Pacific Symposium on Biocomputing 15:421-432(2010)

RETROVIRUS HTLV-1 GENE CIRCUIT: A POTENTIAL OSCILLATOR FOR EUKARYOTES

ALBERTO CORRADIN[1], BARBARA DI CAMILLO[1], FRANCESCA RENDE[2],
VINCENZO CIMINALE[2], GIANNA MARIA TOFFOLO[1], CLAUDIO COBELLI[1]
[1]*Department of Information Engineering, University of Padua*
via Gradenigo 6, Padua, 35131, Italy
[2]*Department of Oncology and Surgical Sciences, University of Padua*
Via Gattamelata 64, Padua, 35128, Italy

Retrovirus HTLV-1 gene circuit is characterized by positive and negative feedback phenomena, thus candidating it as a potential relaxation oscillator deliverable into eukaryotes. Here we describe a model of HTLV-1 which, by providing predictions of genes and proteins kinetics, can be helpful for designing gene circuits for eukaryotes, or for optimizing gene therapy approaches which are currently carried out by means of lentiviral vectors or re-engineered adenoviruses. Oscillatory patterns of HTLV-1 gene circuit are predicted when positive feedback is faster than negative feedback. Techniques to mutate the retroviral genome in order to implement practically the above conditions are discussed. Finally, the effect of stochasticity on the system behavior is tested by means of Gillespie algorithm. Simulations show the difficulties to preserve synchronization in viral expression for a multiplicity of cells, while the long tail of the density probability function of the master regulator gene *tax/rex*, due to its steady state fluctuations, suggests an activation mechanism of HTLV-1 similar to that recently proposed for HIV[1]: the virus tends to latency but under certain circumstances, the master regulator gene reaches high values of expression, whose persistence induces the viral replication.

1. Introduction

Synthetic gene circuits have already been delivered into bacterial cells, proving that it is possible to design and implement synthetic biological systems. A genetic toggle switch, designed by Gardner et al.[2] in 2000, was soon followed by the gene oscillator of Elowitz and Leibler[3]. An essential ingredient of this last work was the preliminary circuit characterization by modeling gene and protein expression; in particular, bifurcation analysis has allowed identifying the ranges of parameter values corresponding to periodic patterns. Nevertheless, the modeling strategy must be sound; otherwise misleading results can be obtained, which can heavily affect the subsequent circuit design. Recently, Kaern et al.[4] supported the use of systems of differential equations based on mass action, i.e. the approach previously adopted by Hasty et al.[5,6] for modeling the λ phage. In moving from bacteria to eukaryotes, the design of gene circuits becomes more difficult because of the more complex regulatory mechanisms. However, significant contributions have become available, e.g. Ramachandra et al.[7] re-engineered adenoviruses to hit tumor cells selectively, i.e. without impairing the healthy cells; Bainbridge et al.[8] addressed the Leber's Congenital Amaurosis by means of recombinant adeno-associated virus vectors. Also, of note is that mathematical models of viral kinetics are potentially valuable for optimizing gene therapy approaches, in particular by improving the design of retroviral vectors by predicting the gene expression following their delivery. In this paper we propose a novel model of HTLV-1 viral kinetics[9]. This model is characterized by positive and negative feedback phenomena, similarly to synthetic relaxation oscillators delivered into prokaryotes and able to exhibit limit cycles. In order to investigate the potential use of HTLV-1 circuit as a novel oscillator for eukaryotes, we analyze the periodic behavior of HTLV-1 model. This represents a preliminary step that can be instrumental for designing gene circuits to be delivered into eukaryotic cells, or for optimizing retroviral vector design in gene therapy.

Results previously obtained with bacteria suggest that a deterministic model may be not adequate to predict the true behavior of a biological system, e.g. a remarkable variability was observed in the repressilator period of oscillation by Elowitz and Leibler[3], Thattai et al.[10] described the noise in transcription and translation whereas Elowitz et al.[11] experimentally highlighted the effects induced by variable quantities of metabolites in the single cells of the same sample. On the other hand, Gillespie[12] pointed out that if a system is small enough that the molecular populations of some reactant species are not too many orders of magnitude larger than one, discreteness and stochasticity may play important roles, so that the predictions coming from deterministic differential equations do not accurately describe the system's true behavior. This is the case for HTLV-1 kinetics since one of the basic mechanisms is transactivation, i.e. the enhancement of transcription caused by the interaction between the viral promoter and a viral protein, and the number of promoters in a cell corresponds to the number of viral genomes integrated in the host cell, which is small. Thus, stochastic simulations are important for an adequate understanding of the system behavior.

In the following, after presenting an HTLV-1 model, which provides deterministic predictions about viral kinetics, we will discuss first the periodic patterns revealed by the bifurcation analysis, and then, the results of stochastic simulations performed by the Gillespie algorithm. Finally, the practical feasibility of an oscillator for eukaryotes will be considered, together with some recent biotechnological techniques potentially helpful to mutate the HTLV-1 genome so as to obtain periodic oscillations of gene and protein expression.

2. The model

The main mechanisms of the HTLV-1 gene circuit[13] are summarized in Figure 1:

1. The full-length genomic RNA of the single stranded retrovirus HTLV-1 encodes for the primary transcript *gag* (compartment 1) which undergoes either single or double splicing in the nucleus or, alternatively, remains unspliced.
2. The doubly spliced mRNA *tax/rex* (compartment 2) is considered the master regulator of viral gene expression since it encodes for two distinct regulatory proteins, p40*Tax* and p27*Rex* (compartments 3 and 4), from ORF III and IV, respectively; in the following we will refer to them simply as *Tax* and *Rex*.
3. *Tax* boosts the transcription of the primary transcript *gag*, generating a positive feedback phenomenon called transactivation.
4. *Rex* prevents the multiple splicing of *gag*, causing a decrease in the amount of *tax/rex* in favor of unspliced and singly spliced genes, and generating a negative feedback phenomenon with respect to *tax/rex*.
5. A variety of *Rex*-dependent viral genes deriving from the single splicing of *gag* were identified, e.g. *1-B*, *p13*, p21Rex (compartments 5.1, 5.2… 5.n), but the mechanisms of splitting up are still unclear.

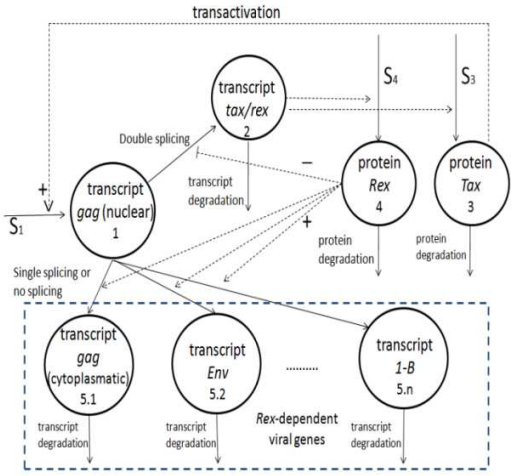

Figure 1. Background knowledge on the HTLV-1 gene circuit. Solid arrows represent fluxes, dashed arrows controls.

The system of differential equations which describe the above mechanisms, based on one-step reactions and mass action (detailed in the Appendix), is:

$$\begin{cases} \frac{dq_1}{dt} = \mathrm{m\,S} + \mathrm{m}\,S\,\beta'_1\,\frac{\mathrm{q_3(t)^2}}{\mathrm{h_1^2+q_3(t)^2}} - \mathrm{k_s}q_1(\mathrm{t}) & \mathrm{q_1(0)=0} \quad (1) \\ \frac{dq_2}{dt} = \{1 - \mathrm{g(q_1(t), q_4(t), z)}\}\,\mathrm{k_s}q_1(\mathrm{t}) - k_{02}\,q_2(\mathrm{t}) & \mathrm{q_2(0)=0} \quad (2) \\ \frac{dq_3}{dt} = \beta'_3 q_2(\mathrm{t}) - k_{03}\,q_3(\mathrm{t}) & \mathrm{q_3(0)=0} \quad (3) \\ \frac{dq_4}{dt} = \beta'_4 q_2(\mathrm{t}) - k_{04}\,q_4(\mathrm{t}) & \mathrm{q_4(0)=0} \quad (4) \end{cases}$$

where the state variables are the concentrations [molecules/l] of nuclear *gag*, *tax/rex*, *Tax*, *Re* corresponding to the compartments 1-4 of Figure 1. Initial conditions were set to zero, i.e. we supposed to deliver the HTLV-1 gene circuit into eukaryotic cells which were not infected previously. System parameters are: the transcription rate S [1/h], the concentration of viral genomes integrated in the host cell m [molecules/l], the transactivation constant β'_1 (adimensional), Michaelis constant h_1 (which is the product of many equilibrium constants, as described in the Appendix), the nuclear export rate k_s [1/h], the order of *Rex* multimerization z (adimensional), degradation rates of the transcript *tax/rex* k_{02} [1/h], of proteins *Tax* k_{03} [1/h] and *Rex* k_{04} [1/h], and parameters β'_3 and β'_4 [protein molecules/(transcript molecules*h)], which are the products of the *Tax* and *Rex* gains in protein translation multiplied for the rate constants of the translation processes (as detailed in the Appendix). The function g(·) is defined in Eq. 5-7:

$$\mathrm{g(q_1(t), q_4(t), z)} = \begin{cases} \frac{\mathrm{q_4^z}}{\mathrm{h_3^z}}\frac{1}{\mathrm{q_1}} & \text{If } \frac{\mathrm{q_4^z}}{\mathrm{h_3^z}} < \mathrm{q_1} \text{ and } \mathrm{q_1} > 0 \quad (5) \\ 1 & \text{If } \frac{\mathrm{q_4^z}}{\mathrm{h_3^z}} > \mathrm{q_1} \text{and } \mathrm{q_1} > 0 \quad (6) \\ 0 & \text{If } \mathrm{q_1} = 0 \quad (7) \end{cases}$$

where h_3 is the product of the some equilibrium constants, as described in the Appendix. Unfortunately, only k_{01} was measured by Grone et al.[14] equal to 0.069/h. As concerns other parameters, approximate values can be derived from the literature, based on measurements of similar biological processes, reported below with the corresponding model parameters between brackets:

1. Rosin-Arbesfeld et al.[15] and Lewis[16] reported the half-lives of about 10 and 4 minutes, respectively, for the nucleo-cytoplasmatic transport (k_s).
2. Weinberger and Shenk[17] estimated the transactivation constant (β'_1) in the HIV gene circuit by fitting an ODE model to normalized data of fluorescence intensities from single-cell measurements, obtaining a value of about 8.
3. As regards protein degradation rates, the *Tax* ubiquitination was confirmed by many laboratories[18] supporting the thesis of a ubiquitin-mediated degradation; moreover, Peloponese et al.[19] proved the *Tax* inactivation induced by ubiquitin. Jeong et al.[20] observed similar decays for *Tax* and β-Galactosidase, whose half-life was estimated in about 13 hours[21,22]. Therefore, we considered a half-life of about 10 hours for *Tax*, and also for *Rex* (k_{03}, k_{04}).
4. Kugel and Goodrich[23] measured the transcription rate induced by polymerase II in eukaryotes (S): 1.9e-3/s.

As regards Michaelis constant h_1, no information is available as well as for the parameter h_3, thus the amount of *Tax* and *Rex* were scaled by h_1 and h_3, respectively: $\widetilde{q_3} = \frac{q_3}{h_1}$ and $\widetilde{q_4} = \frac{q_4}{h_3}$; $\widetilde{q_3}$ and $\widetilde{q_4}$ can be viewed as the effective proteins, present in the nucleus and effectively acting for transactivation and RNA nuclear export, respectively.

Moreover, to preserve the validity of the study front of future HTLV-1-specific measurements resulting in different values of the transcription rate, the time t [h] was scaled by S as done in Ref. 4: $\tau = t*S$; τ is adimensional because the unit of measurements of S is 1/h. After scaling, the model of differential equations became:

$$\begin{cases} \frac{dq_1}{d\tau} = m + m\,\beta'_1 \frac{\widetilde{q_3}(\tau)^2}{1+\widetilde{q_3}(\tau)^2} - k'_s q_1(\tau) & q_1(0)=0 \quad (8) \\ \frac{dq_1}{d\tau} = \{1 - \tilde{g}(q_1(\tau), \widetilde{q_4}(\tau), z)\}\, k'_s q_1(\tau) - k'_{02}\, q_2(\tau) & q_2(0)=0 \quad (9) \\ \frac{d\widetilde{q_3}}{d\tau} = \widetilde{\beta'}_3 q_2(\tau) - k'_{03}\widetilde{q_3}(\tau) & \widetilde{q_3}(0)=0 \quad (10) \\ \frac{d\widetilde{q_4}}{d\tau} = \widetilde{\beta'}_4 q_2(\tau) - k'_{04}\widetilde{q_4}(\tau) & \widetilde{q_4}(0)=0 \quad (11) \end{cases}$$

where the function $\tilde{g}(\cdot)$ is defined in Eq. 12-14:

$$\tilde{g}(q_1(t), \widetilde{q_4}(t), z) = \begin{cases} \frac{\widetilde{q_4}^z}{q_1} & \text{If } \widetilde{q_4}^z < q_1 \text{ and } q_1 > 0 \quad (12) \\ 1 & \text{If } \widetilde{q_4}^z > q_1 \text{ and } q_1 > 0 \quad (13) \\ 0 & \text{If } q_1 = 0 \quad (14) \end{cases}$$

Model parameters were fixed to the following nominal values: $\beta'_1=10$, $k'_{01}=k_{01}/S=0.01$, $k'_s=k_s/S=1$, $k'_{02}=k_{02}/S=0.01$, $k'_{03}=k_{03}/S=0.01$, all of which are adimensional. Parameter m was initially set equal to 1 molecule/l, to reflect the hypothesis of low multiplicity of infection, which underlies the model development (see the Appendix). As regards the parameters $\widetilde{\beta'}_3$ and $\widetilde{\beta'}_4$, since no information was available in the literature, computational simulations of the deterministic system (Eq. 8-11) were performed, and the parameter values were fixed so as to obtain gene expression time course consistent with some experimental measurements[24,25]. This happened for $\widetilde{\beta'}_3$ and $\widetilde{\beta'}_4$ in the range [1e-3, 1e-1], thus we chose the median value $\widetilde{\beta'}_3 = \widetilde{\beta'}_4 = 0.01$ as the default value to be used for the subsequent analyses. Since it was not possible to establish if *Rex* forms dimers (z=2), pentamers (z=5), or something else, different values of z were considered.

3. Bifurcation analysis

We tested if the model of the HTLV-1 gene circuit (Eq.8-11), can exhibit periodic patterns by performing bifurcation analysis with the MatCont software package[26]. Among the system parameters, some were considered tunable on the basis of experimental observations, i.e. the protein degradation rates and the concentration m of viral genomes integrated in the host cell, whereas other parameters were set to their default values. The rational is that *Tax* undergoes ubiquitination[18] and experimental observations support the tunability of the proteins half-life when their degradation involves the ubiquitin pathway[27]. As regards m, it can be easily regulated at the time of virus delivery following an estimation of the titer of viral particles.

3.1 Periodic patterns

With the default parameter values the system falls into a stable steady state. By varying the ratio between the two protein degradation rates, RD=k'_{03}/k'_{04}, two Hopf bifurcations were detected confirming the possibility for the HTLV-1 gene circuit to oscillate[28]; moreover, the periodic patterns were stable because both bifurcations are supercritical. With z=2 the critical values of RD and the Lyapunov coefficients were: RD_{H1}=2.3, RD_{H2}=26.2 with L_{H1}=-4.6e-3 and L_{H2}=-7.6e-4, whereas with z=5 they were: RD_{H1}=7.0, RD_{H2}=13.7 with L_{H1}=-1.5e-3 and L_{H2}=-7.6e-4. In Figure 2a, the trajectories of the state variables *gag* and *tax/rex* for z=2 and RD=3 are shown.

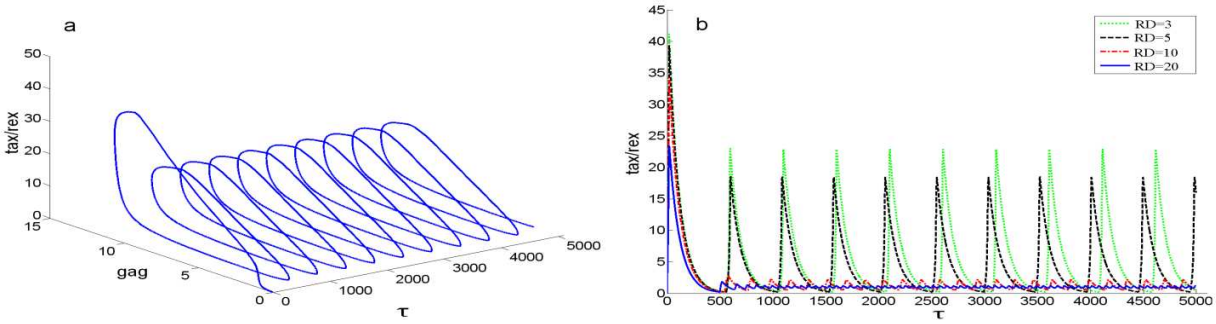

Figure 2. a) The trajectory of the system for z=2 and RD=3. b) Periodic patterns of *tax/rex* for z=2 and the following values of RD: 3,5,10,20.

Higher values of *Tax* degradation rate (with respect to *Rex*) result in smaller amplitudes and periods of oscillation, as shown in Figure 2b, where the periodic patterns of *tax/rex* corresponding to z=2 and RD=3,5,10,20, respectively, are plotted. If multiplicity of infection m is increased, periodic patterns arise if RD is within specific limits which depends on m, as shown in Figure 3a, for z=2 and z=5. This figure suggests the relevant role of z on system behavior. To have a better insight on the role of z, Hopf continuation was performed by varying RD and z, for specific m values. Results (Figure 3b) indicate that periodic oscillations of the model state variables are prevented for z>5.

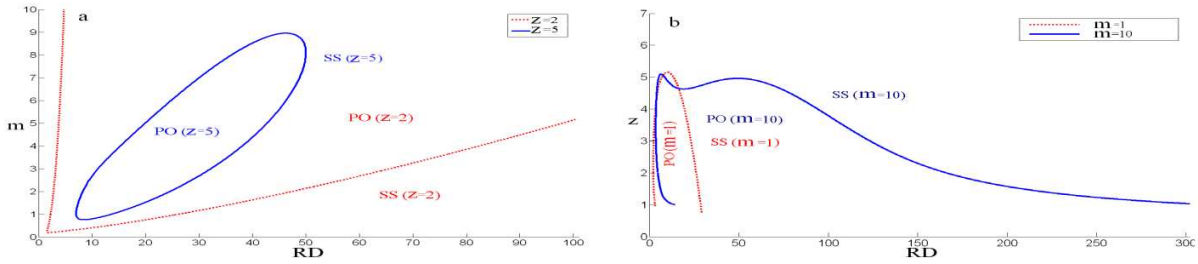

Figure 3. a) Hopf curve continuation with free parameters RD and m, for z=2 and 5. The areas corresponding to parameter settings allowing periodic oscillations are signaled with the symbol PO whereas the areas corresponding to parameter settings for which the system falls into steady states are signaled with the symbol SS. b) Hopf curve continuation with free parameters RD and z, for m=1 and 10.

4. Stochastic fluctuations

Since stochasticity and discreteness can cause deviations of the true system behavior from the predictions of deterministic differential equations when the molecular populations are small, as is the case of the viral promoter sites in our model, stochastic simulations were performed by Gillespie algorithm (direct method[29,30]), with the parameter settings corresponding to the periodic patterns and to the steady state solution of the system (Eq. 8-11). Gillespie algorithm describes the number X(t) of molecules of chemical species involved in the reactions R_j characterizing the system. The key to simulate trajectories of X(t) is the probability function $p(\tau, j \mid x,t)$[12], which is the probability, given X(t) = x, that the next reaction in the system will occur in the infinitesimal time interval $[t+\tau, t+\tau+d\tau)$, and will be an R_j reaction. This probability is related to the number of molecules composing the chemical species involved in the reaction as reactant. Thus, the variability concerns which reaction takes place and when this happens. Two software packages, providing distinct implementation of the algorithm, were considered: Dizzy[31] (ver 1.11.4) and Cain[32] (ver 0.12).

4.1 Periodic patterns

The effects of stochasticity on the periodic patterns of Figure 2b were verified by simulating 1000 trajectories of the state variables. Results evidenced different time points at which peaks of *tax/rex* rise in distinct realizations (see Figure 4a) causing a lack of synchronicity. Consequently, the *tax/rex* mean time course resulted to be leveled, as shown in panel b; moreover, a high variability in gene expression was observed (see panel c, where the standard deviations corresponding to each time point are reported). The leveling of the *tax/rex* mean time course appeared also for other values of z, as shown in panel d, where it is plotted the mean of 1000 realizations with parameter setting z=5 and RD=8, for which periodic oscillations arise in the deterministic system (Eq. 8-11). To verify if the leveling depended on m, the same simulations were repeated with m=10 and 100, instead of 1, but no better result was observed.

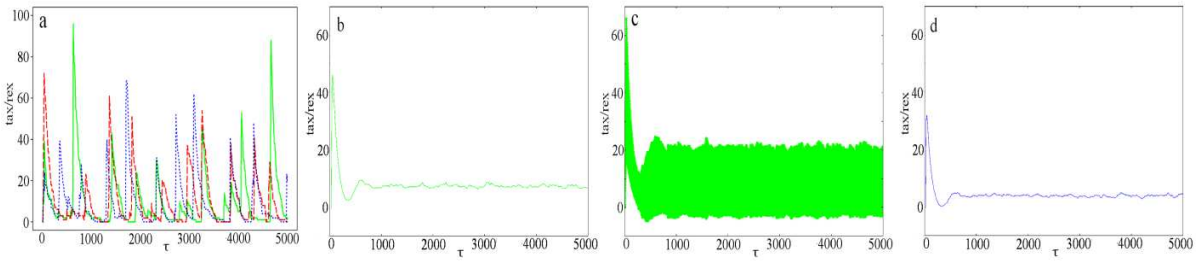

Figure 4. Stochasticity in chemical reactions causes the lack of synchronicity and the leveling of the *tax/rex* mean time course. a) 3 trajectories of the state variable *tax/rex*, b) the mean time course of 1000 realizations, and c) the standard deviations corresponding to each time point for z=2 and RD=3. d) The *tax/rex* mean time course of 1000 realizations for z=5 and RD=8.

4.2 Steady state fluctuations

Since the addition of stochastic phenomena on the periodic patterns provided surprising results, also stochastic fluctuations of the steady state solution were examined by means of 10000 stochastic simulations. All state variables resulted to be affected by remarkable variability. In particular, *tax/rex* values presented a coefficient of variation (CV) higher than 100%, as it is shown in Figure 4, panel a and b, where the results obtained with Dizzy and Cain are reported. To have a better insight, we investigated the density probability function of *tax/rex* that resulted to be characterized by a long tail (see panel c). With different values of z similar distributions appeared, as shown in panel d, where it is plotted the density probability function of *tax/rex* for z=5.

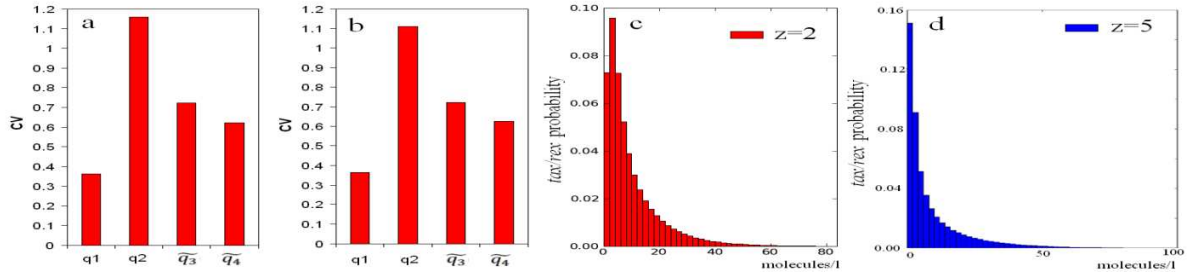

Figure 5. Steady state fluctuations of the steady state solution showed a high variability. a) The CVs obtained from stochastic simulations performed by Dizzy; b) The CVs obtained from stochastic simulations performed by Cain. For these simulations z was set to 2. Density probability functions of the values of *tax/rex* for : c) z=2, and d) z=5.

5. Discussion and conclusions

The gene circuit of the retrovirus HTLV-1 is characterized by positive and negative feedback phenomena, due to the regulatory proteins *Tax* and *Rex*, thus candidating it as a potential relaxation oscillator. To test this hypothesis, a model of the retroviral gene and protein kinetics was developed on the basis of well-established knowledge. The model incorporates the cascade of interactions involving the viral promoter and the biological processes of transcription, translation and degradation, assuming the former to be faster and in equilibrium with respect to the latter. Reasonable simplifications were introduced to limit the number of parameters, and approximate but reasonable numerical values were assigned to them based on information derived in the literature, with the only

exception of order of *Rex* multimerization z, which is unknown. Thus the analysis of the system behavior was for different values of z. To test *in silico* the possibility of observing periodic patterns, bifurcation analysis was performed on the deterministic system of differential equations, considering two model parameters as tunable on the basis of experimental observations: the degradation rate of regulatory proteins and the parameter m, which is related to the multiplicity of infection. Our results show that oscillatory behaviors take place if the kinetics of the positive feedback are faster than those of the negative feedback, as expected from a relaxation oscillator; moreover, higher values of *Tax* degradation rate result in smaller amplitudes and oscillation periods. Obviously, other parameters influence the system behavior: the values of RD compatible with oscillatory behavior depend on m as well as on z. In particular, periodic oscillations are prevented if z is greater than 5.

To reach periodic oscillations, the *Tax* degradation rate should be increased with respect to that of *Rex*. Three recent experimental techniques support the possibility of altering protein degradation rate in practice, two based on enhancing the ubiquitin degradation pathway and the latter on protein tagging. Bachmair et al.[33] pointed out the important role of the amino-terminus in stabilizing/destabilizing the proteins which undergo ubiquitination: in their experiments the β-Galactosidase half-life lowered from more than 20 hours to less than 3 minutes depending on the amino-terminus. Rogers and Rechsteiner[34,35] observed the correlation between the high presence of PEST sequences – where PEST is the nice abbreviation of proline (P), glutamic acid (E), serine (S), and threonine (T) - and the short half-lives of proteins degraded by the ubiquitin pathway. Consequently, the substitution of PEST amino acids with more stable ones should increase the protein half-life. McGinness et al.[36] suggested the import of the E.coli ClpXP protease into eukaryotic cells and the addition of an appropriate ssrA tag to the protein under exam to modulate its degradation rate. Briefly, the tag should have weak affinity for the protease so the introduction of the SspB adaptor protein can be used as a control lever to increase this affinity and, consequently, the proteolysis of the tagged proteins, induced by the ClpXP protease. However, application to the HTLV-1 gene circuit of these three methods is not immediate because *Tax* and *Rex* are translated from the same transcript *tax/rex* and, consequently, the mutation of one protein implies the mutation of the other. To address this problem and make *Tax* kinetics faster than *Rex* kinetics, a possible solution is supplied by the PEST hypothesis. Since the coding sequences of *Tax* and *Rex* are (4829..4832, 6951..8008) and (4773..4832, 6950..8008), respectively (data from the NCBI Reference Sequence NC_001436.1), the sequence (4773..4832) is present in *Rex* but absent in *Tax*. Moreover, it configures as a PEST region since it includes one glutamine, one serine, two threonines and five prolines out of 19 aa, which are all destabilizing amino acids. Therefore, their substitution with more stable amino acids, by site-directed mutagenesis[37,38], should decrease the *Rex* degradation rate, allowing to obtain periodic oscillations. Conversely, practical applicability of the methods proposed by Bachmair et al.[33] and McGinness et al.[36] to our system is still an open issue, due to the overlapping of *Rex* and *Tax* protein sequences.

Since stochasticity and discreteness can cause deviations of the true system behavior from the predictions of deterministic differential equations when the molecular populations are small, as is the case of the viral promoter sites in our model, the Gillespie algorithm was used to perform stochastic simulations. Simulations revealed the leveling of *tax/rex* time course essentially due to the lack of synchronization among the oscillators delivered in distinct cells. In particular, peaks occur at different times in the distinct realizations, as experimentally observed by Stricker et al.[39] by measuring single-cell fluorescence trajectories. The problem of cell synchronization can be addressed by electroporation, but only partially, since the recently developed methodology of transfection[40], which allows the delivery of genes of interest directly into the nuclear compartment in a time period of microseconds, does not guarantee the persistence of synchronization. A continuous synchronizing signal[41,42] may be needed to preserve the forced initial synchronization over time, but currently this is not available. As a consequence, experimental validation is not straightforward and will require single-cell measurements of out of phase oscillators, by time-lapse microscopy and using GFP reporters. Recent findings support the applicability of this technique to viral genes, since some lentiviral vectors with GFP as reporter of the transactivator gene *Tat* were described for the HIV gene circuit[1]. Particularly interesting is the wild type HIV-1 with the gene *Nef* substituted by the GFP; a fascinating testable hypothesis is the realization of a reporter version of the HTLV-1 genome with *tax/rex* substituted by the GFP, to be delivered in addition to the appropriately mutated virus. We expect the RNA *gag* transcribed from this reporter

construct, be either doubly spliced into the GFP, supplying fluorescence intensities proportional to the presence of *tax/rex*, or transferred to the cytoplasm and degraded.

Stochastic fluctuations of the steady state solution were also examined. All state variables are affected by remarkable variability. In particular, *tax/rex* values have a CV higher than 100%, with a long tail of their density probability function, indicating that *tax/rex* gene expression is likely to sometimes assume very high values because of stochastic fluctuations, suggesting mechanisms of retroviral activation similar to those recently proposed by Weinberger for HIV[1].

In conclusion, the bifurcation analysis of the proposed model of the HTLV-1 gene circuit revealed that periodic patterns are possible, provided that *Tax* kinetics is faster than *Rex* kinetics. The next step is the experimental validation of these predictions. However, the stochastic simulations pointed out the problem of cell synchronicity; consequently, single-cell measurements are necessary to observe oscillatory patterns of genes or proteins. Moreover, the high variability at steady state suggests mechanisms of retroviral activation similar to those proposed for HIV.

Appendix

Following Ref. 5,6, the HTLV-1 chemical reactions[13] were divided into two categories: fast and slow; in particular, protein multimerization and complex formation were assumed to be faster than the processes of transcription, translation and degradation of proteins and transcripts. It is reasonable to assume that fast reactions are of the order of seconds, similarly to λ phage's[6] and thus, although not exactly known, much faster than protein degradation in eukaryotes, that is of the order of hours or days[27]. Therefore, faster reactions can be safely assumed to be in equilibrium with respect to the slower ones. In the following paragraphs, we will introduce the fast reactions and then the slower ones.

A.1 Fast reactions

In this paragraph, the following HTLV-1 biological processes[13] will be described: (1) dimerization and complex formation, which lead to transactivation; (2) cooperative interactions, which are necessary for transcription; and (3) *Rex* multimerization.

Dimerization and complex formation: HTLV-1 transactivation is due to the binding of a complex, composed of dimers of *Tax* and dimers of the cellular transcription factor CREB, to the Tax Responsive Element[43] (TRE) in the viral Long Terminal Repeat, where the viral promoter is located. A set of chemical reactions are used to describe:

1. *Tax* dimers formation and their transfer to the nucleus (*Tax* is a shuttling protein[44], i.e. it transfers from the cytoplasm to the nucleus and viceversa, so dimers are present in the whole cell but only the nuclear fraction is involved in transactivation).
2. CREB dimers formation and their transfer to the nucleus.
3. The formation of a $CREB_2$-TAX_2 complex, which subsequently binds to the TRE inducing transactivation, and the alternative interaction $CREB_2$-TRE from which the basal transcription follows. In the following we will indicate the transactivated promoter sites and the nontransactivated ones with the abbreviations $TPrS_0$, and $NTPrS_0$, respectively, and with TRE the inactivated viral promoter sites. Table A1 summarizes the chemical reactions and the corresponding equilibria.

Table A.1. Chemical reactions and the corresponding equilibria concerning molecular dimerization and complex formation.

Tax dimerization	$2\,TAX \underset{}{\overset{K_1}{\rightleftharpoons}} TAX_2$	$[TAX_2] = K_1[TAX]^2$
Tax transfer to the nucleus	$TAX_2 \overset{K_2}{\rightleftharpoons} TAX_2^n$	$[TAX_2^n] = K_2[TAX_2]$
CREB dimerization	$2\,CREB \overset{K_3}{\rightleftharpoons} CREB_2$	$[CREB_2] = K_3[CREB]^2$
CREB transfer to the nucleus	$CREB_2 \overset{K_4}{\rightleftharpoons} CREB_2^n$	$[CREB_2^n] = K_4[CREB_2]$
$CREB_2$-TAX_2 complex formation in the nucleus	$CREB_2^n + TAX_2^n \overset{K_5}{\rightleftharpoons} Complex$	$[Complex] = K_5K_4K_3[CREB]^2\,K_2K_1[TAX]^2$

Binding of the complex to TRE (transactivated promoter)	$\underset{\text{Complex} + \text{TRE} \rightleftharpoons \text{TPrS}_0}{K_{6T}}$	$[\text{TPrS}_0] = K_{6T}K_5K_4K_3[\text{CREB}]^2\ K_2K_1[\text{TAX}]^2\ [\text{TRE}]$
Binding of $CREB_2$ to TRE (non-transactivated promoter)	$\underset{\text{CREB}_2^n + \text{TRE} \rightleftharpoons \text{NTPrS}_0}{K_{6NT}}$	$[\text{NTPrS}_0] = K_{6NT}K_4K_3[\text{CREB}]^2\ [\text{TRE}]$

Cooperative interactions: viral transcription involves a cascade of co-activators and general transcription factors, CBP/p300, PCAF, TFIIA, TFIIB,TFIID[13], whose binding reactions with the promoter regions are described in Table A.2. The transactivated and the non-transactivated promoter sites will be indicated with $TPrS_i$, and $NTPrS_i$, where the suffix i denotes the step in the cascade. Following the same line of reasoning which underlies the formulation of the well-known pseudo-first order rate equations[45], the concentrations of all co-activators and general transcription factors are assumed to be constant or in great excess with respect to the viral promoters they interact with, so that the effects of the variations of their concentrations on the viral kinetics are negligible. Consistently, a low multiplicity of infection is assumed, i.e. we suppose that few viral genomes are integrated in the host cells. Table A.2 summarizes the chemical reactions and the corresponding equilibria.

Table A.2. Chemical reactions and the corresponding equilibria concerning cooperative interactions.

Binding of CBP/p300 to the non-transactivated promoter site	$\underset{\text{NTPrS}_0 + \text{CBP/p300} \rightleftharpoons \text{NTPrS}_1}{K_{7A}}$	$[\text{NTPrS}_1] = K_{7A}[\text{NTPrS}_0]\ [\text{CBP}] = K'_{7A}[\text{NTPrS}_0]$
Binding of PCAF to the non-transactivated promoter site	$\underset{\text{NTPrS}_1 + \text{PCAF} \rightleftharpoons \text{NTPrS}_2}{K_{8A}}$	$[\text{NTPrS}_2] = K_{8A}[\text{NTPrS}_1]\ [\text{PCAF}] = K'_{8A}[\text{NTPrS}_1]$
Binding of TFIIA to the non-transactivated promoter site	$\underset{\text{NTPrS}_2 + \text{TFIIA} \rightleftharpoons \text{NTPrS}_3}{K_{9A}}$	$[\text{NTPrS}_3] = K_{9A}[\text{NTPrS}_2]\ [\text{TFIIA}] = K'_{9A}[\text{NTPrS}_2]$
Binding of TFIIB to the non-transactivated promoter site	$\underset{\text{NTPrS}_3 + \text{TFIIB} \rightleftharpoons \text{NTPrS}_4}{K_{10A}}$	$[\text{NTPrS}_4] = K_{10A}[\text{NTPrS}_3]\ [\text{TFIIB}] = K'_{10A}\ [\text{NTPrS}_3]$
Binding of TFIID to the non-transactivated promoter site	$\underset{\text{NTPrS}_4 + \text{TFIID} \rightleftharpoons \text{NTPrS}_5}{K_{11A}}$	$[\text{NTPrS}_5] = K_{11A}[\text{NTPrS}_4]\ [\text{TFIID}] = K'_{11A}\ [\text{NTPrS}_4]$
Binding of CBP/p300 to the transactivated promoter site	$\underset{\text{TPrS}_0 + \text{CBP/p300} \rightleftharpoons \text{TPrS}_1}{K_{7B}}$	$[\text{TPrS}_1] = K_{7B}[\text{TPrS}_0]\ [\text{CBP}] = K'_{7B}[\text{TPrS}_0]$
Binding of PCAF to the transactivated promoter site	$\underset{\text{TPrS}_1 + \text{PCAF} \rightleftharpoons \text{TPrS}_2}{K_{8B}}$	$[\text{TPrS}_2] = K_{8B}[\text{TPrS}_1]\ [\text{PCAF}] = K'_{8B}[\text{TPrS}_1]$
Binding of TFIIA to the transactivated promoter site	$\underset{\text{TPrS}_2 + \text{TFIIA} \rightleftharpoons \text{TPrS}_3}{K_{9B}}$	$[\text{TPrS}_3] = K_{9B}[\text{TPrS}_2]\ [\text{TFIIA}] = K'_{9B}[\text{TPrS}_2]$
Binding of TFIIB to the transactivated promoter site	$\underset{\text{TPrS}_3 + \text{TFIIB} \rightleftharpoons \text{TPrS}_4}{K_{10B}}$	$[\text{TPrS}_4] = K_{10B}[\text{TPrS}_3]\ [\text{TFIIB}] = K'_{10B}[\text{TPrS}_3]$
Binding of TFIID to the transactivated promoter site	$\underset{\text{TPrS}_4 + \text{TFIID} \rightleftharpoons \text{TPrS}_5}{K_{11B}}$	$[\text{TPrS}_5] = K_{11B}[\text{TPrS}_4]\ [\text{TFIID}] = K'_{11B}[\text{TPrS}_4]$

Rex multimerization: protein *Rex* multimerizes[46], but the exact kind of multimer it forms is not known; in particular, there is evidence that Rex at least dimerizes[47], but the formation of complexes of higher orders like pentamers or hexamers is likely as well. To describe *Rex* multimerization, the Helfferich procedure for multistep reactions[48] is applied, summarizing with K_{12} the ratio of the overall forward and backward kinetics constants. Like *Tax*, also *Rex* is a shuttling protein[49] and only its nuclear fraction is involved in the nuclear export of incompletely spliced transcripts. Table A.3 summarizes the chemical reactions and the corresponding equilibria; with the symbol z we indicate the number of *Rex* molecules involved in the multimer formation, e.g. z=2 for dimers and z=5 for pentamers.

Table A.3. Chemical reactions and the corresponding equilibria concerning *Rex* multimerization and its transfer to the nucleus.

Rex multimerization	$\underset{z\ \text{REX} \rightleftharpoons \ldots \rightleftharpoons \text{REX}_z}{K_{12}}$	$[\text{REX}_z] = K_{12}[\text{REX}]^z$
Rex transfer to the nucleus	$\underset{\text{REX}_z \rightleftharpoons \text{REX}_z^n}{K_{13}}$	$[\text{REX}_z^n] = K_{13}[\text{REX}_z] = K_{13}K_{12}[\text{REX}]^z$

A.2 Slow reactions

Transcription, translation and the degradation of transcripts and proteins are irreversible and slow reactions; in the following we will describe: (1) the transcription of the primary transcript *gag*; (2) the alternative splicing of *gag* and the nuclear export of mRNAs; (3) the kinetics of the transcript *tax/rex*; (4) the kinetics of the proteins *Tax* and *Rex*; and (5) the kinetics of the incompletely spliced transcripts.

The transcription of the primary transcript *gag*: *gag* synthesis $S_1(t)$ is due to the basal transcription $S_{11}(t)$ and to the transcription induced by transactivation $S_{12}(t)$. Chemical reactions of the transcription processes $S_{11}(t)$ and $S_{12}(t)$, with their mathematical formulations, are shown in Table A.4, where the concentration of RNA polymerase (RNAp), is assumed to be constant or in great excess with respect to the viral promoter sites it interacts with. In Table A.4, β_1 indicates the gain in transcription, i.e. the number of transcripts generated by the binding of a molecule of RNAp to DNA and the subsequent process of *gag* transcription, k_t is a reaction rate constant [1/h] and c_0 a multiplicative constant.

Table A.4. Chemical reactions of transcriptions and the corresponding syntheses of nuclear *gag*.

Basal transcription	$\mathrm{NTPrS_5 + RNAp \xrightarrow{k_t} NTPrS_5 + RNAp} + \beta_1$ nuclear *Gag*	$S_{11}(t) = \beta_1 k_t p_0 [\mathrm{NTPrS_5}](t) = \\ = \beta_1 k_t p_0 \mathrm{K'_{11A} K'_{10A} K'_{9A} K'_{8A} K'_{7A}} [\mathrm{NTPrS_0}](t)$
With transactivation	$\mathrm{TPrS_5 + RNAp \xrightarrow{c_0 k_t} TPrS_5 + RNAp} + \beta_1$ nuclear *Gag*	$S_{12}(t) = \beta_1 c_0 k_t p_0 [\mathrm{TPrS_5}](t) = \\ = \beta_1 c_0 k_t p_0 \mathrm{K'_{11B} K'_{10B} K'_{9B} K'_{8B} K'_{7B}} [\mathrm{TPrS_0}](t)$

From the transcription processes $S_{11}(t)$ and $S_{12}(t)$, in molecules/h, of Table A.4, we derive the transcription rates $S=\beta_1 k_t p_0 \mathrm{K'_{11A} K'_{10A} K'_{9A} K'_{8A} K'_{7A}}$ and $S'=\beta_1 c_0 k_t p_0 \mathrm{K'_{11B} K'_{10B} K'_{9B} K'_{8B} K'_{7B}}$, in 1/h, for the basal transcription and transcription following transactivation, and we call c_1 their ratio, i.e. $c_1 = S'/S$. To have a better insight on the effects of the *Tax*-induced transactivation on the total *gag* synthesis $S_1(t)$, we introduce the multiplicity of infection (MOI), which is the mean number of viral genomes integrated in the host genome per cell. Then, we indicate with [cells] the cell concentration in a sample, in number of cells/l. Consequently, the concentration of retroviral genomes integrated in the host cells, m, equals MOI*[cells], in number of viral molecules/l. Now, making the working hypothesis that all the promoters are demethilated, i.e. none of them is *a priori* prevented from being involved in transcription, we derive that:

$$\mathrm{m} = [\mathrm{TRE}](t) + [\mathrm{NTPrS_0}](t) + [\mathrm{TPrS_0}](t) \tag{A.1}$$

The total transcription $S_1(t)$ is the sum of $S_{11}(t)$ and $S_{12}(t)$. Therefore, by summing up we have:

$$S_1(t) = \mathrm{S}\,[\mathrm{NTPrS_0}](t) + S'\,[\mathrm{TPrS_0}](t) = S\{[\mathrm{NTPrS_0}](t) + c_1[\mathrm{TPrS_0}](t)\} \tag{A.2}$$

where the constant c_1 was introduced to obtain the right-hand side of the equation. Then, from Eq. A.1 and A.2, and by some algebraic passages we obtain:

$$S_1(t) = S\{\mathrm{m} - [\mathrm{TRE}](t) + (c_1 - 1)[\mathrm{TPrS_0}](t)\} = S\{\mathrm{m} - [\mathrm{TRE}](t)\}\left\{1 + \frac{[\mathrm{TPrS_0}](t)}{\mathrm{m} - [\mathrm{TRE}](t)}(c_1 - 1)\right\} \tag{A.3}$$

Now, we focus on the term $\frac{[\mathrm{TPrS_0}](t)}{\mathrm{m}-[\mathrm{TRE}](t)}$ of Eq. A.3 and call $q_3(t)$ the *Tax* concentration. From Eq A.1 and equilibria of Table A.1 we obtain:

$$\frac{[\mathrm{TPrS_0}](t)}{\mathrm{m} - [\mathrm{TRE}](t)} = \frac{[\mathrm{TPrS_0}](t)}{[\mathrm{NTPrS_0}](t) + [\mathrm{TPrS_0}](t)} = \frac{\mathrm{q_3}(t)^2}{\mathrm{h_1^2} + \mathrm{q_3}(t)^2} \qquad \text{with: } \mathrm{h_1^2} = \frac{K_{6NT}}{K_{6T}} \frac{1}{K_5 \mathrm{K_2 K_1}} \tag{A.4}$$

Eq. A.4 shows that an elevated *Tax* concentration increases the number of transactivated promoter sites among the promoters involved in transcription, which are $[\mathrm{TPrS_0}](t)+[\mathrm{NTPrS_0}](t)$. Then, by inserting the right term of Eq. A.4 in Eq. A.3 the synthesis of nuclear *gag* becomes:

$$S_1(t) = S\{\mathrm{m} - [\mathrm{TRE}](t)\} + \{\mathrm{m} - [\mathrm{TRE}](t)\}\, S\, \beta'_1 \frac{\mathrm{q_3}(t)^2}{\mathrm{h_1^2} + \mathrm{q_3}(t)^2} \tag{A.5}$$

where $\beta'_1 = c_1 - 1$ is the transactivation constant (adimensional). Eq. A.5 shows a saturative effect of *Tax* concentration on *gag* synthesis given by the term $\frac{\mathrm{q_3}(t)^2}{\mathrm{h_1^2} + \mathrm{q_3}(t)^2}$, which can be due to the limited number of integrated viral promoter sites; in other words once they have been all transactivated, the *gag* transcription $S_1(t)$ saturates and no further increase is possible. Now, let $p_1 = [\mathrm{CREB}]$, which is assumed to be constant or in great excess with respect to the viral promoter sites it interacts with, and focus on the term m-[TRE](t). From equilibria of Table A.1 and Eq. A.1, and by several algebraic passages we obtain:

$$\mathrm{m} - [\mathrm{TRE}](t) = \mathrm{m} - \frac{m}{\mathrm{f}(\mathrm{p_1}, \mathrm{q_3}(t))} \tag{A.6}$$

where:

$$f(p_1, q_3(t)) = 1 + \frac{{p_1}^2}{h_2^2} + \frac{{p_1}^2}{h_2^2}\frac{1}{h_1^2} q_3(t)^2 \qquad \text{with } h_2^2 = \frac{1}{K_{6NT} K_4 K_3} \tag{A.7}$$

Eq. A.6 implies that if there is a lot of *Tax* or CREB then $f(p_1, q_3(t)) \gg m$ and, consequently, $m - [TRE](t) \approx m$, i.e. the number of promoter sites which are not involved in transcription becomes negligible. Therefore, by inserting the right term of Eq. A.6 in Eq. A.5, the synthesis of *gag* becomes:

$$S_1(t) = S\left\{m - \frac{m}{f(p_1, q_3(t))}\right\} + \left\{m - \frac{m}{f(p_1, q_3(t))}\right\} S\,\beta_1' \frac{q_3(t)^2}{h_1^2 + q_3(t)^2} \tag{A.8}$$

The alternative splicing of *gag* and the transfer of mRNAs to the cytoplasm: in presence of nuclear *Rex*, the transcripts are transferred incompletely spliced to the cytoplasm whereas, in absence of *Rex*, *gag* is doubly spliced into *tax/rex*[13]. Quantitatively, the amount of incompletely spliced RNAs transferred to the cytoplasm depends on the fraction of *gag* molecules which interact with multimers of *Rex* in the nucleus. This fraction is represented by the function $g(q_1(t), q_4(t), z)$, see Eq. A.9-11.

$$g(q_1(t), q_4(t), z) = \begin{cases} \frac{K_{13} K_{12} q_4^z}{q_1} = \frac{q_4^z}{h_3^z}\frac{1}{q_1} & \text{with } h_3^z = \frac{1}{K_{13} K_{12}} \quad \text{If } \frac{q_4^z}{h_3^z} < q_1 \text{ and } q_1 > 0 \quad \text{(A.9)} \\ 1 \quad \text{If } \frac{q_4^z}{h_3^z} > q_1 \text{ and } q_1 > 0 \quad \text{(A.10)} \\ 0 \quad \text{If } q_1 = 0 \quad \text{(A.11)} \end{cases}$$

where the variable $q_4(t)$ represents the *Rex* concentration in the sample, and the algebraic passages derive from the equilibria of Table A.3. The transcripts that are not transferred to the cytoplasm are degraded by the nuclear enzymes. Therefore, the total decay of g*ag* in the nucleus is due to the nuclear export plus the nuclear degradation (see reactions in Table A.5, where k_s and k_{01} are reaction rate constants, and c_2 is a multiplicative constant; for simplicity, the splicing and the nuclear export are condensed into a unique reaction).

Table A.5. Nuclear *gag* decay is due to the nuclear export of incompletely spliced RNAs (with Rex case), the splicing and nuclear export of doubly spliced RNAs (without *Rex* case) and to the nuclear degradation of the transcripts.

with *Rex*	$\xrightarrow{k_s}$ Nuclear *gag* + Rex_z → Cytoplasmatic incompletely spliced RNA + Rex_z	$L_{11}(\tau) = g(q_1(t), q_4(t), z)\, k_s q_1(t)$
without *Rex*	$\xrightarrow{c_2 k_s}$ Nuclear *gag*+ $\sum_i splicing\ factor_i$ → Cytoplasmatic *tax/rex* +$\sum_i splicing\ factor_i$	$L_{12}(\tau) = \{1 - g(q_1(t), q_4(t), z)\}\, c_2\, k_s q_1(t)$
Nuclear degradation	$\xrightarrow{k_{01}}$ Nuclear *gag*→	$D_1(\tau) = \{1 - g(q_1(t), q_4(t), z)\}\, k_{01} q_1(t)$

The sum of the three addends of Table A.5 is:

$$\text{Nuclear } gag \text{ decay}(t) = L_{11}(t) + L_{12}(t) + D_1(t) = \tag{A.12}$$
$$= g(q_1(t), q_4(t), z)\, k_s q_1(t) + \{1 - g(q_1(t), q_4(t), z)\}\, c_2\, k_s\, q_1(t) + \{1 - g(q_1(t), q_4(t), z)\}\, k_{01} q_1(t)$$

Therefore, by considering *gag* synthesis (Eq. A.8) and nuclear decay (Eq. A.12), the rate equation for nuclear g*ag* is:

$$\frac{dq_1}{dt} = S_1(t) - Nuclear\ gag\ decay(t) = \tag{A.13}$$
$$= S\left\{m - \frac{m}{f(p_1, q_3(t))}\right\} + \left\{m - \frac{m}{f(p_1, q_3(t))}\right\} S\,\beta_1' \frac{q_3(t)^2}{h_1^2 + q_3(t)^2} - g(q_1(t), q_4(t), z) k_s q_1(t) - \{1 - g(q_1(t), q_4(t))\}\{c_2\, k_s + k_{01}\} q_1(t)$$

where $q_1(t)$ represents the concentration of nuclear *gag*, in molecules/l.

The kinetics of the transcript *tax/rex*: *tax/rex* kinetics is given in Eq. A.14, where $S_2(t)$ is the synthesis following the double splicing of nuclear *gag*, i.e. $S_2(t) = L_{12}(t)$ (Table A.5), and $D_2(t)$ represents the degradation; the reactions are reported in Table A.6, where k_{02} is the reaction rate constant of the *tax/rex* degradation process. Therefore, the rate equation for *tax/rex* is:

$$\frac{dq_2}{dt} = S_2(t) - D_2(t) = \{1 - g(q_1(t), q_4(t), z)\} c_2\, k_s\, q_1(t) - k_{02}\, q_2(t) \tag{A.14}$$

where $q_2(t)$ represents the *tax/rex* concentration, in molecules/l.

Table A.6. Transcript *tax/rex* kinetics is due to two terms: synthesis, by double splicing of nuclear *gag*, and degradation.

tax/rex synthesis	$c_2 k_s$ Nuclear *Gag*+ $\sum_i splicing\ factor_i \rightarrow$ $\rightarrow$ Cytoplasmatic *tax/rex* +$\sum_i splicing\ factor_i$	$S_2(t) = L_{12}(\tau) = \{1 - g(q_1(t), q_4(t), z)\}\ c_2\ k_s\ q_1(t)$
tax/rex degradation	k_{02} *tax/rex* $\rightarrow$	$D_2(t) = k_{02}\ q_2(t)$

The kinetics of the proteins *Tax* and *Rex*: *Tax* and *Rex* are translated from the same transcript *tax/rex*; their synthesis reactions are shown in Table A.7, where k_{T3} and k_{T4} are reaction rate constants [1/h], β_3 and β_4 the gains in protein translations [protein molecules/transcript molecules], and β'_3, β'_4 their products [molecules/(molecules*h)].

Table A.7. Reactions of translation of the regulatory proteins *Tax* and *Rex*.

Tax synthesis	k_{T3} *tax/rex* $\rightarrow$ *tax/rex* + β_3 *Tax*	$S_3(t) = k_{T3}\ \beta_3\ q_2(t) = \beta'_3 q_1(t)$
Rex synthesis	k_{T4} *tax/rex* $\rightarrow$ *tax/rex* + β_4 *Rex*	$S_4(t) = k_{T4}\ \beta_4\ q_2(t) = \beta'_4 q_1(t)$

Therefore, the rate equations for *Tax* and *Rex* are:

$$\frac{dq_3}{dt} = S_3(t) - D_3(t) = \beta'_3 q_2(t) - k_{03} q_3(t) \quad (A.15)$$

$$\frac{dq_4}{dt} = S_4(t) - D_4(t) = \beta'_4 q_2(t) - k_{04} q_4(t) \quad (A.16)$$

where $q_3(t)$ and $q_4(t)$ represent the concentrations [molecules/l] of *Tax* and *Rex*, respectively, and k_{03} and k_{04} are the rate constants of the corresponding degradation reactions.

The kinetics of the incompletely spliced transcripts: analogously to *tax/rex*, the rate equations of the incompletely spliced RNAs are:

$$\frac{dq_{5i}}{dt} = S_{5i}(t) - D_{5i}(t) = f_{5i} g(q_1(t), q_4(t), z)\ k_s\ q_1(t) - k_{05i}\ q_{5i}(t) \qquad \text{with i=1,..,n} \quad (A.17)$$

where $q_{5i}(t)$ represent the concentrations of incompletely spliced RNAs, k_{05i} are the rate constants of the corresponding degradation reactions, and f_i are the fractions of nuclear *gag* singly spliced into each transcript (or remained unspliced in the case of the cytoplasmatic *gag*) and transferred to the cytoplasm.

From the rate equations (Eq. 13-17), a system of differential equations was derived. To limit the complexity of the model, a number of assumptions were made:

Simplification 1: Grone et al.[14] observed that the total viral RNA in the sample is *Rex*-independent. From Eq. A.13, A.14 and A.17, total RNA is:

$$\int_0^t \left(\frac{dq_1}{d\tau} + \frac{dq_2}{d\tau} + \Sigma_i \frac{dq_{5i}}{d\tau}\right) d\tau = \int_0^t [\, S_1(\tau) - D_1(\tau) - D_2(\tau) - \Sigma_i D_{5i}(\tau)] d\tau \quad (A.18)$$

If the nuclear degradation $D_1(t)$ is negligible with respect to the other addends, the total viral RNA does not depend on *Rex*, thus leading to k_{01} in Eq. A.13 much smaller than the transcript degradation rates. For simplicity, k_{01} was set to 0 in Eq. A.13.

Simplification 2: we assume that the multiplicity of infection is so low that $\frac{m}{f(p_1, q_3(t))} \approx 0$ in Eq. A.13. Since the MOI can be regulated at the time of virus delivery by the titer of viral particles, this condition is attainable.

Simplification 3: we assume that the nuclear-cytoplasmatic transport take place with the same rate for every transcript, i.e. $k_s \approx c_2 k_s$ in Eq. A.13.

Simplification 4: since the kinetics of incompletely spliced RNAs do not affect the possible oscillatory behavior of *tax/rex* and *gag*, their kinetics were excluded from further considerations.

From Simplifications 1-4, we obtain the system of differential equations of Eq. A.19-22:

$$\begin{cases} \frac{dq_1}{dt} = m\,S + m\,S\,\beta'_1 \frac{q_3(t)^2}{h_1^2 + q_3(t)^2} - k_s q_1(t) & (A.19) \\ \frac{dq_2}{dt} = \{1 - g(q_1(t), q_4(t), z)\}\, k_s q_1(t) - k_{02}\, q_2(t) & (A.20) \\ \frac{dq_3}{dt} = \beta'_3 q_2(t) - k_{03} q_3(t) & (A.21) \\ \frac{dq_3}{dt} = \beta'_4 q_2(t) - k_{04} q_4(t) & (A.22) \end{cases}$$

References

1. L. S. Weinberger, R. D. Dar, M. L. Simpson, *Nat Genet* **40**, 466 (2008).
2. T. S. Gardner, C. R. Cantor, J. J. Collins, *Nature* **403**, 339 (2000).
3. M. B. Elowitz, S. Leibler, *Nature* **403**, 335 (2000).
4. M. Kaern, W. J. Blake, J. J. Collins, *Annu Rev Biomed Eng* **5**, 179 (2003).
5. J. Hasty, J. Pradines, M. Dolnik, J. J. Collins, *Proc Natl Acad Sci U S A* **97**, 2075 (2000).
6. J. Hasty, F. Isaacs, M. Dolnik, D. McMillen, J. J. J. Collins, *Chaos* **11**, 207 (2001).
7. M. Ramachandra, et al., *Nat Biotechnol* **19**, 1035 (2001).
8. J.W. Bainbridge, et al., *N Engl J Med*. **358**,2282 (2008)
9. A. Corradin, et al., *The 10th International Conference on Systems biology,* Aug 30-Sept 4, 2009. Stanford, California (accepted).
10. M. Thattai, A. van Oudenaarden, *Proc Natl Acad Sci U S A* **98**, 8614 (2001).
11. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183 (2002).
12. D. T. Gillespie, *Annu Rev Phys Chem* **58**, 35 (2007).
13. M.D. Lairmore, G. Franchini, *In Fields Virology, Fifth Edition. Ed. David M. Knipe and Peter M. Howley. Lippincott Williams and Wilkins, Philadelphia*, pp. 2071-2106 (2007).
14. M. Gröne, et al., *Virology* **218,**316 (1996).
15. R. Rosin-Arbesfeld, et al., *EMBO J.* **22,**1101 (2003).
16. J. Lewis, *Current biology : CB* **13**, 1398 (2003).
17. L. S. S. Weinberger, T. Shenk, *PLoS Biol* **5** (2006).
18. F. Kashanchi, J.N. Brady, *Oncogene* **24**,5938(2005).
19. J. M. Peloponese Jr, et al. *J Virol.* **78**,11686 (2004).
20. S. J. Jeong, et al., *Biochem Biophys Res Commun.* **381**,294(2009)
21. N. Yildirim, M. C. Mackey, *Biophys J* **84**, 2841 (2003).
22. K. D. Jacobsen, B. M.Willumsen. *J. Mol. Biol.* **252**, 289 (1995)
23. J. F. Kugel, J. A. Goodrich, *Proc Natl Acad Sci U S A* **95**,9232 (1998).
24. A. Corradin, et al., *Proceedings of the ENFIN Symposium at the Functional Genomics & Disease Conference*, Oct 2-4, 2008. Innsbruck, Austria
25. F. Rende, et al., *Proceedings of the HERN meeting*, June 1-2, 2008. Bruges, Belgium.
26. A. Dhooge, et al., *ACM TOMS.* **29**,141(2003).
27. T. S. Olson, J. F. Dice, *Curr Opin Cell Biol.* **1,**1194 (1989).
28. Y.A. Kuznetsov, *Elements of Applied Bifurcation Theory, Springer-Verlag, New York* (2004).
29. D. T. Gillespie, *Journal of Computational Physics* **22**, 403 (1976).
30. D. T. Gillespie, *The Journal of Physical Chemistry* **81**, 2340 (1977).
31. S. Ramsey, D. Orrell, H. Bolouri, *J Bioinform Comput Biol* **3**, 415 (2005).
32. S. Mauch, M. Stalzer, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **99**, (5555).
33. A. Bachmair, D. Finley, A. Varshavsky, *Science (New York, N.Y.)* **234**, 179 (1986).
34. S. Rogers, et al., *Science* **234**,364 (1986).
35. M. Rechsteiner, S. W. Rogers, *Trends in biochemical sciences* **21**, 267 (1996).
36. K. E. McGinness, T. A. Baker, R. T. Sauer, *Mol Cell* **22**, 701 (2006).
37. T. A. Kunkel, *Proc Natl Acad Sci U S A* **82**,488 (1985).
38. J. W. Taylor, et al., *Nucleic Acids Res.* **13**,8765 (1985).
39. J. Stricker, et al., *Nature* (2008).
40. F. Rende, et al., *Proceedings of the 14th ICHR: HTLV and Related Viruses.* July 1-4, 2009. Salvador, Brazil.
41. J. Hasty, et al., *Phys Rev Lett.* **88**,148101 (2002).
42. T. Zhou, et al., *Chaos* **18**,037126 (2008).
43. F. Tie, et al., *J Virol.* **70,**8368 (1996).
44. M. Burton, et al., *J Virol.* **74**,2351 (2000).
45. K. A. Connors, *Chemical Kinetics, the study of reaction rates in solution*.VCH Publishers (1991)
46. L. Fu, et al., *FEBS Lett.* **396,**47 (1996).
47. R. E. Smith, et al., *Virolog.* **237**,397 (1997).
48. F.G. Helfferich, *J. Phis. Chem.* **93**,6676 (1989)
49. D. Palmeri, M.H. Malim, *J Virol.* **70**,6442.