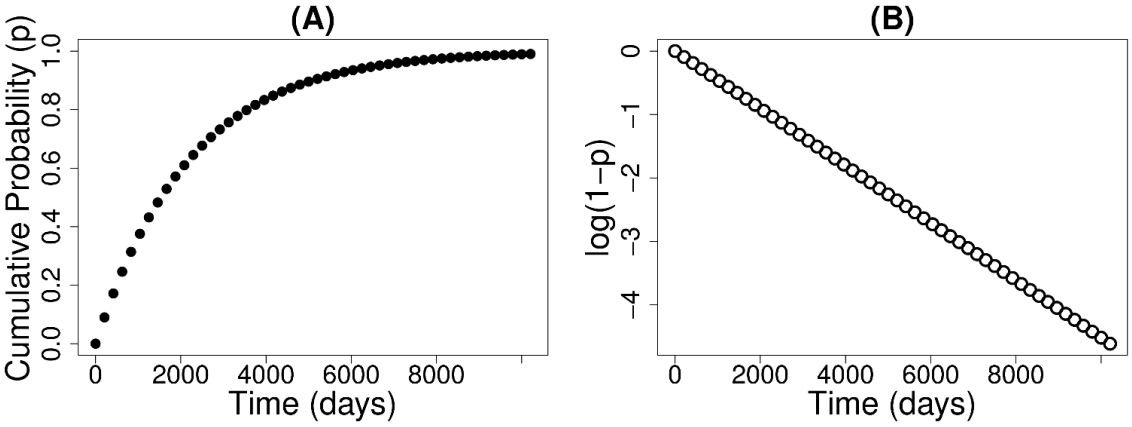

Figure S9. Cumulative probability distribution (A) of first passage time from latency to transactivation and the exponential decay (B) of probability mass that are not yet absorbed into the transactivated state.

References

1. Cao Y, Terebus A, & Liang J (2016) State space truncation with quantified errors for accurate solutions to discrete Chemical Master Equation. *Bull. Math. Biol.* 78(4):617-661.
2. Cao Y & Liang J (2008) Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability. *BMC Syst. Biol.* 2:30.
3. Cao Y, Lu HM, & Liang J (2010) Probability landscape of heritable and robust epigenetic state of lysogeny in phage lambda. *Proc. Natl. Acad. Sci. U. S. A.* 107(43):18445-18450.
4. Cao Y, Terebus A, & Liang J (2016) Accurate Chemical Master Equation solution using multi-finite buffers. *Multiscale Model Simul* 14(2):923-963.
5. Josefsson L, *et al.* (2011) Majority of CD4+ T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. *Proc. Natl. Acad. Sci. U. S. A.* 108(27):11199-11204.
6. Josefsson L, *et al.* (2013) Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4+ T-cells contain one HIV-1 DNA molecule. *PLoS Pathog.* 9(6):e1003432.
7. Redner S (2001) *A guide to first-passage processes* (Cambridge University Press).
8. Nobile AG, Ricciardi LM, & Sacerdote L (1985) Exponential trends of Ornstein–Uhlenbeck first-passage-time densities. *J. Appl. Probab.* 22(2):360-369.
9. Karn J (2011) The molecular biology of HIV latency: breaking and restoring the Tat-dependent transcriptional circuit. *Curr. Opin. HIV AIDS* 6(1):4-11.
10. Weinberger LS, Burnett JC, Toettcher JE, Arkin AP, & Schaffer DV (2005) Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* 122(2):169-182.
11. Althaus CL & De Boer RJ (2010) Intracellular transactivation of HIV can account for the decelerating decay of virus load during drug therapy. *Mol. Syst. Biol.* 6:348.
12. Razooky BS, Pai A, Aull K, Rouzine IM, & Weinberger LS (2015) A hardwired HIV latency program. *Cell* 160(5):990-1001.
13. Razooky BS, *et al.* (2017) Nonlatching positive feedback enables robust bimodality by decoupling expression noise from the mean. *PLoS Biol.* 15(10):e2000841.
14. Reddy B & Yin J (1999) Quantitative intracellular kinetics of HIV type 1. *AIDS Res. Hum. Retroviruses* 15(3):273-283.
15. Kim H & Yin J (2005) Robust growth of human immunodeficiency virus type 1 (HIV-1). *Biophys. J.* 89(4):2210-2221.
16. Efthymiadis A, Briggs LJ, & Jans DA (1998) The HIV-1 Tat nuclear localization sequence confers novel nuclear import properties. *J. Biol. Chem.* 273(3):1623-1628.
17. Maiuri P, *et al.* (2011) Fast transcription rates of RNA polymerase II in human cells. *EMBO Rep* 12(12):1280-1285.
18. Marcello A (2012) RNA polymerase II transcription on the fast lane. *Transcription* 3(1):29-34.
19. Wang S, Huber PW, Cui M, Czarnik AW, & Mei HY (1998) Binding of neomycin to the TAR element of HIV-1 RNA induces dissociation of Tat protein by an allosteric mechanism. *Biochemistry* 37(16):5549-5557.
20. Schwartz S, Felber BK, & Pavlakis GN (1992) Distinct RNA sequences in the gag region of human immunodeficiency virus type 1 decrease RNA stability and inhibit expression in the absence of Rev protein. *J. Virol.* 66(1):150-159.