CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD MONTERREY

Study of the molecular mechanisms involved in HIV latency

Tesis presentada por:

M. en C. Luis Ubaldo Aguilera de Lira.

Para obtener el grado de:

Doctor en Ciencias.

En la especialidad de:

Ingeniería y Física Biomédicas.

Director de tesis:

Dr. Jesús Rodríguez González.

Monterrey, N.L. Mayo 2015

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD MONTERREY

Estudio de los mecanismos moleculares implicados en la latencia del VIH

Tesis presentada por:

M. en C. Luis Ubaldo Aguilera de Lira.

Para obtener el grado de:

Doctor en Ciencias.

En la especialidad de:

Ingeniería y Física Biomédicas.

Director de tesis:

Dr. Jesús Rodríguez González.

Monterrey, N.L. Mayo 2015

This dissertation is respectfully dedicated to Prof. Dr. Ursula Kummer and Dr. Jesús Rodríguez González.

Acknowledgements

Thanks to CONACYT for have supported this project.

I want to thank to Dr. Mosés Santillan Zerón, Dr. Sven Sahle, Dr. Hilda J. Mercado Uribe and Dr. Bruno A. Escalante Acosta for the thorough review of this dissertation.

I want to express my sincere thank you to the scientific and administrative staff in Cinvestav-Monterrey for helping me during all the time of my studies.

I wish to express my sincere thanks to the people in the department of Modeling of Biological Processes in Bioquant at Heidelberg University, for make me feel an important part of you during my internship.

I place on record, my sincere thank you to my family (my parents Raymundo and Martha, and my wife Linda), for the continuous encouragement to believe in myself and fulfill my goals.

Abstract

Even in the presence of successful antiretroviral therapy stalling the progress of AIDS, developing a cure for this disease is still an open question. One of the major barriers to viral eradication is the presence of latent HIV reservoirs in the infected patients. It is suggested that eradicating latent HIV reservoirs will provide the next major step forward in controlling the HIV pandemic. Although, the fundamental role of HIV latency in the disease its origins are only partially understood. This thesis aims to increase the quantitative understanding of the molecular mechanisms suggested as responsible for latency. In the first part of this thesis, we study the effects of the virus and host interactions. In the second part, we will look at the effects of the genotypic variability in the mechanisms of the virus gene expression. Both studies implemented computational models to describe the virus and host interactions. Those models were solved under deterministic and stochastic dynamics. Additionally, massive sets of parameter values were evaluated to ensure the robustness in the predictions derived from those models. The results obtained support two independent mechanisms able to dictate the fate in an infected cell. The first mechanism is given by the bistable dynamics that arise from the network motif that appears by the mutual negative virus and host interactions. The second mechanism is given by bimodal dynamics encoded in the mechanisms of HIV gene expression.

Resumen

La terapia antiretroviral inhibe la progresión del SIDA. Sin embargo, el desarrollo de una cura para esta enfermedad aún es una pregunta abierta. Una de las principales barreras para la erradicación vírica es la presencia de reservorios latentes de VIH en los pacientes. Se ha sugerido que la eliminación de dichos reservorios latentes será el siguiente mayor paso en el control de la pandemia por VIH. A pesar del papel fundamental de los reservorios latentes de VIH en la progresión de la enfermedad, sus orígenes son parcialmente conocidos. Esta tesis tiene como objetivo incrementar el entendimiento cuantitativo de los mecanismos moleculares responsables de la latencia en el VIH. En la primera sección de este proyecto fueron estudiados los efectos de la interacción molecular entre las biomoléculas del virus y el hospedero. En la segunda parte fueron estudiados los efectos de la variabilidad genética en los mecanismos de expresión génica del VIH. En ambos estudios fueron desarrollados modelos computacionales, estos modelos fueron resueltos usando dinámicas deterministas y estocásticas. Adicionalmente, múltiples combinaciones de parámetros fueron evaluados para garantizar la robustez de las predicciones generadas. Los resultados obtenidos sugieren dos mecanismos independientes capaces de dictar el destino celular. El primer mecanismo está dado por dinámicas biestables que son originadas por un motivo en la red de interacción entre el virus y el hospedero. El segundo mecanismo está dado por las dinámicas bimodales que están codificadas en los mecanismos de expresión génica del VIH.

Table of contents

List of figures

List of tables

List of abbreviations

Abbreviation	Definition
AIDS	Acquired Immune Deficiency Syndrome
HIV	Human Immunodeficiency Virus
HAART	Highly Active Retroviral Therapy
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
ssRNA	single strand RNA
tRNA	transfer RNA
mRNA	messenger RNA
LTR	Long terminal repeat
μL	micro liters
Kb	Kilo base pairs
nm	nanometer
CD4	Cluster of differentiation 4
T4 cells	Lymphocytes CD4+
IFN	Interferon
GFP	Green Fluorescent Protein
IRFS	IFN regulatory factors
JAK1	Tyrosine kinases Janus kinase 1
TYK2	Tyrosine kinase 2
ISGF3	IFN-stimulated gene factor 3
ISREs	IFN-stimulated response elements
ISG	IFN-stimulated genes
V	HIV proteins
C, CRF	Cell Restriction Factors
VNP	Virus Non-Productive State or Latent State
VP	Virus Productive State
r	Speed
V	Volume
Ω	System size
m	Particle mass
d	Radius
Δ	Change

k_B	Boltzmann's constant
N_A	Avogadro's constant
T	Temperature
G	Gibbs energy
H	Enthalpy
S	Entropy
N	Total number of chemical species
n	Molecular count
μ	Specific chemical reaction
M	Total reaction channels
S	Chemical species
R	Chemical reaction or reaction channel
$\varnothing$	Non consider chemical species
φ	Stoichiometric matrix
c	Reaction parameter
k_μ	Constant of reaction
X	State of the stochastic system
Z	State of the determinist system
CME	Chemical Master Equation
CLE	Chemical Langevin Equation
RRE	Rate Reaction Equation
SSA	Stochastic Simulation Algorithm
DSSA	Delay Stochastic Simulation Algorithm
DDE	Delay Differential Equation
ODE	Ordinary Differential Equation
a_μ	Propensity
P	Probability
η	Biochemical noise
t	Time
τ	Time interval
θ	Time delay
ss	Steady state
λ	Eigenvalue
$\boldsymbol{I}$	Identity matrix
$\boldsymbol{v}$	Eigenvector
J	Jacobian matrix
J	Flux
U	Potential function

1 Theoretical framework

The Human Immunodeficiency Virus (HIV) targets and kills cells involved in the innate and adaptive immune response (T lymphocytes, macrophages and dendritic cells). Damage in the immune system leaves the body susceptible to infection with a wide range of opportunistic pathogens. This condition is called acquired immune deficiency syndrome (AIDS) (Carter & Saunders 2007). Treating patients with an HIV infection has strongly progressed over the last few years. The most fertile event in HIV research was achieved in 1996 with the introduction of highly active retroviral therapy (HAART). HAART consist on a combination of competitive and noncompetitive inhibitors that target specific steps in the HIV intracellular replication cycle (Choudhary & Margolis 2011). Nevertheless, HAART does not eliminate HIV infection, if treatment is interrupted the virus rebounds and hence lifetime HIV infection persists (Davey Jr. et al. 1999; Havlir et al. 2005; Maldarelli et al. 2007; Palmer et al. 2008). The main source of HIV persistent infections and the major barrier for HIV eradication is the existence of infected cells carrying latent reversible non-productive provirus, a state termed latency (Chun et al. 1997; Finzi et al. 1997). It is suggested that eradicating latent HIV reservoirs will provide the next major step forward in controlling the HIV pandemic (Coiras et al. 2009; Chomont et al. 2009). During the last decade, important experimental and theoretical efforts have increased the understanding of the causal molecular mechanisms responsible for HIV latency.

This thesis aims to design and solve mathematical models that represent those key molecular mechanisms. The aim of those models is to increase the quantitative understanding on how viruses and host interaction networks affect latency. A long-term goal of this thesis is to provide theoretical evidences to help researchers and pharmacologist to discover new therapeutic targets in HIV therapy.

1.1 Molecular Biology of HIV

1.1.1 Virion characteristics

The HIV virion size is a macromolecule of around 100 nm (see Figure 1). Each virion is made of: a lipid membrane, membrane proteins gp41 and gp120, a nucleocapsid, a Reverse

Transcriptase, Integrase, Tat protein, tRNA and two single strand RNA (ssRNA) positive genomes (Carter & Saunders 2007).

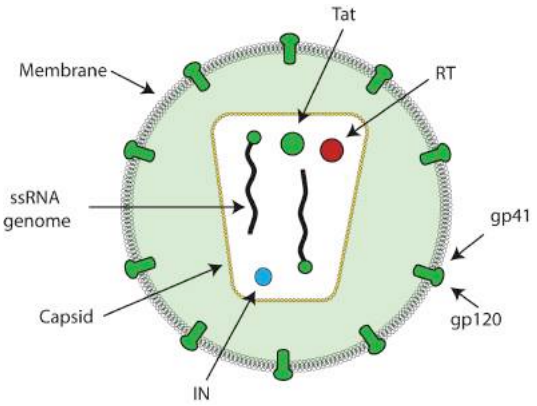


Figure 1. Elements in the HIV virion
Integrase (IN), Nucleocapsid (NC), Reverse Transcriptase (RT), and the glycoproteins gp41and gp120.

HIV ssRNA genome has 9.3 kb in length (see Figure 2), which consists of nine open reading frames that code for nineteen proteins. All three reading frames in the HIV genome are used and there is extensive overlapping. Multiple splicing events lead to two additional short sizes of HIV mRNA (van Dijk et al. 2010).

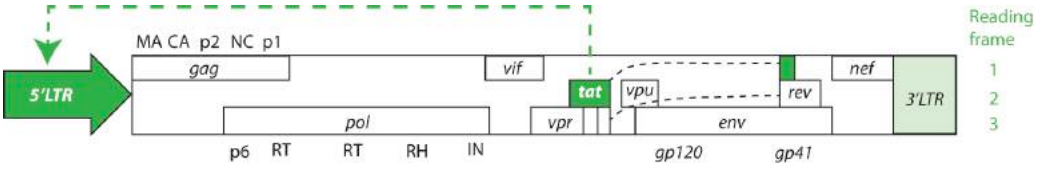


Figure 2. The HIV genome
HIV genome consists on nine open reading frames that code for nineteen different proteins. The three reading frames are used. By alternative splicing the genes tat and rev are produced.

1.1.2 HIV replication circle

The HIV infection starts with the interaction between the virus and the host cell. This interaction is given by the attachment of the viral protein gp120 and the cell receptor CD4. Then, the membranes of the virion and the cell are fused. The contents of the virion are released into the cell cytoplasm. In the cytoplasm the synthesis of the viral DNA is carried out by the Reverse Transcriptase. After reverse transcription has been completed, the pre-integration complex (viral DNA and proteins) is moved to the nucleus. In the nucleus, the Integrase cuts the DNA of the cell chromosome and integrates the provirus into the gap. Integrated provirus acts as a transcription template and its expression is divided in early and late phase. In the early phase, viral transcripts are spliced and Nef, Tat and Rev proteins are

expressed. Nef reduces the detection of HIV-infected cells from immune surveillance, Tat induces provirus transcription and Rev promotes the expression of late genes. In the late phase, Gag and Pol polyproteins are produced. Then, from those polyproteins the accessory proteins Vif, Vpr and Vpu are cleaved. In the last step of the virus replication cycle, the viral proteins and genome become anchored to the plasma membrane. The capsid is polymerized and structural proteins gp120 and gp41 are cleaved from the polyprotein Env. Finally, after the virion packages all its components it crosses the cell membrane obtaining in this way its lipid envelope (Dimmock et al. 2007; Madigan et al. 2010; Reddy & Jhon 1999). HIV replication cycle is given in Figure 3.

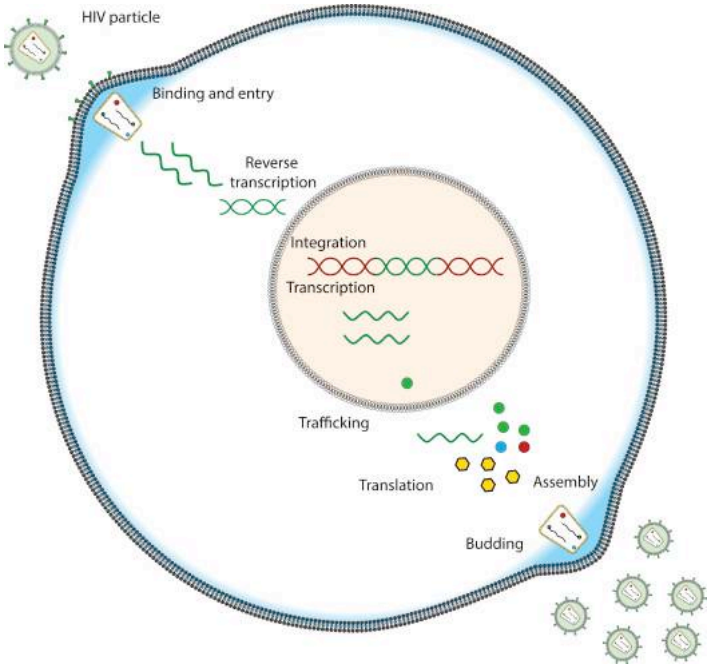


Figure 3. The HIV replication cycle
The HIV replication cycle consists on the following steps: Attachment and entry, reverse transcription and transport to the nucleus, integration of the provirus, transcription and genome replication, assembly and release. Image made using data from the following references (Dimmock et al. 2007; Madigan et al. 2010; Reddy & Jhon 1999).

1.2 HIV infection

After HIV diagnosis, the disease progression in the patient is continuously monitored using clinical markers. The essential clinical markers include the quantification viral load and lymphocytes CD4+, T4 cells. Those markers indicate the state of the immune system and replication state of the virus. The periodic monitoring of those markers is essential to prevent HIV-associated symptoms and as a guideline to initiation of therapy (Vajpayee & Mohan 2011).

1.2.1 Dynamics of HIV infection without HAART

Patients without therapy show extensive infection of immune cells and a rapid viral growth. Several weeks after the beginning of the HIV infection the viral load reaches a maximum concentration. This peak of viral infection stimulates the innate and the specific immune systems. Immune system stimulation leads to virus neutralization. During the following months after the initial infection, the virus and immune cell concentrations stabilize reaching an asymptomatic phase, see Figure 4 A (O'Connell et al. 2009). Nevertheless, during this apparent stable phase, T4 cell concentrations continue decreasing, and the viral concentration increases. After some years of infection, the system descends into AIDS when the patient exhibits severe immune degradation. AIDS is reflected when a T4 cell concentration in the bold is below 200 cells/ μL , see Figure 4 B (Wasserstein-Robbins 2010).

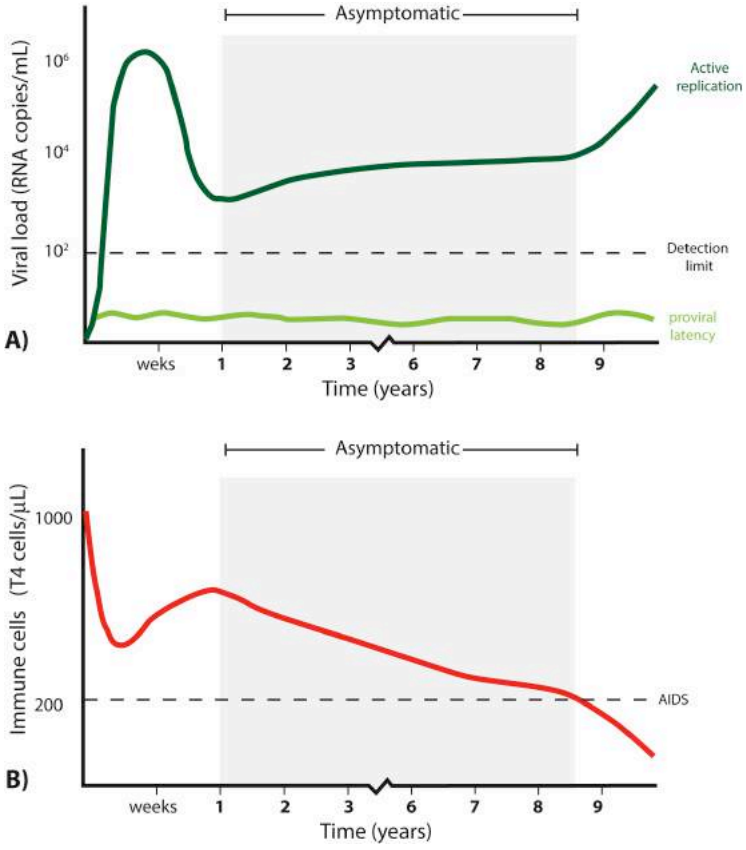


Figure 4. Dynamics of the HIV infection without HAART
A) Representative progression of plasma viral load (green line). Time post-infection is indicated in weeks and then in years. The first weeks show a peak in the viral concentration. Then, an asymptomatic phase of infection is present during the following years (gray window). Here the virus concentration reaches an apparent steady state. Notice that during all infection a constant presence of proviral latency is present. B) Representative dynamics of the T4 cell counts (red line). The progression of the infection is characterized by the continuous decrease in immune cells, progression into AIDS occurs when T4 cell counts drop to <200 cells/ μL . Image made with data from reference (O'Connell et al. 2009; Wasserstein-Robbins 2010)

1.2.2 Dynamics of HIV infection with HAART

HIV dynamics in patients under HAART is characterized by the decrease in viremia and a temporal recovery in the concentration of immune cells. As is shown in Figure 5 A. After, HAART administration the decrease in viremia is abrupt. This viral decrement is caused by the death of infected active T4 cells, macrophages and dendritic cells that are the main source of virion production (Perelson et al. 1997). Nevertheless, during the time HAART is administrated there are low levels of viral concentration, this remaining viremia is maintained by the presence of infected long half-life resting memory T4 cells that will remain in the patient for life. On the other hand, given the action of HAART the viral active replication is stopped and hence the immune system recovers and maintains a constant concentration, see Figure 5 B. Notice, that the removal of HAART is associated with a rebound in the viral infection, this last phenomenon is given by the stochastic activation of resting memory T4 cells infected in latent state that survived during the time that HAART was applied (Finzi et al. 1999; Ho et al. 2013).

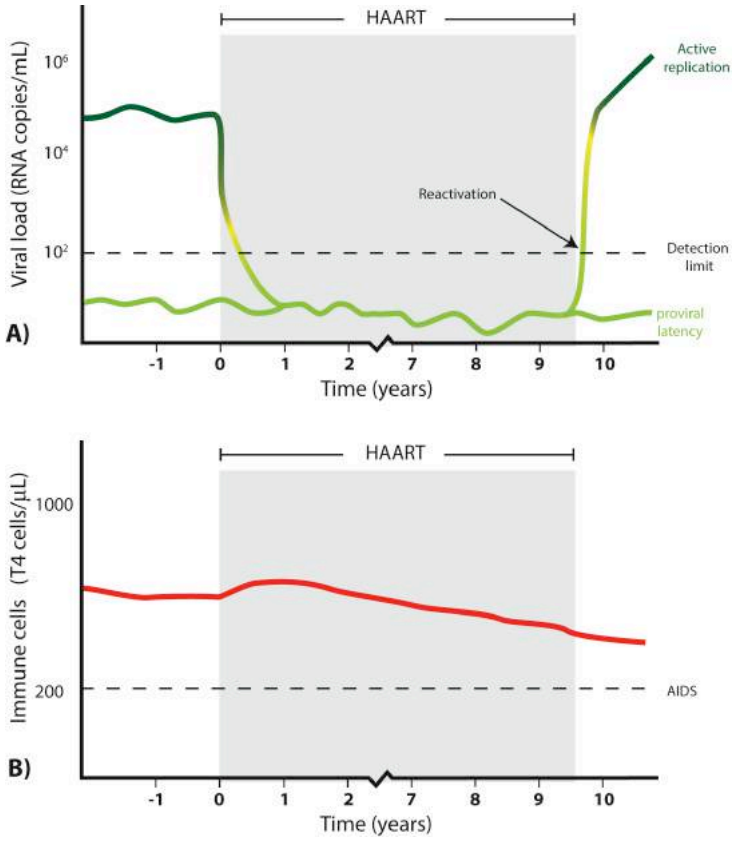


Figure 5. Dynamics of the HIV infection with HAART
A) Representative progression of plasma viral load (green line). After the HAART the decay in the viral levels is given for the death of infected activated T4 cells, macrophages and dendritic cells. Basal levels of HIV are maintained by the presence of latently infected T4 cells in resting memory state (light green line). After HAART

removal there is a rebound in the concentration of the viral load. B) Representative dynamics of the T4 cell counts (red line). The immune system is in apparent steady state where the concentration of the immune cells is maintained constant. Image made with data from reference (Palmer et al. 2011; Finzi et al. 1999; Ho et al. 2013).

1.3 HIV latency

The final cell fate in an HIV infection is given by a probabilistic decision between the viral productive state or latent state, see Figure 6 (Marcello 2006; Weinberger & Weinberger 2013). The viral productive state is characterized by an active expression of viral gene products (Kim & Yin 2005). HIV latency is characterized by an integrated provirus that is replication competent but expresses no or not all viral gene products (Donahue & Wainberg 2013).

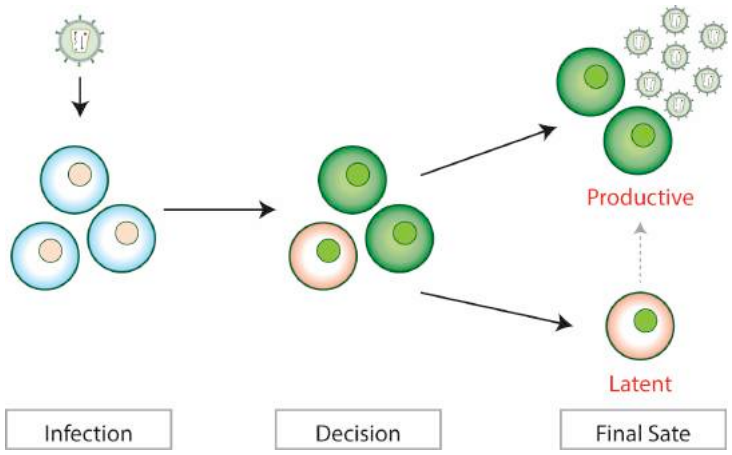


Figure 6. Model for the establishment of latency and reactivation
Each viral infection is a probabilistic decision the output could be entering latency or the viral productive state (image modified from reference (Weinberger & Weinberger 2013; Donahue & Wainberg 2013).

1.3.1 Molecular mechanisms responsible of HIV latency

1.3.1.1 Tat circuit

HIV-Tat transactivation is a key mechanism involved in HIV latency (Bohan, C. A., F. Kashanchi, B. Ensoli, L. Buonaguro, K. A. BorisLawrie 1992). HIV-Tat transactivation is a positive feedback loop given between the HIV-5'LTR and HIV-Tat protein (see Figure 7). Transactivation increases until 100-fold HIV-transcription over basal expression levels (Nabel & Baltimore 1987). Absence of HIV-Tat protein results in premature termination of the transcription. Overexpression of HIV-Tat protein is sufficient to induce virus production from resting latently infected cells (Donahue et al. 2012).

The central role of HIV-Tat protein in HIV-latency was confirmed by studies conducted by Weinberger's workgroup. They hypothesized that the low molecular counts of Tat proteins at the beginning of infection exhibits large fluctuations in the transactivation processes. Hence highly noisy transactivation processes dictate a latent or productive infection (Razooky & Weinberger 2011; Weinberger et al. 2005). To prove this hypothesis they used reduced HIV-based vectors that coexpressed the green fluorescent protein (GFP) and the HIV-Tat protein controlled by the HIV-5'LTR promoter into $CD4^+$ Jurkat T cells obtaining two cell phenotypes: bright and off, that resemble the productive and latent HIV-infected cell states. Then, modeling the Tat-circuit and adjusting it to experimental data it was determined a monostable architecture, that after its solving in a stochastic regime showed bimodal dynamics (Weinberger et al. 2005).

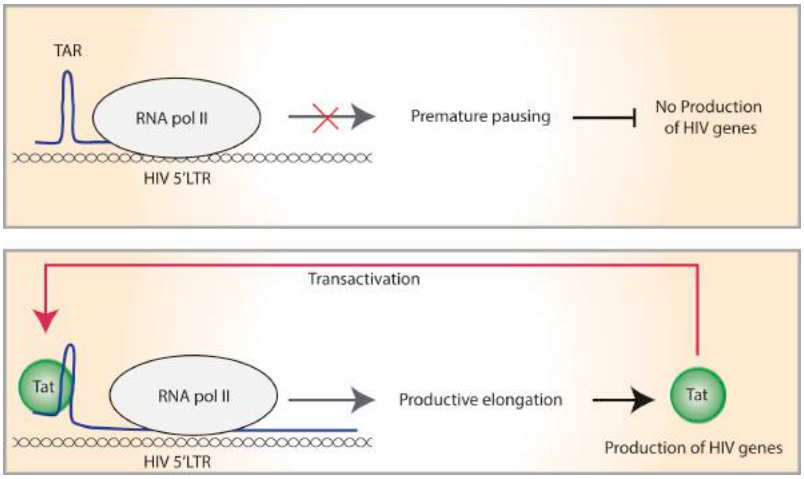


Figure 7. Representations of the transcriptional Tat circuit
A) Premature pausing in the transcription of the HIV genome. This abrupt pause is given by the formation of a secondary structure in the new synthetized viral transcripts. B) Tat protein acts as a transcriptional factor binding the new synthetized transcripts, this relaxes the abortive secondary structure in the new synthetized RNA allowing the expression of complete viral transcripts. Complete viral transcripts will produce the viral elements starting a productive infection. One of those viral products is the same Tat protein. Tat protein form a positive feedback circuit in the expression of HIV, this last process is termed transactivation.

Subsequently, the same group proved that HIV expression is dictated by the time that the Tat protein remains in the inactive (desacetylated) state, which is longer than the time that it remains in the active state (acetylated), see Figure 8. In that study, they experimentally and theoretically tested the effects of the inactive state in the Tat protein. In the theoretical model they considered a time delay in the transactivation reaction. The aim of this time delay was to represent the large differences between states in the Tat protein. After testing delayed stochastic dynamics they observed a system with two possible outcomes. In the first outcome transactivation is dissipated and latent infection was established. Conversely, in the second

outcome periods of transactivation were demonstrated and a viral active infection was established (Weinberger & Shenk 2007).

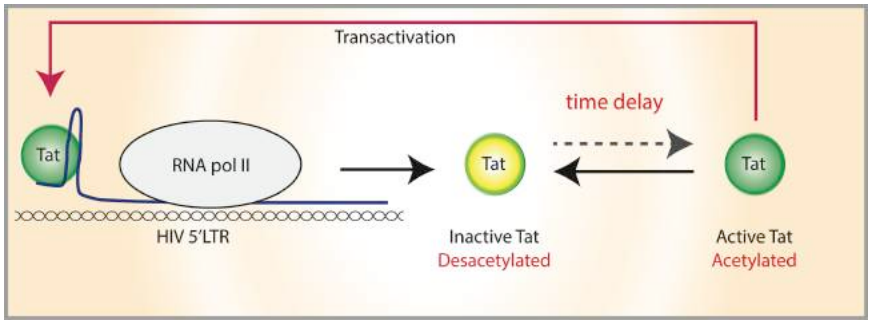


Figure 8. The active and inactive states in the Tat protein
The transitions between states are given at different rates, generating periods where the inactive state is more common than the active state. Image made with data taken from reference (Weinberger & Shenk 2007).

1.3.1.2 Tat genotypic variability

It is suggested that Tat genotypic variants occur as important variables of viral replication and hence affect the cell fate after HIV infection (Mayne et al. 1998; Burnett et al. 2009; Zhu et al. 1993). Tat protein sequence diversity is generated by low fidelity of reverse transcriptase and selective pressures applied by the host immune response and HAART (Knipe et al. 2007). Different Tat protein sequences from HIV infected patients show variation in transactivation activity (van Marle & Power 2005; Mayne et al. 1998).

1.3.1.3 The innate immune response

Cellular innate immune responses provides a powerful defense against viral infections (Ning et al. 2005). It is suggested that innate immune responses could prevent the emergence of virus from latency (Di Santo 2001; Haaland et al. 2005). Innate immune responses that interfere with HIV-intracellular dynamics are characterized by cell restriction factors (CRF). CRF are host cellular biomolecules that block specific stages of the HIV replication cycle (Duggal & Emerman 2012). Most CRF have an interferon (IFN)-induced expression (Barr 2010).

1.3.1.3.1 The IFN system

The IFN system is composed of two phases, early and late (Takeuchi & Akira 2009). In the early phase, pathogen recognition receptors, such as RIG-I, stimulate an intracellular signaling cascade that leads to the activation of IRFs 3 and 7, resulting in their homo- or

heterodimerization (Rand et al. 2012). These dimers then translocate to the nucleus and induce the expression of IFN (Honda & Taniguchi 2006). Then, in the late phase (also known as the JAK STAT signaling pathway) new synthesized IFN bind the IFN receptor (IFNAR). Subsequently, IFNAR activate the cytoplasmic tyrosine kinases JAK1 and TYK2, which phosphorylate the signal transducer and activator of transcription 1 (STAT1) and STAT2. Phosphorylated STAT1 and STAT2 dimerize and translocate to the nucleus, where they assemble with IRF9 to form a trimolecular complex called IFN-stimulated gene factor 3 (ISGF3). ISGF3 binds to the IFN-stimulated response elements (ISREs) in the cell genome. (Ivashkiv & Donlin 2014) Multiple ISREs in the cell genome are stimulated and the activation transcription of a set of several hundred IFN-stimulated genes (ISGs) ocurrs. Some of those ISGs are part of the intrinsic immune response. The specific genes with antiviral activity are termed cell restriction factors (CRF). CRF are directly responsible for establishing a cellular antiviral state, see Figure 9 (Honda et al. 2005)(Horner & Gale 2013).

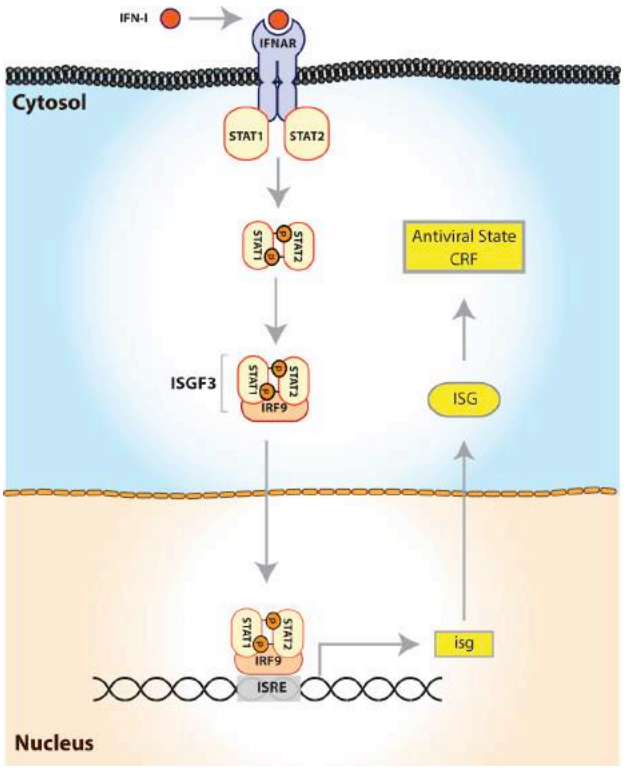


Figure 9. The JAK-STAT signaling pathway
Here, IFN type I binds IFNAR. Subsequently, IFNAR activate JAK1 and TYK2 that phosphorylate STAT1 and STAT2. Phosphorylated STAT1 and STAT2 dimerize and assemble with IRF9 to form ISGF3, and then it translocate to the nucleus. ISGF3 induces the transcription of Interferon-Stimulated Genes (ISG). Some ISG function as cell restriction factors (CRF) establishing the antiviral cellular state. Image made with data taken from the following references: (Ivashkiv & Donlin 2014; Honda et al. 2005; Horner & Gale 2013).

1.3.1.3.2 Virus and host interactions

The cellular stratagems against the virus are moderated by HIV encoded countermeasures. This complex intricate strike-counterstrike interactions govern the ultimate outcome of HIV infection, see Figure 10 (Strebel et al. 2009). On one hand, experimentally validated CRF that interfere with HIV-transcription include: Murr1 (Ganesh et al. 2003), ISG20 (Espert et al. 2005) and PKR (Sadler et al. 2009). CRF that interfere with HIV-translation are OAS1/RNAseL (Maitra & Silverman 1998) and specific cellular microRNAs (Cojo et al. 2011; Sanghvi & Steel 2012). CRF that interfere during the HIV trafficking/budding steps are shown by Tim22 (Barr et al. 2008), Tetherin/ BST-2 (Perez-Caballero et al. 2009) and ISG15 (Okumura et al. 2006). On the other hand, experimentally validated counter inhibitions provided by HIV-proteins towards CRF include: Vpu inhibitions against Tetherin/BST-2 (Perez-Caballero et al. 2009). Tat inhibitions against microRNAs pathways (Cojo et al. 2011; Sanghvi & Steel 2012). Tat inhibitions against OAS1/RNAseL (Schröder et al. 1990). HIV-antagonistic activity against ISG20 (Espert et al. 2005). And the indirect mechanisms carried out by Vpu and Vif that independently interfere with the IFN system (Donahue & Wainberg 2013).

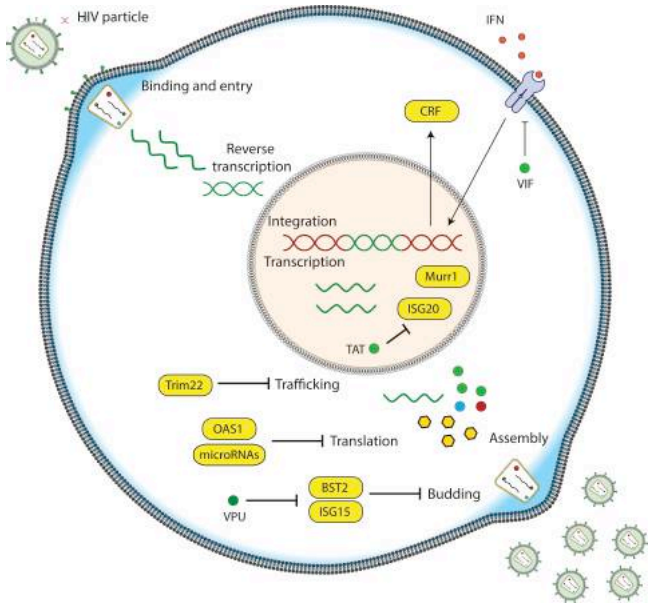


Figure 10. HIV replication cycle considering virus and host interactions
The negative interactions between the virus and host generate a complex biochemical network. In this network HIV intracellular dynamics are determined by the inhibitory effect of CRF in the HIV-replication cycle, and counter-inhibitions of HIV-proteins against CRF. Representative examples of CRF that interfere at postintegration steps include: human microRNAs, Murr1, OAS1/RNAseL, Trim22, ISG15, ISG20 and Tetherin/ BST-2. On the other hand, HIV-proteins that have evolved the ability to antagonize CRF are: Tat and Vpu. In the graph yellow boxes represent CRF. Lines with hammer ends represent negative interactions. Arrows represent positive interactions.

1.4 Fundamentals of chemical kinetics

1.4.1 Elementary and complex reactions

Biochemical networks, such as metabolism, signaling pathways, gene regulation or virus replication cycles are complex webs of interaction among multiple biomolecules. Biochemical networks can be reduced to chemical reactions and can be studied using chemical kinetics (Klipp et al. 2011). Any complex biochemical network can be broken down into one or multiple steps that belong to three elementary processes: association, dissociation and conversion. Association or synthesis is given when two molecules form a complex. Dissociation or decomposition is given when a bond is broken down and a molecule dissociates into two different species. Conversion describes when a molecule is modified in some of its components thus becoming another one (Mcnaught & Wilkinson 1997).

Biochemical networks can be reduced by applying a quasi-steady-state approximation in some intermediary steps, in this way it is possible to lump multiple elementary reactions in a single step (Keizer 1987). Those lumped processes are non-elementary reactions that are termed complex reactions. For example, after assumptions of quasi-steady-state approximation in intermediary steps Michaelis–Menten kinetics (or Hill functions) can represent enzymatic reactions.

1.4.2 Thermodynamics of chemical reactions

Before we formulate a mathematical (deterministic or stochastic) description of a biochemical network, it is necessary to understand why chemical reactions in biochemical networks occur.

1.4.2.1 Collision theory

According to collision theory a chemical reaction will only takes place between two molecules if they collide. But, first it is important to understand why particles collide. Microscopically, any system in the liquid phase consists of particles in constant random motion (thermal energy). The speed, r , of each one individual molecule follows a Maxwell-Boltzmann distribution that depends on the temperature of the system: the higher the

temperature, the greater the mean kinetic energy of the particles. The probability density function for the speed of the molecules in the system is given by:

$$f(r) = \sqrt{\left(m/2\pi k_B T\right)^3}\, 4\pi v^2 e^{-\frac{mv^2}{2k_B T}}. \tag{1}$$

where, m is the particle mass and $k_B T$ is the product of Boltzmann's constant and temperature in Kelvin. The average speed of the molecules can be calculated by:

$$\langle r \rangle = \int_0^\infty r\ f(r)\, dr = \sqrt{k_B T/\pi m}. \tag{2}$$

In Figure 11 can be observed the shape of the Maxwell-Boltzmann distribution displaying that particles have a very wide range of speeds. A general idea obtained from the curve is that very few particles have low speed, whereas most of the particles have speed values near the average and few particles possesses a kinetic energy above a given threshold.

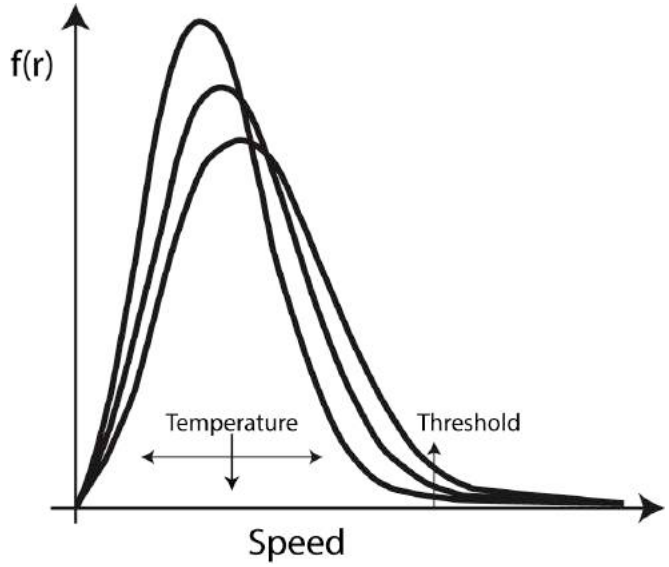


Figure 11. The Maxwell-Boltzmann distribution
Maxwell-Boltzmman distribution specifies the number of particles having a particular speed. The example shows three independent graphs made using equation (1) with three arbitrary values for the temperature. A general idea obtained from the shape in the three curves is that very few particles have low speed values whereas most of the particles have average speed and a fraction possesses a kinetic energy above a given threshold to achieve a successful reactive collision. The shape of the distribution is dependent of the temperature. Higher temperature decreases the amplitude in the distribution and at the same time increases its dispersion. It is important to mention that strictly, this graph only applies to gases, but in general the idea behind it can be applied to reactions involving liquids.

Then, considering a system composed of molecular species, S_1, inside a volume, V. As can be observed in Figure 12 a collision occurs whenever the center-to-center distance between two molecules decreases to $d_{12} = d_1 + d_1$, where d_1 represents the radii of the molecules.

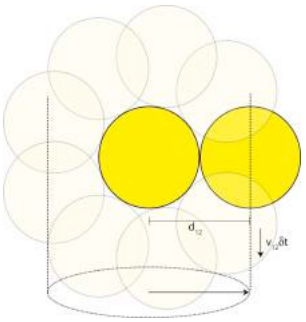


Figure 12. The collision volume between two molecules
A molecular collision is defined when the distance between two molecules is reduced to d_{12} . In the infinitesimal time interval δt the molecules form the collision volume.

By picking an arbitrary S_1 molecular pair it can be calculated the rate at which such collisions are occurring. First, denoting by r_{12} the speed of one molecule relative to the other. Then, in the next small time interval δt , it is defined a collision volume V_{coll} at time δt by:

$$\delta V_{Coll} = \pi d_{12} r_{12} \delta t. \tag{3}$$

Now, substituting the average particle speed obtained from the Maxwell-Boltzmann distribution and dividing by the total volume it is obtained the average probability that this particular molecular pair will collide in the next time interval δt :

$$\left\langle \delta V_{Coll} / V \right\rangle = V^{-1} \pi d_{12} \sqrt{k_B T / \pi m_{12}} \delta t, \tag{4}$$

where, $m_{12} = m_1 m_2 / \left(m_1 + m_2 \right)$. Here, the brackets denote the average.

When does a collision result in a reaction? As it can be easily intuited: in a constant volume and temperature a major amount of molecules results in a major number of collisions, as is illustrated in Figure 13. However, it is important to stress that not all collisions result in a chemical reaction. Most collisions just result in nonreactive collisions. To obtain a successful reactive collision the molecules need to collide in the right way, and with enough energy to solve or form new bonds (Laidler 1987).

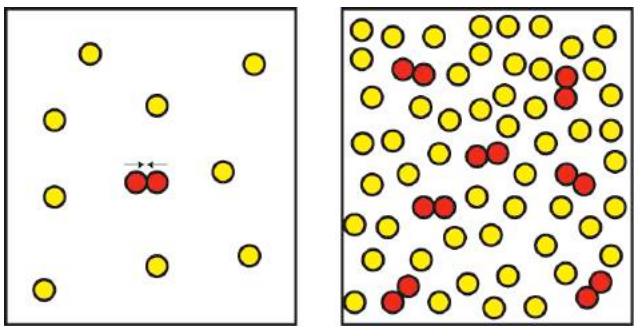


Figure 13. Collision theory

Two systems of chemical species homogeneously distributed in a compartment of constant volume V and at thermal equilibrium. Yellow circles represent the reactants and two bounded red circles represent the products. In the left a low number of reactants decreases the rate of collision reflecting a few successful reactive collisions. On the right, a large number of reactants increase the chances of successful reactive collisions and hence a major number of products are formed.

1.4.2.2 Transition state theory

Exactly, how much energy must a collision have to achieve a chemical reaction? Transition state theory explains this as follows: during the course of a reaction, reactants transit by a continuum of energy states, as presented in Figure 14, and at some stage they must pass through a barrier of maximum energy (transition state) in which bonds are solved or newly formed concluding in the formation of products. The potential barrier determines the rate at which the process occurs. When the barrier is low, the needed energy to pass this barrier is low and hence the reaction will be fast. However, when the barrier is high, only a few reactants will have sufficient energy, and the reaction will be slower (Zhou 2010).

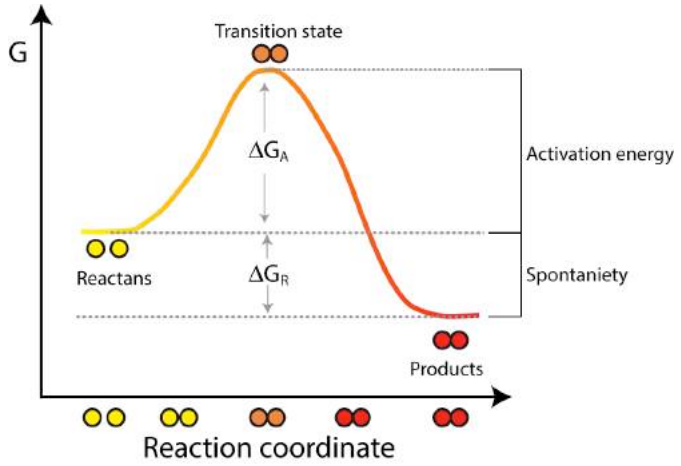


Figure 14. Reaction profile according to transition state theory
The x-axis indicates the reaction coordinate containing different states of the reactants and products, and the y-axis indicates the change in free energy along the course of the reaction. In the graph ΔG_A represents the activation energy, that is, the needed energy that the reactants must pass in order to the process occurs, notice that the maximum is the unstable transition state. Finally, ΔG_R describes the spontaneity of the reaction. In this case because $\Delta G_R<0$ it is exergonic process.

To describe the driving force of a chemical reaction it is used the Gibbs energy (G), which is the energy capable of carrying out work. Gibbs energy is made of the balances between enthalpy (H) minus the product of the temperature and the entropy (S, a measure how much the energy of atoms and molecules become more spread out in a process). So, the Gibbs energy is given by:

$$G = H - TS. \tag{5}$$

and the change in this quantity is described by:

$$\Delta G = \Delta H - T\Delta S. \tag{6}$$

This difference describes the activation energy, ΔG_A, this is the difference between the free energy in the reactants and the energy in the transition state. An empirical mathematical expression to describe the rate of the reaction in function of the temperature and the activation energy is given by the Arrhenius equation:

$$\tilde{c}(T) = A_0 \exp^{\left(-\frac{\Delta G_A}{k_B T}\right)} \tag{7}$$

where, A_0 is the pre-exponential factor.

1.4.2.3 The reaction parameter

In summary, combining the collision theory and transition state theory it can be obtained a mathematical expression termed the reaction parameter, c, that describes the rate of the reaction:

$$c(T) = V^{-1}\pi d_{12}\sqrt{8k_B T/\pi m_{12}}\exp^{\left(-\frac{\Delta G_A}{k_B T}\right)}. \tag{8}$$

From now on, this function will be considered as a constant, nevertheless, it is important to empathize that it is a function dependent on the temperature. However, this assumption will be valid because most of the biological processes are given at constant temperatures or present temperature variations in an extremely constrained range.

1.5 Stochastic dynamics

As previously mentioned, on a microscopic scale a chemical reaction is given by a reactive collision between reactants. Given the constant stochastic motion of the solvent and reactant molecules, chemical reactions are characterized by two properties: i) a random occurrence in time, and ii) no correlations between successive reactions. Satisfying those properties it is obtained a Markov process that describes a system evolution only dependent on the actual state. As can be seen in Figure 15 the time evolution of a chemical system is caused by a stochastic variable in the continuous time that satisfies the Markov property (Adrews et al. 2009).

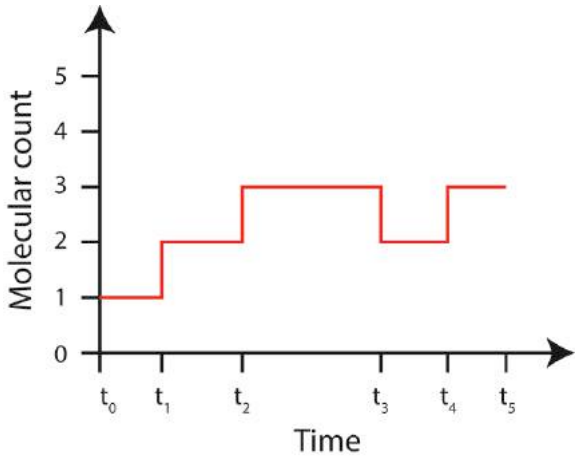


Figure 15. Time evolution of a Markov process in the continuous time.

To formally describe the state of a chemical system made by molecular species $\{S_1,\ldots,S_N\}$ that chemically interact through $M \geq 1$ reaction channels $\{R_1,\ldots,R_M\}$, where each chemical reaction is represented as follows:

$$R_\mu : s_1 \cdot S_1 + ,\ldots, + s_N \cdot S_N \rightarrow products. \tag{9}$$

where, s_N represent the number of S_N molecules involved in the reaction.

The state of a chemical system is given by:

$$X(t) = (X_1(t),\ldots,X_N(t)), \tag{10}$$

where, $X_i(t)$ is a discrete random variable that represent the number of S_i molecules in the system at time t.

To describe the dynamics of $X(t)$ given an initial state $X(t_0) = x_0$, the fundamental premise of stochastic chemical kinetics was introduced by (Gillespie 1976), this premise describes the average probability that a particular combination of S_i reactant molecules will react in the next time interval $[t + dt)$. This is given by:

$$P_A \equiv a_\mu(x)dt, \tag{11}$$

where, the propensity function a_μ is the product of the reaction parameter c_μ and the number of reactant combination h_μ, that is: $a_\mu(x) = c_\mu h_\mu(x)$.

Using the propensity function and the stoichiometric coefficient, $\varphi_{i\mu}$, that indicates the number of molecules of each reactant and product appearing in the reaction equation; the reaction channel R_μ is completely described.

1.5.1 The Chemical Master Equation

The time evolution of $X(t)$ is a homogeneous Markov process in continuous time. To analyze this process; take the conditioned probability function:

$$P(x,t|x_0,t_0) = \mathrm{P}\{X(t) = x | X(t_0) = x_0\} \tag{12}$$

Then, suppose that $x(t)$ is the system's state at time $t > t_0$, the value of each propensity function $a(\mu, x(t)) > 0$, for each index μ, it is represented as follows:

$$x(t) \xrightarrow{a(\mu,x(t))} x(t) + \varphi[\mu], \tag{13}$$

notice, that for each time increase dt, the product $a(\mu,x(t))dt$ it is the probability that μ-reaction occurs in the infinitesimal time interval $[t+dt)$.

Then, to determine $P(x;t+dt)$ is needed to determine the sum of all the probabilities in which the system can arrive to state $x(t+dt)$.

Defining the different forms to arrive to state x as:

$x(t) \longrightarrow x(t+dt)$	with probability	$1 - \sum_{\mu=1}^{M} a(\mu,x)dtP(x,t),$
$x(t) - \varphi_1 \longrightarrow x(t+dt)$	with probability	$a(1, x-\varphi_1)dtP(x-\varphi_1,t),$
$x(t) - \varphi_2 \longrightarrow x(t+dt)$	with probability	$a(2, x-\varphi_2)dtP(x-\varphi_2,t),$
$x(t) - \varphi_\mu \longrightarrow x(t+dt)$	with probability	$a(\mu, x-\varphi_\mu)dtP(x-\varphi_\mu,t).$

Notice, that the sum of all probabilities is equal to one. Then, the probability of the state $x(t+dt)$ at time $t+dt$ is calculated by:

$$P(x,t+dt) = P(x,t)\left[1 - \sum_{\mu=1}^{M} a(\mu,x)dt\right] + \sum_{\mu=1}^{M} a(\mu, x-\varphi_\mu)dtP(x-\varphi_\mu,t). \tag{14}$$

In equation (14) the fist term represent the probability that no reaction occurs in the time interval $[t+dt)$ and the second term represents the probability that the μ reaction occurs in the time interval $[t+dt)$, starting in state $x-\varphi_\mu$ and ending in the state x.

Regrouping terms in equation (14), dividing by *dt* and applying the limit $dt \to 0$ it is possible to obtain the temporal dynamics of the probability $P(x,t)$. Obtaining in this way the CME. The CME is a differential–difference equation (differential in time and difference in states) that describes the probability that the system has a specific copy number for each type of chemical species at a given point in the future (Gillespie 2007). The CME is:

$$\frac{\partial P(x,t)}{\partial t} = \lim_{x \to 0} \frac{P(x,t+dt) - P(x,t)}{dt}, \tag{15}$$

replacing terms it is obtained the expanded form of the CME:

$$\frac{\partial P(x,t)}{\partial t} = \sum_{\mu=1}^{M} \left[a(\mu, x-\varphi_\mu) P(x-\varphi_\mu, t) - a(\mu,x) P(x,t) \right], \tag{16}$$

where, the two terms within brackets give the rate at which the probability of being in state x increases or decreases over time because of reactions into or out of state x, respectively.

For a system of *N* interacting chemical species with *n* molecules the CME is made by a set of coupled ODEs for each state of the system. Considering the combination of n and N, a very rough approximation of the size of the state space can be given by:

$$states = n^N. \tag{17}$$

In a simple biochemical system (for example a bacterial cell) the number of chemical species (proteins) is in the order of thousands. So, the numbers of possible system's states even in simple biochemical systems is huge. For that reason, attempting to solve numerically or analytically the CME for realistic systems is infeasible or in many cases impossible (Qian & Bishop 2010). Nevertheless, there is an alternative way to not explicitly solve the CME but still exactly genrate the stochastic time evolution of a chemical system. This is determined by the repetitive generation of many individual trajectories of the CME.

1.5.2 The Stochastic Simulation Algorithm

To achieve this alternative solution, the problem it is formulated in a different way (Gillespie 1976). First, it is introduced the next-reaction probability density function that describes the probability that the next reaction in the system will occur in the infinitesimal time interval $[t+\tau, t+\tau+d\tau)$ and the reaction will be R_μ:

$$P(\tau,\mu|x,t)d\tau = P_A(\tau)\cdot P_0(\tau) \tag{18}$$

where, $P_0(\tau) = \exp^{-\sum_{\mu=1}^{M} a_\mu(x)\tau}$, this is the probability that no reaction occurs during the interval $(t, t+\tau)$. Then, it is possible to arrive to the exact expression for the reaction probability density function:

$$P(\tau,\mu) = a_\mu(x)\cdot \exp^{-a_0(x)\tau} \tag{19}$$

with, $a_0(x) = \sum_{\mu=1}^{M} a_\mu(x)$.

Equation (19) provides the basis for a stochastic simulation approach. It implies that τ is an exponential random variable with mean (and standard deviation) $1/a_0(x)$, while μ is a statistically independent integer random variable with point probabilities $a_\mu(x)/a_0(x)$.

Then, by Monte Carlo procedures, samples of τ and μ are generated according to these distributions. For example, using the direct method, τ and μ are given as:

$$\tau = (1/a_0(x))\ln(1/u_1) \tag{20}$$

$$\sum_{i=1}^{\mu-1} a_i(x) < u_2 a_0(x) \le \sum_{i=1}^{\mu} a_i(x) \tag{21}$$

where, u_1 and u_2 denote two uniformly distributed random numbers in the unit interval. To have a better understanding of this selection processes see Figure 16.

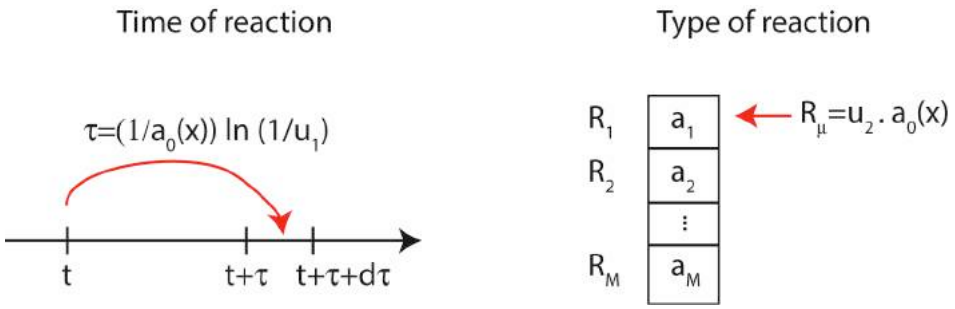


Figure 16. Time of the next reaction and the selection of the next reaction
After repetitively select those quantities and update the system it is possible to reconstruct a trajectory of the CME.

Iteratively generating values for (τ,μ) and updating the system in each iteration step it is obtained the stochastic simulation algorithm (SSA) that constructs an exact numerical realization of the CME (Gillespie 1976). The steps in the SSA are:

0. Begin the simulation with $(x = x_0, t = t_0)$.

1. Evaluate all $a_\mu(x)$ and $a_0(x)$.

2. Generate values for (τ,μ) using equations (20) and (21).

3. Update the system by replacing $(t \leftarrow t+\tau)$ and $(x \leftarrow x+\varphi_\mu)$.

4. Save (x,t) and return to Step 1, or else end the simulation.

1.5.3 The Chemical Langevin Equation

The previous described CME and SSA exactly describe the stochastic dynamical behavior of a chemical system. Nevertheless, there are still computational limitations to fully implement those methods in realistic systems. In this section, an approximate method based on stochastic differential equations is discussed (Gillespie 2000).

For this approximated method the system needs to fulfill the leap condition, this condition establishes that it is possible to define a time interval τ to be small enough that the change in the propensity functions values are not affected during the interval $[t,t+\tau)$; and at the same time this τ must be large enough that the expected number of occurrences of each reaction channel R_μ in $[t,t+\tau)$ to be much larger than 1 (the leap condition is depicted in Figure 17). Notice that both conditions are apparently contradictory but for systems with reactant

populations sufficiently large both conditions can be satisfied simultaneously (Gillespie 2000).

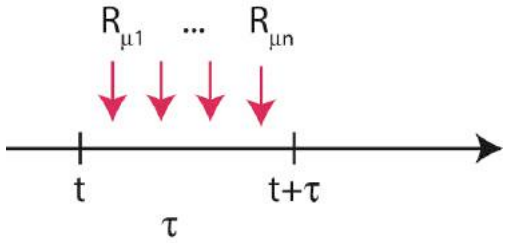


Figure 17. The leap condition in a chemical system
The leap condition is given by a time interval τ that is small enough that all propensities functions are not affected after this time interval, at the same time, this interval must ensure that multiple reaction occur.

The leap condition is necessary because if $a_\mu(x)$ remains constant during $[t,t+\tau)$ the number of times that the reaction channel R_μ occurs in $[t,t+\tau)$ is a Poisson random variable with mean and variance $a_\mu(x)\tau$. Therefore, it can be approximately leap the system ahead by a time τ by:

$$X(t+\tau) = x(t) + \sum_{\mu=1}^{M} \tilde{P}_\mu\left(a_\mu(x)\tau\right)\varphi_\mu, \tag{22}$$

where, $\tilde{P}_\mu(\vartheta_\mu)$ is a Poisson random variable with mean and variance ϑ_μ.

Now, considering the fact that a Poisson random variable with large mean and variance can be approximated as a normal random variable with that same mean and variance, it is possible to arrive to the canonical form of a Langevin equation. For chemical systems, it is termed Chemical Langevin Equation (CLE):

$$X(t+\tau) = x + \sum_{\mu=1}^{M} \varphi_\mu a_\mu(x)\tau + \sum_{\mu=1}^{M} \varphi_\mu \sqrt{a_\mu(x)}\zeta_\mu(0,1)\sqrt{\tau}, \tag{23}$$

where, $\zeta_\mu(0,1)$ is a normal random variable with mean 0 and variance 1.

Then, deriving a time-evolution equation for equation (23) the temporal dynamics of the system are calculated by the following stochastic differential equation:

$$\frac{d}{dt}X(t) = \sum_{\mu=1}^{M} \varphi_\mu a_\mu\left(X(t)\right) + \sum_{\mu=1}^{M} \Gamma_\mu(t)\varphi_\mu \sqrt{a_\mu\left(X(t)\right)}, \tag{24}$$

where, the first term is simply the reaction rate equation. The second term adds Gaussian noise to the deterministic result. Here, $\Gamma_\mu(t)$ represents a temporally uncorrelated, statistically independent Gaussian white noise with mean 0 and variance 1.

In the thermodynamic limit the species population $X(t)$ and the volume V approach to infinity, as the propensities are proportional to $X(t)$ they also tend to infinity. So, in the thermodynamic limit the last term in equation (24) becomes negligibly small compared with the other term obtaining a deterministic system (Pahle 2009).

1.6 Deterministic dynamics

A deterministic system is a system in which no randomness is involved in the development of future states, thus, always producing the same result from a given starting condition (Ullah & Wolkenhauer 2011). If the chemical system defined in (10) resides on a constant volume V and the molecules populations of all involved chemical species are large enough to neglect fluctuations, $X_i(t)$ can be approximate to a continuous quantity termed concentration:

$$Z_i(t) = X_i/\Omega. \tag{25}$$

where, for molar concentration the system size is: $\Omega = N_A V$, and N_A is the Avogadro's constant. Or, alternatively defining $\Omega = V$ it is obtained the number of molecules per volume (Mcnaught & Wilkinson 1997).

And generalizing the deterministic state of the complete chemical system at time t is given by:

$$Z(t) = (Z_1(t),\ldots,Z_n(t)), \tag{26}$$

The basis of the deterministic approach is the mass-action kinetics, where the reaction velocity is proportional to the product of the substrate concentrations to the power of their respective stoichiometric coefficient, $\varphi_{i\mu}$. So, the rate of each reaction R_μ is given by the following kinetic function:

$$f_\mu(Z_1,\ldots,Z_n) = k_\mu \prod_{i=1}^{n} Z_i^{\varphi_{i\mu}}. \tag{27}$$

where, k_μ is the rate constant of reaction R_μ, that is associated to reaction parameter as follows:

$$k_\mu = V \cdot c_\mu, \tag{28}$$

1.6.1 ODE systems

A chemical system comprises one ordinary differential equation (ODE) for each S_i. The vector form that describes the dynamics of the entire system is termed the rate reaction equation, RRE:

$$\frac{d}{dt}Z_i(t) = \sum_{\mu=1}^{M} \varphi_{i\mu} \cdot f_\mu, \tag{29}$$

here, the right hand sides of these equations contain one summand for each μ. Each summand in turn is the product of the kinetic function of the μ reaction and the $\varphi_{i\mu}$ (Adrews et al. 2009).

1.6.2 Solving the ODE-system

The general solution for the previous ODE-system is achieved by finding the functions that fulfill equation (29), this solution depend on the integral constants and represents a set of curves that describe all possible states in the state space. A particular solution is derived from the general solution by setting the parameters to particular values and setting the initial conditions. This particular solution describes a trajectory inside the state space. Solving the system under the condition $dZ/dt = 0$ gives the steady states (or fixed points) that are points in the state space where the system is not changing with the time. The time course represent the dynamics of the variables in the time and it can be obtained by the numerical, or if possible, by the analytical integration of the system equation (Klipp et al. 2011; Murray 2002).

1.6.3 ODE-system linearization

Linearization refers to find a linear approximation to the system at a given point. This is a needed condition before the calculation of the local stability for a nonlinear system. The linear approximation of a nonlinear system is arrived at by the following procedure (Strogatz 2001). First, it is calculated the steady states of the nonlinear system by analytical or numerical methods. Then, considering a deviation ξ from those steady states Z_i^*. Obtaining:

$$Z = Z^* + \xi \tag{30}$$

Then, calculating the temporal change of this deviation by:

$$\frac{d}{dt}Z = \frac{d}{dt}\left(Z^* + \xi\right), \tag{31}$$

reorganizing and expanding terms in the previous expression it is obtained the temporal change of the deviation:

$$\frac{d}{dt}\xi_i = f_i\left(Z_1^* + \xi_1, \ldots, Z_N^* + \xi_n\right) \tag{32}$$

Then, the temporal change of the deviation can be expressed as a Taylor series:

$$\frac{d}{dt}\xi_i = f_i\left(Z_1^*, \ldots, Z_n^*\right) + \sum_{j=1}^{n} \frac{\delta f_i}{\delta Z_j}\xi_j + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \frac{\delta^2 f_i}{\delta Z_j \delta Z_k}\xi_j \xi_k + \ldots, \tag{33}$$

considering steady state and neglecting higher order terms, the previous expression is reduced to:

$$\frac{d}{dt}\xi_i = \sum_{j=1}^{n} \frac{\delta f_i}{\delta Z_j}\xi_j = \sum_{j=1}^{n} a_{ij}\xi_j \tag{34}$$

where, the coefficients a_{ij} form the so-called Jacobian matrix, $\mathbf{J}$:

$$\mathbf{J} = \mathrm{a}_{ij} = \begin{pmatrix} \frac{\delta f_1}{\delta Z_1} & \cdots & \frac{\delta f_1}{\delta Z_n} \\ \vdots & \ddots & \vdots \\ \frac{\delta f_n}{\delta Z_1} & \cdots & \frac{\delta f_n}{\delta Z_n} \end{pmatrix} \tag{35}$$

Notice that after apply those previous concepts a linear system is obtained:

$$\frac{d}{dt}Z = \mathbf{A}Z \tag{36}$$

where, $\mathbf{A}$ is a square matrix.

1.6.4 ODE system stability

After linearizing the ODE system is possible to determine its stability, that is, the robustness of the dynamics of the system under small perturbations to its initial conditions (Strogatz 2001). To determine stability at a given point in the linear system it is applied the Eigen-decomposition theorem, which states that for some scalar λ called the eigenvalue of a linear square matrix **A** there is a corresponding eigenvector **v** that satisfy the following relation:

$$\mathbf{A}v = \lambda \mathbf{v}, \tag{37}$$

which, is equivalent to:

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = 0. \tag{38}$$

where, **I** is the identity matrix. This system has nontrivial solution if the determinant is zero:

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0, \tag{39}$$

equation (39) is known as the characteristic equation of **A**. And in a general form the characteristic polynomial is:

$$f(\lambda) = \det(\lambda \mathbf{I} - \mathbf{A}) = \lambda^n - a_1 \lambda^{n-1} + \ldots + (-1)^n a_n \tag{40}$$

where a_i is the trace of matrix **A**, a_n is the determinant of matrix **A** and a_i is the sum of the i-rowed diagonal minors of matrix **A**.

Solving equation (40) for λ it is determined the stability in the system. For $\lambda > 0$ the steady state is unstable and for $\lambda < 0$ the steady state is stable.

1.7 Bistability and bimodality

Bistability is given when the system has two stable steady states separated by unstable steady state (Markevich et al. 2004). Having two steady states the system can evolve towards one of them in dependence on the initial conditions. On the other hand, the analytical or numerical solution of a stochastic system is given by a probability distribution. If this distribution has two different modes the system shows bimodality (Craciun et al. 2006). A representative example showing the bimodality and bistability concept is given in Figure 18.

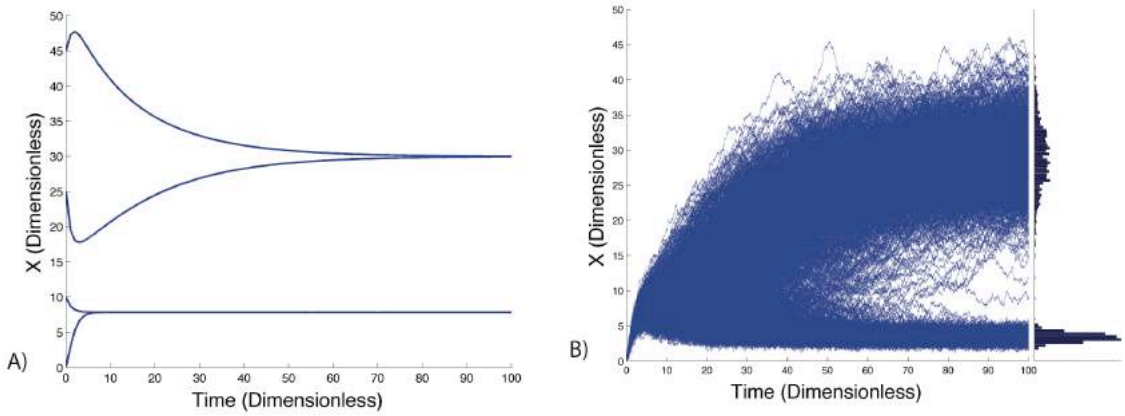


Figure 18. Representative example showing bistability and bimodality in a system
A) Example of a dynamical system with deterministic bistability. The system is given by the following differential equations: $\dot{X} = k_1 - X\left(Y^2 / \left(1 + Y^2\right)\right)$; $\dot{Y} = k_2 - Y\left(X^2 / \left(1 + X^2\right)\right)$. After numerical solving the system using different initial conditions the system evolves to two different steady states. B) Solving the equivalent stochastic system by repetitively running the SSA and building histograms for the trajectories at the final time it is obtained bimodal distributions. Image obtained by numerical simulations in Matlab.

Most of the times, bimodality and bistability can be used interchangeably, that is, most systems with bistability show bimodal dynamics in the stochastic regime. Nevertheless, there are some exceptions where purely stochastic dynamics without deterministic bistabilities can show bimodal dynamics (Artyomov et al. 2007).

2 Objectives

2.1 General objective

- To increase the quantitative understanding of the HIV and host molecular mechanisms involved in HIV latency.

2.2 Particular objectives

- To study how interactions between the virus and host biomolecules affect HIV latency.
- To determine how biochemical noise affects HIV latency.
- To determine how cellular innate immune responses affects HIV latency.
- To determine how the mechanisms of the virus gene expression affect HIV latency.
- To determine how genotypic variability in the mechanisms of virus gene expression affects HIV latency.
- To determine how latently infected cells influence the life span of a culture of infected cells.

3 Model I. Virus and host interactions

3.1 Model aim

In this project it is hypothesized that: the productive and latent cell phenotypes in a HIV infected cell are caused by the interplay between the virus and host biomolecules. To test this hypothesis the following specified theoretical questions were formulated:

1) Are the virus and host negative interaction enough to generate the latent and productive HIV phenotypes?
2) How does biochemical noise affects the cell fate?

3.2 Model construction

3.2.1 Deterministic models

A mathematical model was developed as a tool to solve the hypothesis. The model was developed using existing biological information regarding to the HIV replication cycle and the interaction between the virus and host biomolecules. The specific details of those interactions were given in Section 1.3.1.3.2 and in Figure 10. Then, an abstraction of the process was carried out, this abstraction consisted in the combination and lumping of the intermediate molecules and feedback loops in this pathway. In this way, it was obtained an elemental circuit that summarizes the complex regulatory network can be reduced into a model with two components, the HIV proteins (V) and the CRFs (C) in an HIV-infected cell. The main characteristics of this circuit are a positive feedback loop in V (basal expression and Tat transactivation) and mutual negative feedback loops between V and C (the mutual negative interaction between the innate immune response and HIV proteins).

This model consists of the following two differential equations that account for the temporal evolution of the concentrations of V and C, respectively:

$$\frac{d}{dt}V = k_{BV} + k_V\left(V/\left(V + k_M\right)\right) - k_{VC}VC - k_{dV}V, \tag{41}$$

$$\frac{d}{dt}C = k_C - k_{CV}VC - k_{dC}C. \tag{42}$$

The meaning of the functions, variables and parameters in the above equations is as follows. Parameter k_{BV} represents the basal expression of V (Nabel & Baltimore 1987). The Michaelis-Menten function $k_V\left(V/\left(V+k_M\right)\right)$ denotes the transcriptional Tat circuit (Razooky & Weinberger 2011). The parameter k_V is the maximal V production rate, and the parameter k_M is a half-saturation constant. The parameter k_{VC} represents the negative interaction of C with V (Hsu 1981). The parameter k_{dV} denotes a first-order decay rate of V (Tan et al. 2012). The parameter k_C denotes the basal production rate of C. Notice that parameter k_C contains IFN effects, $k_C = k_C$'[IFN], where k_C' and [IFN] are constants. The second term represents the negative interaction of V with C. Parameter k_{dC} represents the first-order decay of C (Tan et al. 2012).

Then, assuming quasi-steady-state in equation (42) the original system was reduced to:

$$\frac{d}{dt}V = k_V/100 + k_V\left(V/\left(V+k_M\right)\right) - k_{VC}V\left(k_C/\left(k_{dC}+k_{CV}V\right)\right) - k_{dV}V. \tag{43}$$

Notice, that parameter k_{BV} was redefined as $k_{BV} = k_V/100$, according to (Nabel & Baltimore 1987).

To analytically solve the previous minimal model, parameters were substituted by the dimensionless parameters $a_v = k_v/k_{dV}k_M$, $p_{vc} = k_C k_{VC}/k_{dV}k_{dC}$, $p_{CV} = k_M k_{CV}/k_{dC}$ and variables were substituted by the dimensionless variables $v = V/k_M$ and $\tau = k_{dV}t$ obtaining the following dimensionless model:

$$\frac{d}{d\tau}v = a_v/100 + a_v(v/(v+1)) - (p_{vc}v/(1+p_{cv}v)) - v. \tag{44}$$

Notice that parameters $p_{vc} \propto k_{VC}$ *and* $p_{cv} \propto k_{CV}$. Therefore, they represent the strength of the negative interactions between V and C.

3.2.2 Stochastic models

3.2.2.1 Langevin equation

A first stochastic model was build introducing a Gaussian white noise source, η, into equation (44) obtaining:

$$\frac{d}{d\tau}v = (a_v / 100 + a_v(v/(v+1)) - (p_{vc}v/(1+p_{cv}v)) - v) + \eta, \quad (45)$$

Subsequently, this model was used to calculate the transition between states in function of η; a formal description of this procedure is given in Section 6.1.1.

3.2.2.2 Stochastic simulations

Then, a second stochastic model was generated, here, the system given by equations (41) and (42) was converted into a reaction rate scheme (see Table 1). To obtain this reaction scheme, the terms of production and degradation in the system equations were used as elementary complex reactions. Then, it was solved by direct method of the SSA.

Table 1. Reaction scheme for the negative interaction between the virus and the host

Reactions	Description	Effective propensity
$\varnothing \rightarrow V$	Basal production of V	k_{BV}
$V \rightarrow V+V$	Production of V enhanced by transactivation	$k_V\,(V / (V + k_M))$
$C+V \rightarrow$ C	Interference competition between C and V	$k_{VC}\,V\,C$
$V \rightarrow \varnothing$	Decay of V	$k_{dV}\,V$
$\varnothing \rightarrow C$	Constant production of C	k_C
$V+C \rightarrow$ V	Interference competition between V and C	$k_{CV}\,V\,C$
$C \rightarrow \varnothing$	Decay of C	$k_{dC}\,C$

3.3 Methods

3.3.1 Initial conditions

Four initial conditions were used to describe the relevant biological cases: Case 1, a cell with an integrated HIV provirus but null expression of HIV proteins and no CRFs (*V(0)=0, C(0)=0*). Case 2, represents a system with constitutive CRF expression (*V(0)=0, C(0)=1000*) (Huang et al. 2007; Wu 2012). Case 3 represents the preintegration transcription of the HIV provirus (Wu 2004). Here, a low initial number of HIV proteins and no CRF was assumed (*V(0)=5, C(0)=0*). Finally, in case 4, was made by combining the previous two scenarios to represent a system that begins with a low count of HIV proteins and prior expression of CRFs (*V(0)=5, C(0)=1000*).

3.3.2 The virus productive (VP) and latent (Virus nonproductive, VP) states

In the simulations, it was used the dynamics of V to define a threshold to differentiate latent and virus-productive states. Latent state is characterized by null expression of viral gene products (Donahue & Wainberg 2013). In the simulations it was used a null concentration of V ($V < 1$ *Molecules/Cell*) to define the latent state. In contrast, virus-productive state is characterized by active expression of viral gene products. In the simulations it was assumed the virus-productive state if V reached values inside the biological limit determined in some HIV-biomolecules (i.e. $V \in [1000, 10000]$ Molecules/ Cell) (Kim & Yin 2005).

3.3.3 Nominal values for the parameters

All the parameters in the equations have positive ranges defined by experimental reports and are tabulated in Table 2.

Table 2. Parameter definitions and range values

Parameter	Description	Nominal Value	Range	Units	Reference
k_{BV}	Basal expression rate of V	1.34×10^{-3}	$[0, 1\times10^{-2}]$	Molecules/s	(Nabel & Baltimore 1987).
k_M	Michaelis Menten-constant	380	$[1\times10^{2}, 1\times10^{4}]$	Molecules	(Slice et al. 1992)
k_V	Maximal expression rate of V	1.34×10^{-1}	[0, 1]	Molecules/s	(Reddy & Jhon 1999)
k_C	Expression rate of C.	7×10^{-2}	$[0, 1\times10^{-1}]$	Molecules/s	(Schwanhäusser et al. 2011)
k_{VC}	Interference coefficient from C against V	2.95×10^{-2}	[0, 1]	1/Molecules*s	This study
k_{CV}	Interference coefficient from V against C	9.27×10^{-1}	[0, 1]	1/Molecules*s	This study
k_{dV}	Decay rate of V	6.85×10^{-5}	$[0, 1\times10^{-3}]$	1/s	(Reddy & Jhon 1999)
k_{dC}	Decay rate of C	5.01×10^{-5}	$[0, 1\times10^{-4}]$	1/s	(Phillips et al. 2010)

3.3.4 The complete parameter space

To determine the generality of the model results, a massive exploration in the parameter space was performed. Here, parameter values that vary in orders of magnitude were defined (see Table 2). Then 10,000 pseudorandom parameter sets were generated. Each parameter set was iteratively tested on the previously described deterministic and stochastic models. Those searches were used to sample: 1) the parameter space that reproduces bistable behavior, 2) the effects of the biochemical noise in the transition between states along the parameter space, and 3) the parameter space that reproduces bimodal behavior.

The complete theoretical design is given in the flow diagram given in Section 6.3.1.

3.3.5 Numerical methods

Function ode45 from Matlab (The Mathworks Inc. 2010) was used to solve the ordinary differential equations (ODEs), this function implements a Runge-Kutta method (4 and 5 order) with a variable time step. Symbolic methods in Matlab were used to analytically solve the system. The SSA was coded in Matlab and also tested in COPASI 4.11 (Hoops et al. 2006) , the pseudocode is given in Section 6.2.1. Parallel computing was used in Matlab to iteratively solve the stochastic models under each specific set of parameters. Additionally, analysis methods were coded to determine bimodality in the distribution histograms obtained after the SSA.

3.4 Results

3.4.1 Deterministic dynamics

First, it was tested if multiple steady states were present in the system provided by equations (41) and (42). Here, The right-hand side in equation (42) was set to zero, solved for C and substituted the result into equation (41). Obtaining three roots for V^*. Substituting these V^* values in equation (42) and solving for C, the corresponding values for C^* were obtained. Keeping all the parameter in their nominal values the steady states of the system (ss) were determined. Additionally, the stability of those steady states was determined obtaining a system with three steady states, two stable and one unstable (see Table 3). This is characteristic of a bistable system (Murray 2002).

Table 3. Steady states in the system

Steady state	Values	Stability
ss1	$V^*=8X10^{-5}$, $C^*=559$	Stable
ss2	$V^*=3.19$, $C^*=2X10^{-2}$	Unstable
ss3	$V^*=1551$, $C^*=4.8\ X10^{-5}$	Stable

Furthermore, numerical simulations were used determining that the system evolved to two different stable steady states that correspond to ss_1 and ss_3 (see Figure 19, black solid lines). Then, using the values obtained for ss_2 and two additional values close to this point as initial conditions, an unstable steady state was determined (see Figure 19, red line and gray lines).

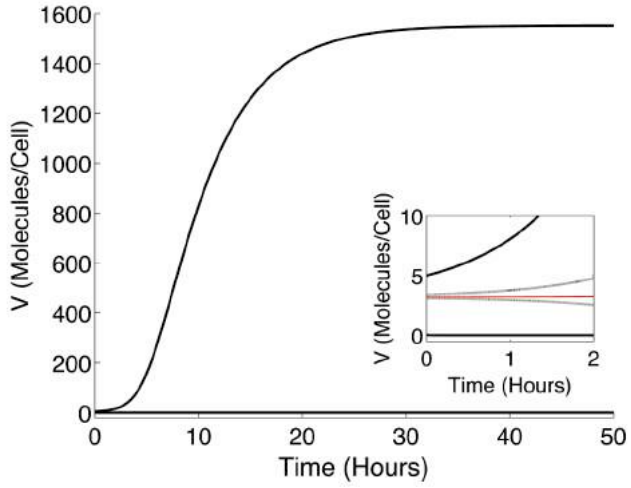


Figure 19. Deterministic numerical simulations
Deterministic simulations using two different initial conditions ($V(0)=0$, $C(0)=0$ and $V(0)=5$, $C(0)=0$) give rise to two stable states (black solid lines). An unstable steady state is reached using the initial conditions $V(0)=3.19$, $C(0)=0.02$ (red dashed line). Using two different sets of initial conditions ($V(0)=3.1$, $C(0)=0.02$ and $V(0)=3.4$, $C(0)=0.02$) close to the previous unstable steady state, the trajectories evolved to the stable states (gray lines). Inset zooms out the y-axis.

Then, the effect of the terms that represent the negative interaction between V and C in the system dynamics was determined. To do this, the complete system (equations (41) and (42)) was solved using nominal parameter values and assuming k_{VC} and k_{CV} as free parameters. Obtaining a bifurcation diagram that contains three different dynamic regions (see Figure 20). The first region is a narrow monostable region that was obtained when low values for the parameter k_{VC} were given. A second region was obtained when intermediate values of k_{VC} were given that reproduces bistable behavior. A third region was obtained when high values of k_{VC} were given that reproduces monostable behavior. In this bifurcation diagram, bistable

behavior was obtained in a wide area but it was biased toward low values of the parameter k_{VC}.

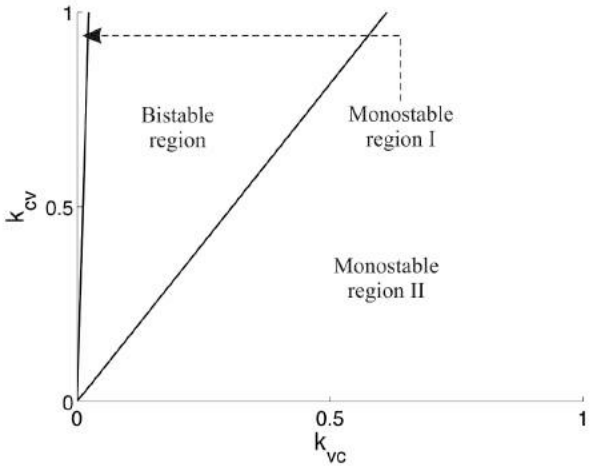


Figure 20. Bifurcation diagram
Simulations were made with the nominal values given in Table 2 and varying the undetermined parameters k_{CV} and k_{VC} in the range [0, 1]. Low values for k_{VC} reproduce a narrow monostable area. Intermediate values of k_{VC} reproduce a bistable region. Large values of k_{VC} reproduce a monostable area.

Then, it was examined whether bistability is maintained over a wide range of parameter values. After testing 10,000 pseudorandom parameters that vary in the ranges defined in Table 2, it was found that approximately 12% of those parameters reproduce bistable dynamics. A complete distribution of those parameters in the parameter space did not reflect a restriction in their values. The results are shown in Figure 21.

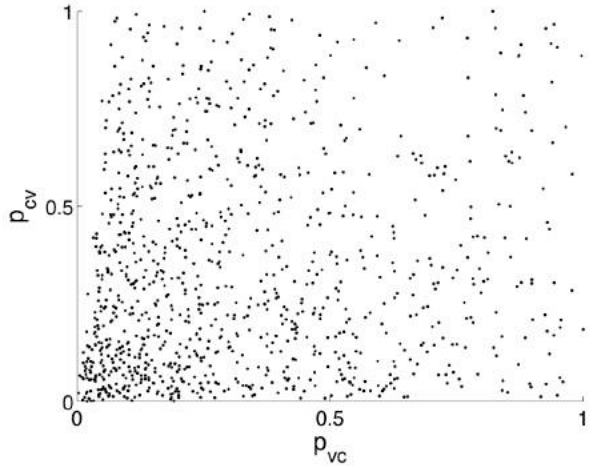


Figure 21. Parameter space that reproduces bistable dynamics
Black dots represent bistable cases. The parameters p_{vc} and p_{cv} correspond to the x- and y- axis, respectively. Every axis is normalized, and extreme values were eliminated.

Then, the distribution of steady states in parameter sets that reproduce bistable dynamics was determined. In Figure 22, every bistable case was plotted along the x-axis and the value for V that corresponds to every steady state in the y-axis. As expected, two stable steady states and one unstable steady state for each bistable case were found. For all cases, the first stable steady state was found at $V \approx 0$ (Figure 22, black dots at the bottom). Most unstable steady states were found distributed in the range $V \in [0,2000]$ (Figure 22, red dots), and the second stable steady state was found distributed over $V > 1000$ (Figure 22, upper black dots).

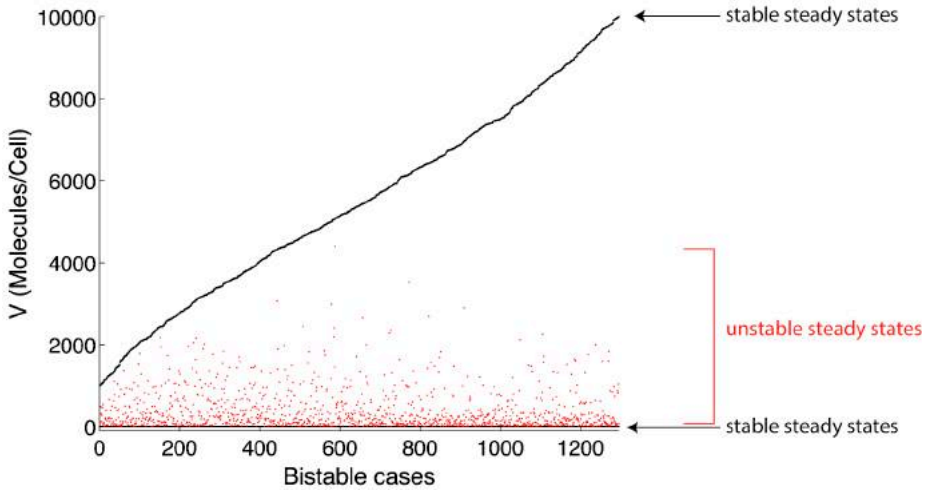


Figure 22. Distribution of steady states in each bistable case
In the graph, every x point represents a bistable case and has three vertical points that correspond to its three steady states. Red dots represent unstable steady states, while black dots represent stable steady states. Bistable cases were ordered by the V value in the third steady state.

3.4.2 Stochastic dynamics

After determining that the system reproduces bistable dynamics, it was tested how do the intensity of biochemical noise affects the transition between steady states. Using the formalism described in Section 6.1.1, a potential-like landscape with two local minima that correspond to the VNP and VP states was obtained (see Figure 23).

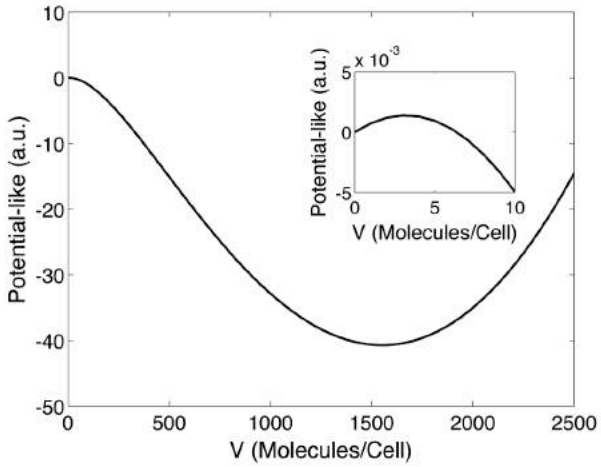


Figure 23. Potential-like landscape
Local minima correspond to stable steady states. Local maxima correspond to unstable steady states. Inset zooms on the x-axis and shows the local maxima. The representative example was made with the nominal parameter values given in Table 2.

Using the potential landscape, the transition between states was determined using a fixed value of intensity for the biochemical noise ($\eta = 0.0032$) obtaining that the system has a high probability of reaching the VP state once in equilibrium ($P(VP) \approx 1$). Then, it was determined the transition between states to all positive bistable cases and using $\eta = 0.0032$. The scatter plot in Figure 24 shows the empirical probability of obtaining the VP state as the colored points. As the inhibition from CRF towards HIV proteins becomes less effective (low nominal value in p_{VC}), the VP state becomes more probable (red dots).

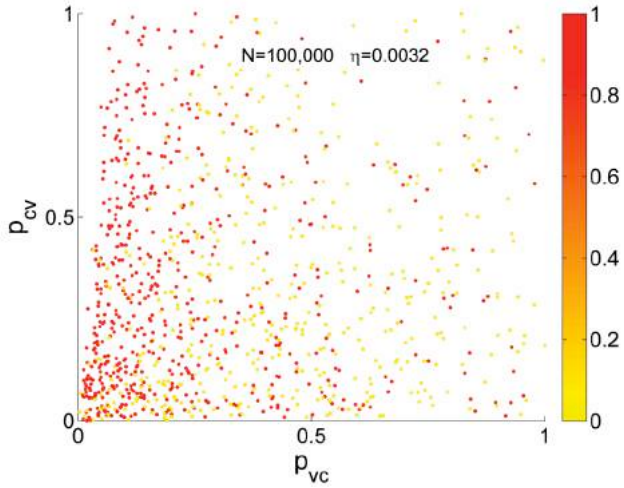


Figure 24. Scatter plot of the probability of obtaining cells in the VP state
The empirical probability of obtaining cells in the VP state according to the parameter values. Simulation results using a molecular count $N = 100000$ and $\eta = 0.0032$. In the scatter plot, each point represents a set of

parameters that resulted in a system with bistable behavior. The point color represents the empirical probability of obtaining cells in the VP state.

Then, it was determined how do different intensities of biochemical noise affect the transition between steady states. To do this, the transition between states was determined using the range $\eta \in [0.0032,1]$. Figure 25 shows the percentage of bistable cases that showed a probability of obtaining cells in the VP state higher than 0.5. This threshold indicates when it is more probable to obtain cells in the VP than in the VNP state. The results showed that as η is increased the percentage of bistable cases with a high probability of obtaining the VP state increases. In contrast, if η is decreased to values near zero, a purely deterministic regime where no decision is expected.

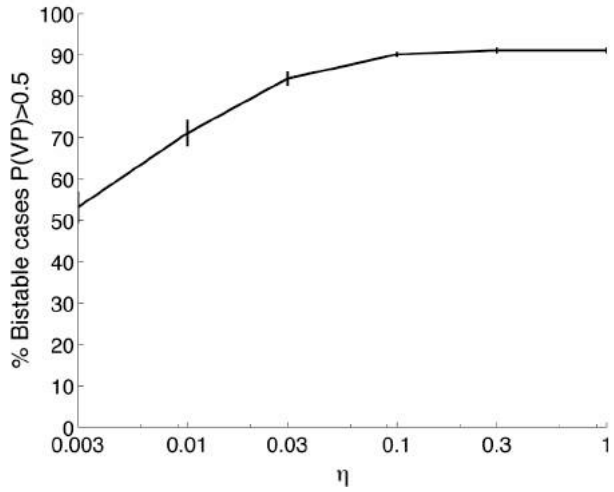


Figure 25. Cases with high probability to show the VP state
Percentage of total bistable cases with an empirical probability of obtaining cells in the VP state higher than 0.5 (P (VP)>0.5) under different intensities of η. The error bars were calculated from three repeats of numerical experiments.

To directly track the effects of biochemical noise in trajectories that represent individual HIV-infected cells, stochastic simulations were carried out. The SSA was implemented (see section 3.2.2). A first set of 1,000 trajectories was made using the nominal parameter values given in Table 2 and a simulated timespan of 50 hours. Using the initial condition (*V(0)=0, C(0)=0)*, a bimodal distribution where the VNP state was more common and a small fraction of trajectories ended in the VP state was obtained (Figure 26 A). For the second initial condition (*V(0)=0, C(0)=1000*), it was found that CRFs interfered with the transition to the VP state, and most trajectories ended in the VNP state (Figure 26 B). In the third initial

condition (*V(0)=5, C(0)=0*), it was found that an initial count of five HIV proteins produced a substantial increase in the number of trajectories that end in the VP state (Figure 26 C). In the fourth initial condition (*V(0)=5, C(0)=1000*), it was found that even with a low count of HIV proteins the transition to the VP state was inhibited if CRFs were already expressed (Figure 26 D).

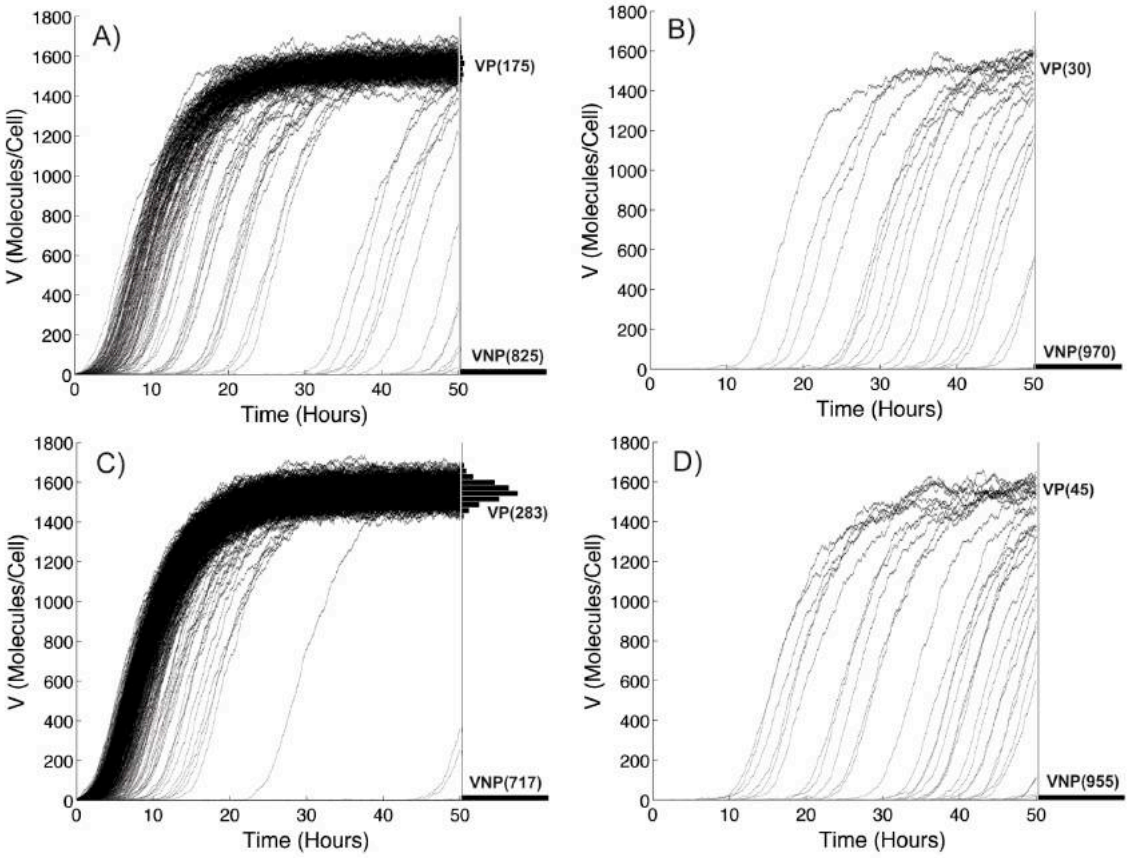


Figure 26. Stochastic simulation for 50 hours
Stochastic simulation results for 1,000 individual trajectories for a simulated timespan of 50 hours. Each trajectory represents a single-cell dynamics. On the right, a histogram of the final trajectories is given. VNP and VP legends are used to indicate those states. A) Using the initial conditions *V(0)=0, C(0)=0*. B) Using the initial conditions *V(0)*=0, *C(0)*=1000. C) Using the initial conditions *V(0)*=5, *C(0)*=0. D) Using the initial conditions *V(0)*=5, *C(0)*=1000. The representative example was made with the nominal parameter values given in Table 2.

To determine the temporary stability of the obtained stochastic dynamics, the SSA was carried out for large periods of time (21 days of simulated time) obtaining similar dynamics to those obtained over 50 hours. However, for the longer simulations, the stochastic transitions from trajectories in the VNP to the VP state were more evident (see Figure 27 A-D).

Finally, to determine the parameter space that reproduces bimodal dynamics, the SSA for all parameters generated by the Monte-Carlo scanning was implemented. Histograms for the trajectories at the end of the simulated time were obtained. In those histograms, it was determined whether bimodal distributions were present. Figure 28 shows the parameter space that reproduces bimodal dynamics. It should be noted that the region that reproduces bimodal dynamics is obtained with low nominal values of the parameter p_{vc}.

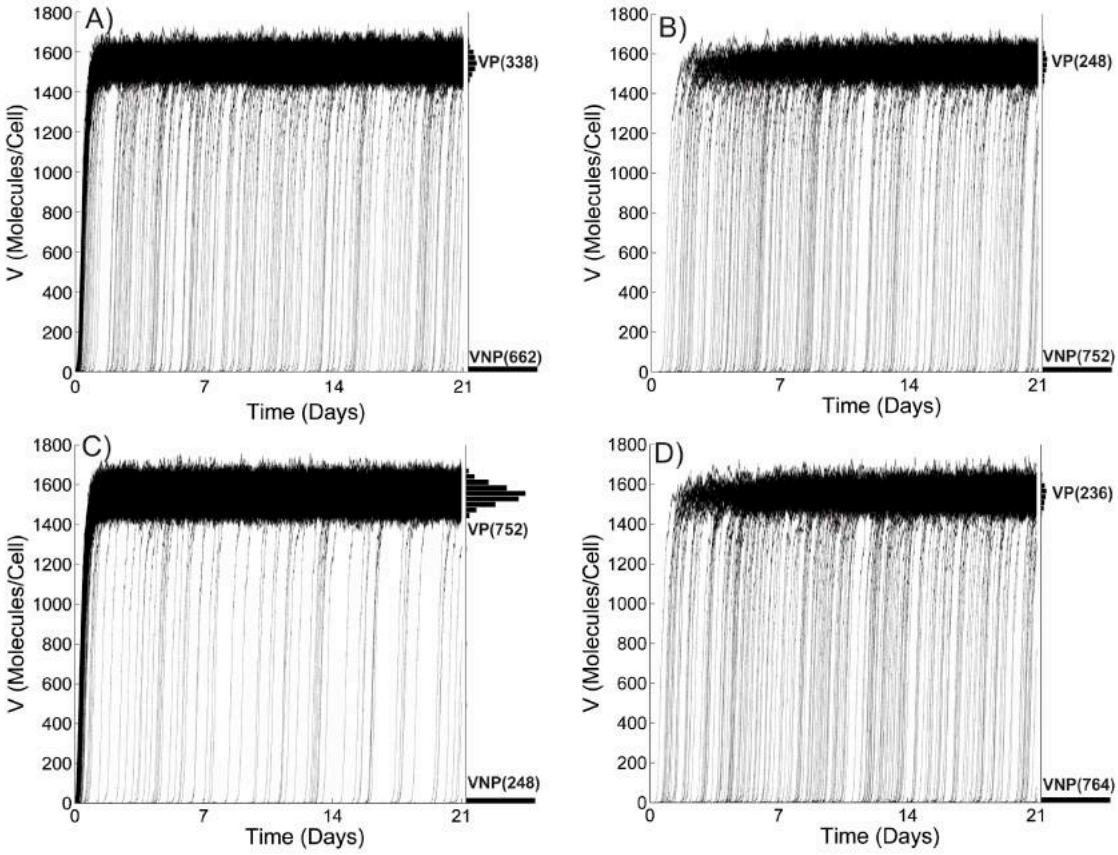


Figure 27. Stochastic simulations for 21 days
Stochastic simulation results for 1,000 individual trajectories for a simulated timespan of 21 days. Each trajectory represents a single-cell dynamics. On the right, a histogram of the final trajectories is given with the number of cells that end in each distribution. VNP and VP legends are used to indicate those states. A) Using the initial conditions $V(0)$=0, $C(0)$=0. B) Using the initial conditions $V(0)$=0, $C(0)$=1000. C) Using the initial conditions $V(0)$=5, $C(0)$=0. D) Using the initial conditions $V(0)$=5, $C(0)$=1000. The representative example was made with the nominal parameter values given in Table 2.

3.5 Discussion

In this model, bistability was obtained as the result of the combined effect of a double negative interaction between the virus and host biomolecules. Additionally, it was determined that multiple parameter combinations reproduce bistable dynamics (see Figure 21). These

results indicated that the virus and host interaction network are robust to reproduce bistable dynamics.

Considering the given initial conditions (section 3.1.1), only two possible outputs were obtained (see Figure 22); the first one was a system that began in the VNP and stayed there, and the other one was a system that started near the VNP state and evolved to the VP state. The intrinsic randomness in the system was needed to explain transitions between states.

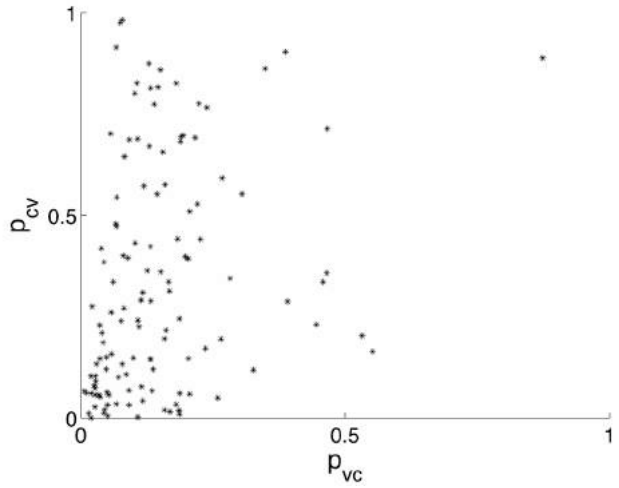


Figure 28. Parameter space that reproduces bimodal dynamics
Black dots represent bimodal cases. The parameters p_{vc} and p_{cv} correspond to the x- and y- axis, respectively. Every axis is normalized, and extreme values were eliminated. Simulations made using different initial conditions generated similar results (data not shown).

Then, it was determined how the biochemical noise affects the system dynamics across the parameter space. The VP state was favored when low nominal values of p_{vc} were given (see Figure 24, red dots). Subsequently, it was determined that, as the intensity in the biochemical noise increased, there was an increase in the probability of obtaining the VP state (see Figure 25). The results agree with previous studies that proved that biochemical noise is detrimental to the latency decision (Weinberger et al. 2008; Weinberger & Shenk 2007; Weinberger et al. 2005; Razooky & Weinberger 2011; Konkoli & Jesorka 2013).

Stochastic simulations reproduced temporal bimodal dynamics with transitions from the VNP to the VP state (see Figure 26). Comparing the results with the stochastic dynamics of the transcriptional Tat circuit, significant differences were identified. The stochastic dynamics of the Tat circuit reproduces non-bistable temporary bimodal dynamics were cells in the VP

state eventually return to the VNP state (Weinberger et al. 2005; Razooky & Weinberger 2011).

The conditions of the cell before infection contribute to the entry into the latent or productive phenotype. An initial molecular count of HIV proteins above zero at the beginning of stochastic simulation reduces the probability of obtaining cells in the latent state and enhances the stochastic transitions to the VP state (see Figure 26 B D). Previous reports considered an initial count of HIV proteins above zero to be required to obtain bimodal behavior (Althaus & De Boer 2010; Weinberger et al. 2005). However, if CRFs are already expressed in the system, the probability of obtaining a latent state increases (see Figure 26 B D). These results are corroborated by previous experimental studies in which the expression of CRF in infected cells increased the probability that those cells exhibited the latent phenotype (Huang et al. 2007; Wu 2012).

Stochastic dynamics for long periods of time (21 days) showed accumulative transitions from the VNP to the VP state (see Figure 27 A-D). Nevertheless, it was not possible to detect transitions from the VP to the VNP state. For this reason, it is expected that simulations could only achieve monomodal behavior. This predicted temporal bimodal behavior may explain the dynamics obtained for long-term infection under HAART characterized by a decline in HIV concentration and intermittent viral blips (Rong & Perelson 2009).

Comparing the results obtained by transition-state theory and the SSA, it was found that the strength of the interactions between innate immune response and HIV proteins affects the final VNP or VP state. Specifically, it was determined that low nominal values of p_{vc} and high nominal values of the parameter p_{cv} favor the VP state (Figure 7, red dots and Figure 28, black dots). The genetic background of each individual could determine p_{vc}. Polymorphisms in genes with innate antiviral function, for example CRF, may affect gene expression and protein function, thus contributing to HIV susceptibility and disease progression (Barr 2010). However, the parameter p_{cv} could be determined by the specific anti-restriction mechanism of the HIV protein, which could also be affected by the genetic background of the given viral strain (Eriksson et al. 2013).

4 Model II. Variability in Tat protein

4.1 Model aim

In this second part of the project it was hypothesized that: genotypic variability in Tat protein debilitate the transcriptional Tat circuit and influence the final cell fate. To test this hypothesis the following specified theoretical questions were formulated:

1) How does genotypic variability in Tat protein affects transactivation activity?
2) How does transactivation activity influence the cell fate?
3) How does transactivation activity modify the proportion of infected cells in latent and productive state?
4) How does latency affects the lifetime of a population of infected cells?

4.2 Model construction

4.2.1 Deterministic model

4.2.1.1 Deterministic model with time delay for the HIV gene expression

A mathematical model was developed as a tool to solve this hypothesis. The model was developed using existing biological information regarding the steps involved in the production of HIV proteins (V). The specific details of this circuit are given in Section 1.3.1.1 and in Figure 8. In the system it was included: a basal production of V, production of V enhanced by the transcriptional Tat circuit and the decay of V.

The deterministic mathematical model is given by the following delay differential equation (DDE) that accounts for the temporal evolution of the overall concentration of V in one HIV infected cell:

$$\frac{d}{dt}V = k_{BV} + k_V\left(V(t-\theta)/\left(V(t-\theta)+k_M\right)\right) - k_{dV}V, \tag{46}$$

The meaning of the functions, variables and parameters in the above equation is as follows: parameter k_{BV} represents the basal expression of V (Nabel & Baltimore 1987). The Michaelis Menten function denotes the transcriptional Tat circuit (Razooky & Weinberger 2011), where it was explicit considered a time delay (θ) to take into account the time that Tat protein

remains in the inactive state (Weinberger & Shenk 2007). The parameter k_V is the maximal V production rate, and the parameter k_M is a half-saturation constant. The parameter k_{dV} denotes a first-order decay rate of V (Tan et al. 2012). All the parameters in the equations have positive ranges defined by experimental reports and are tabulated in Table 5.

To test the variability in the transcriptional Tat circuit parameter k_V was modified in different ranges.

4.2.2 Stochastic model

To consider stochastic effects the determinist system equation (46) was converted into the reaction scheme given in Table 4. To obtain this reaction scheme it was used the terms of production and degradation in the system equation as elementary complex reactions.

Table 4. Reaction scheme for the intracellular dynamics in the viral infection

Reactions	Description	Effective propensity
$\varnothing \rightarrow V$	Basal production of V	k_{BV}
$V \rightarrow V+V$	Delayed Transactivation	$k_V\,(V / (V + k_M))$
$V \rightarrow \varnothing$	Decay of V	$k_{dV}\,V$

4.3 Methods

4.3.1 Model parameters

Table 5. Parameter values for the intracellular dynamics

Parameter	Description	Nominal Value	Range	Units	Reference
k_{BV}	Basal expression rate of V	2.19×10^{-6}	$[0, 1\times10^{-2}]$	Molecules/s	(Nabel & Baltimore 1987)
k_M	Michaelis-constant	844	$[1\times10^{2}, 1\times10^{4}]$	Molecules	(Slice et al. 1992)
$*k_V$	Maximal expression rate of V	9.31×10^{-1}	[0, 1]	Molecules/s	(Reddy & Jhon 1999)

k_{dV}	Decay rate of V	3.81×10^{-4}	$[0, 1\times10^{-3}]$	1/s	(Reddy & Jhon 1999)
θ	Time delay	600	[0,1200]	s	This study

- k_v value used to reproduce cells infected with viruses carrying the H13L mutation in Tat protein was defined as $k_{v\_mutant}=0.64* k_v$ (Donahue et al. 2012).

4.3.2 Experimental data

Experimental data describe a cell culture that considers the establishment and reactivation of HIV gene expression using $CD4^+$ Jurkat T cells infected with full length wild-type viral genomes (NL4-3-Δ*env*-EGFP) or attenuated *tat* viruses carrying the H13L mutation in Tat protein. The cultures were followed at various times postinfection, and the evolution of infected cells in the virus-productive and latent state were determined using flow cytometry (Donahue et al. 2012). Notice that in this experimental setup no secondary infections can be observed because of the produced virus lack env polyproteins, needed to infect new cells. Thus, experimental data describes the percentage of cells that remain infected since the beginning of the infection.

4.3.3 Theoretical design

Having a mathematical model and experimental data the following strategy was employed to test the hypothesis: First, parameter values were defined inside biological feasible limits and combinations of pseudorandom parameter sets were generated. Then, each parameter set was used to solve the model in deterministic and stochastic regimes (see Section 6.2.2, for the DSSA). Subsequently, it was obtained the parameter values for the model by minimizing an objective function that compares simulation results and experimental data for the *wild-type* infection (see Section 6.2.3). Then, it was validated the model by testing if the model was able to reproduce cells infected with viruses carrying the H13L mutation in Tat protein. Subsequently, it was determined robustness in the selected parameter values (see Section 6.2.4). Then, simulations were performed testing variability in Tat protein (see Section 6.2.5). A model that considers HIV intracellular dynamics coupled with cell population dynamics is given in Section 6.4. Finally, cultures of cells infected with *wild-type* and mutant viruses considering cell death were made (see Section 6.2.6). The complete theoretical design is given in the flow diagram given in Section 6.3.2.

4.3.4 Numerical methods

Function dde23 from Matlab was used to solve DDEs with constant delays. Symbolic methods in Matlab (The Mathworks Inc. 2010) were used to analytically solve the system. The DSSA was coded in Matlab. Simulations for cell populations considering cell death were coded in Matlab. COPASI 4.11 was used for data fitting and parameter estimation (Hoops et al. 2006).

4.4 Results

4.4.1 Deterministic dynamics

To determine the dynamical behavior of the system given in equation (46), it was analytically solved in the limit $(\theta = 0)$ obtaining for positive parameters a monostable circuit (see Section 6.1.2). Then, numerically solving equation (46) using different values for the time delay $(\theta \in [0,1200])$ it was determined that the system is only affected in the response time (elapsed time in which the system changes from 10% to 90% in the response), but monostability was maintained (see Figure 29).

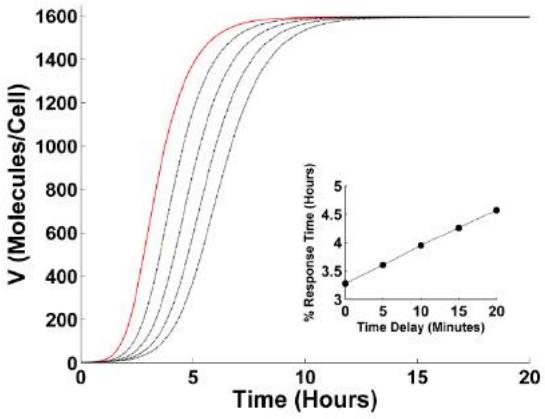


Figure 29. Delayed deterministic simulation
Considering no delay (red line) and assuming the following time delays: 5min, 10min, 15min and 20min (black lines, from left to right). The inset shows the response time vs different time delays.

4.4.2 Stochastic dynamics

Then, to determine the behavior of the system under stochastic dynamics delayed stochastic simulations were used to obtain trajectories that represent the concentration of V in one

infected cell. The simulation results showed bimodal dynamics where cells started in the initial condition (latent state, *V(0)=1 Molecules/Cell*) and as the time goes on, by stochastic transitions, evolve into the virus-productive state (see Figure 30). In the simulations it was assumed the virus-productive state if V reached values inside the biological limit determined in some HIV-biomolecules (i.e. $V \in [1000, 10000]$ Molecules/ Cell) (Kim & Yin 2005).

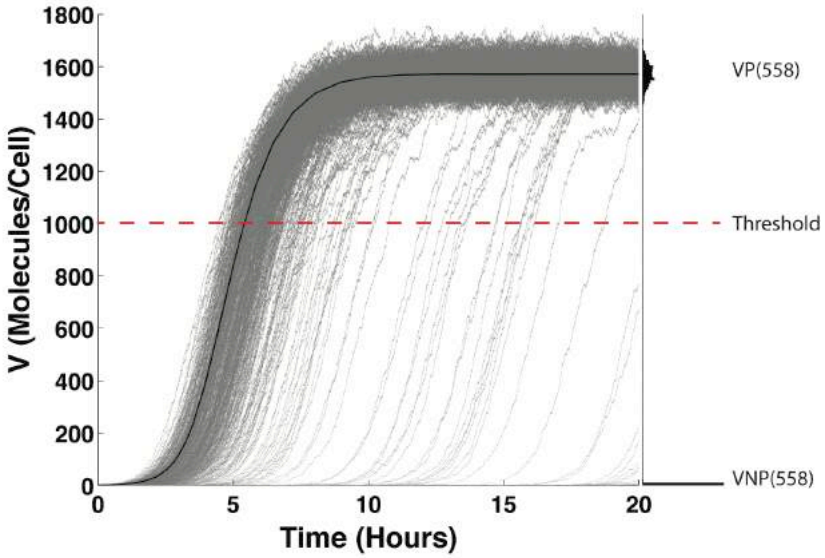


Figure 30. Stochastic simulation results for a delayed system
1000 individual trajectories that represent single cell dynamics are shown for a simulated time-span of 20 hours. At right a histogram of the final trajectories is given. Red doted line represents the threshold used to consider cells in the virus productive state.

To calibrate the model it was determined the parameter values that best reproduce experimental data corresponding to cells infected with *wild-type* viruses. The simulation results showed reasonable similarity with experimental data (see Figure 31, black dots and black lines). In this way, the nominal parameter values given in Table 5 are achieved.

Then, to validate the model and check the feasibility of selected nominal parameter values it was tested if the model reproduced cells infected with viruses carrying the H13L mutation in Tat protein. H13L mutation decreases transactivation activity by approximately 40% (Donahue et al. 2012). In the model simulations keeping the nominal values for all parameter and only decreasing the nominal value in parameter responsible for transactivation activity (k_V) were performed. Obtaining that reducing 36% k_V the model properly reproduces 23 days of experimental data (see Figure 31 red dots and red lines).

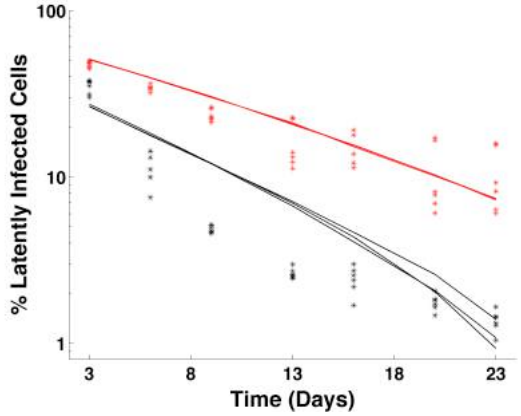


Figure 31. Calibration and validation of simulation results with experimental data
Data points for 23 days of experimental and simulation time. Black points represent experimental data for cells infected with *wild-type* viruses; this data was used for model calibration. Red points represent experimental data for cells infected with mutant viruses; this data was used for model validation. Continue black lines represent three independent simulation results for a population of 10,000 infected cells with *wild-type* viruses. Continue red lines represent three independent simulation results for a population of 10,000 of infected cells with mutant viruses.

Additionally, given the experimental manipulation we proved that a better fitting is achievable if cell division and cell dilution along with the cell death given by the cytopathic effects of virus replication are considered (see the fitting given in Appendix 6.4).

Further, to determine the robustness in the parameter values $1x10^7$ pseudorandom parameter sets were tested to show the reaction using a range of time delays $(\theta \in [0,1200])$. Here it was obtained that in the defined parameter space stable bimodal dynamics (maintained for more that 3 days of simulation time) only occurred under delayed reactions. Interestingly, it was observed that as it was increased the size of the time delay a larger percentage of parameters could reproduce bimodal dynamics (see Figure 32). The rest of the parameters did not show restriction in their values to reproduce bimodality.

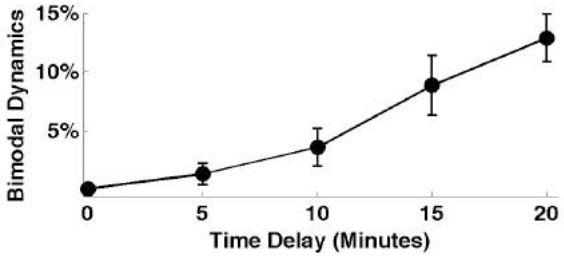


Figure 32. Robustness in the system parameters that reproduce bimodal dynamics
Percentage of obtained parameters with bimodal dynamics after stochastic simulations using the following time delays: 5min, 10min, 15min and 20min. Data taken at day 3 of the simulated time and using parameter values ranges given in Table 5. Bars represent the standard deviation of three independent runs.

To determine if the genotypic variability in Tat protein affects the decision between active replication and latency, it modified the strength in the transcriptional Tat circuit. By varying the value of parameter k_V in a given range it was obtained a major percentage of infected cells in the latent state (see Figure 33).

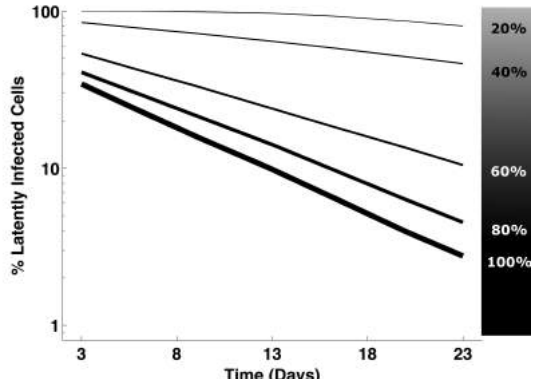


Figure 33. Varying the transactivation activity in a population of infected cells
Each line represents a population of HIV infected cells where the thickness in the line is proportional to the transactivation activity. Simulations were performed using the nominal parameter values given in Table 5 and varying from 20 to 100% parameter k_v. Black bar at the right represents Tat transactivation activity.

Finally, to study the effect of genotypic variability in Tat protein we mimicked a cell culture in a petri dish considering cell death. In those simulations, infected cells made stochastic transitions from latent into the virus-productive state and if they remained in the virus-productive state for more than two days they were removed from the simulation (Perelson AS, Essunger P, Cao Y, Vesanen M, Hurley A, Saksela K, Markowitz M 1997). Following the culture for 23 days it was determined that cells infected with *wild-type* viruses made transition to the virus-productive state in early periods of time and as a consequence of this early transition they died faster than those infected with the mutant viruses (see Figure 34).

4.5 Discussion

In this study the effects of genotypic variability in Tat protein in the dynamics of HIV disease were determined. Stochastic dynamics and delayed reactions in the transcriptional Tat circuit are indispensable to describe the bimodal response in HIV gene expression (Razooky & Weinberger 2011; Weinberger et al. 2005). It was proven that an equivalent ODE-model is monostable under any combination of positive parameter values (see Section 6.1.2 and Figure 29). Bimodal dynamics were obtained when some trajectories remained in the latent state and eventually transition into the virus-productive state as previously reported by (Weinberger & Shenk 2007).

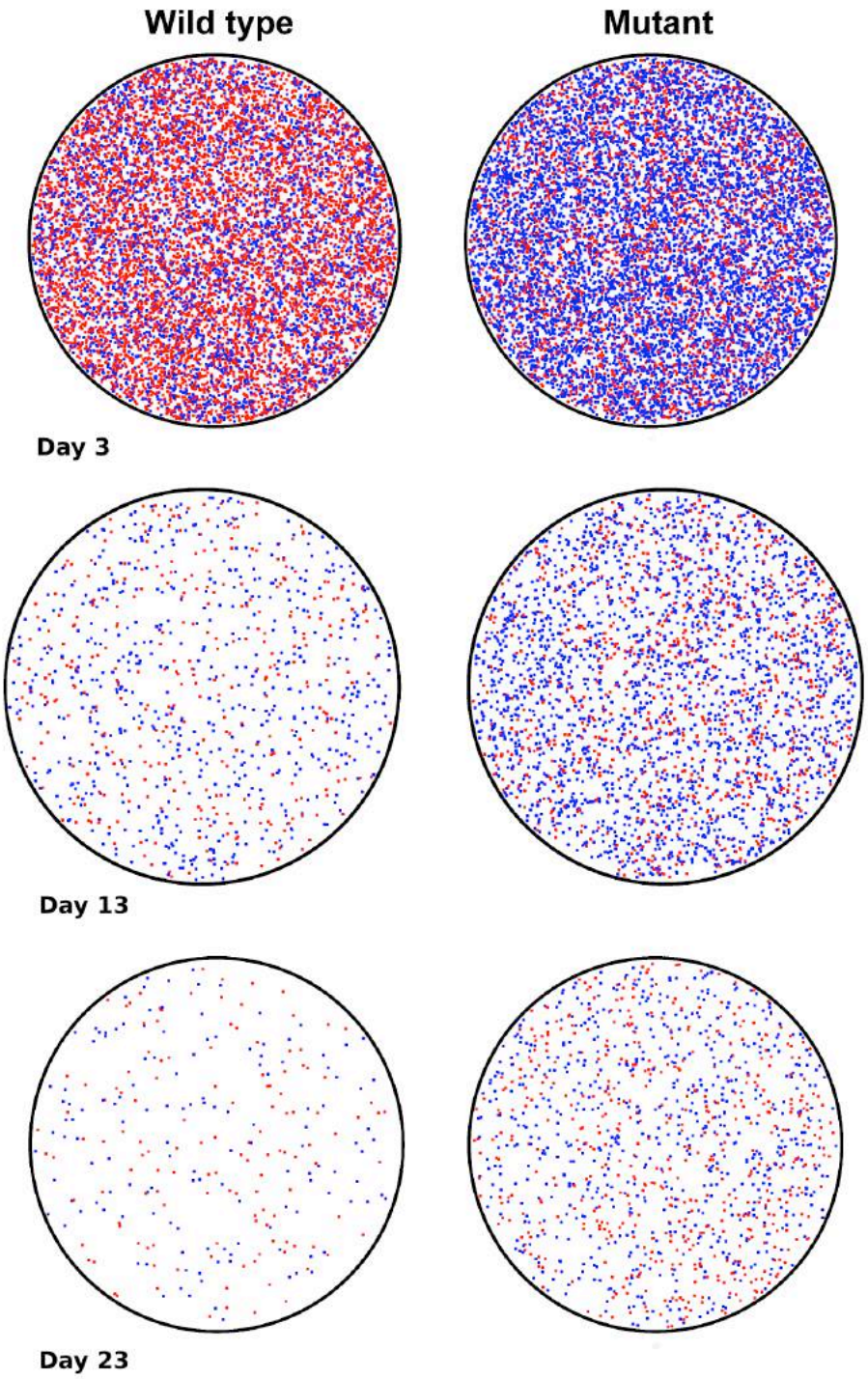


Figure 34. Simulation of cultures of infected cells
Representation of a culture of cells infected with *wild-type* viruses (left column, Wild type) and a culture of cells infected with viruses carrying the H13L mutation in the Tat protein (right column, Mutant). Red dots represent cells in virus-productive state. Blue dots represent cells in latent state. Each row contains two images that represent days 3, 13 and 23 of simulated time. Representative example using the nominal parameter values given in Table 5.

Even considering the simplicity of the mathematical model, we observed a reasonably good fit between the simulations and the experimental measurements of a cell culture infected with *wild-type* viruses (see Figure 31, black lines). Additionally, it was tested whether the model can also explain data that was not used in the model calibration. Following biological constrains, *i. e.* reducing 36% Tat protein activity the model was able to reproduce a culture of cells infected with attenuated viruses carrying the H13L mutation in Tat protein (see Figure 31, red lines).

Then, the importance of delayed stochastic dynamics in HIV latency was determined. It was tested using different delay sizes obtaining that delayed dynamics are a necessary condition to maintain sustained bimodal dynamics (see Figure 32). The biological mining of this finding is important, because it shows the dual role of the states in Tat protein. On one side, it can help the virus to enhance HIV production by orders of magnitude. On the other side, it can also determine the latent state that may allow the maintenance of the infection for long periods of time.

Frequently Tat protein mutations decrease Tat protein transactivation activity in respect of *wild-type* Tat protein (Li et al. 2012). Combining this information with the findings observed in this work it can be predicted that genotypic variability in the Tat protein generated during the course of HIV infection may attenuate Tat activity and decrease the probability that infected cells become into the virus-productive state (see Figure 33).

Analyzing cell populations it was determined that infected cells with *wild-type* viruses died faster than cells infected with mutant viruses (see Figure 34). This indicates that an inefficient blocking of Tat protein is equivalent to an inefficient transactivation activity. Those findings showed that an inefficient transactivation activity resulted in the maintenance of HIV latency and hence being prejudicial for the long-term prognosis of HIV infection.

5 Conclusions and perspectives

In this thesis, a network motif encoded by the mutual negative interaction between the virus and host biomolecules was demonstrated. This, mutual negative interaction generates bistable and bimodal dynamics that reproduces the latent and productive phenotypes in HIV infected cells.

Subsequently, different mechanisms were studied to determine their effect in the cell fate decision. First, biochemical noise was tested showing that it increases the probability that the infected cell adopts the virus productive state. Later, it was proven that the innate immune response during HIV infection increases the probability that the cell adopts the latent state.

Then, the molecular mechanisms responsible for the virus gene expression were studied. Here, it was proven that genotypic variability in the virus gene expression is reflected as inefficient viral transactivation and a decreased probability to HIV gene expression that at the end results in latency. Then, it was proven that cells infected in latent state affect the long-term prognosis of HIV infection.

Finally, the results obtained in this thesis increase the landscape in the understanding of latency. As a future work, it is suggested to modify the presented models, migrating from minimal models to models that consider the complete mechanisms in the virus and host pathways.

6 Appendices

6.1 Additional mathematical calculations

6.1.1 Transition between states

To determine the influence of the biochemical noise in the transition between states, it is considered a stable distribution of cell states in the presence of random fluctuations. This idea is reminiscent of statistical physics, where randomness results from thermal fluctuations and the stable distribution of states reflects a potential energy function (Sisan et al. 2012).

Following (Strogatz 1994), the right hand side of equation (44) was redefined as the potential-like function

$$U_V = -\int F(v)dv, \tag{47}$$

and, after integrating equation (47), a potential-like landscape emerges. In a potential-like landscape, the local minima correspond to stable cell states. Notice that in the project the stable cell states correspond to the VNP and VP states. In a landscape composed of multiple minima, transitions from one stable state to another are possible if there is a force that drives those fluctuations.

Then, using equation (44), standard transition-rate theory was used. Standard transition-rate theory states the escape rate is proportional to the exponential of the negative of the energy-barrier height (ΔU) separating this basin from the surrounding state, divided by the intensity of η. Therefore, the probability per unit time that the molecule state leaves the basin is proportional to the following expression: $exp(-\Delta U/\eta)$ (Kramers 1940).

According to the previous discussion, the cell escape rates from one local minimum to another are defined from the basin VNP into VP ($p_{\text{VNP-VP}}$) and from the basin VP into VNP ($p_{\text{VP-VNP}}$), equations (48) and (49), respectively.

$$p_{VNP-VP} = A_0 \exp^{-\left(\frac{\Delta U_{VNP}}{\eta}\right)}, \tag{48}$$

$$p_{VP-VNP} = A_0 \exp^{-\left(\frac{\Delta U_{VP}}{\eta}\right)}, \tag{49}$$

where A_0 is proportionality constants.

To extend the analysis from a single cell to study the dynamics of cell populations, it was defined the cellular flux that determines the number of cells that escape from the basin VNP into VP ($J_{\text{VNP-VP}}$) and *vice versa* ($J_{\text{VP-VNP}}$) as follows:

$$J_{VNP-VP} = N_{VNP}(p_{VNP-VP}), \tag{50}$$

$$J_{VP-VNP} = N_{VP}(p_{VP-VNP}), \tag{51}$$

N_{VNP} and N_{VP} are the number of cells whose states are trapped in the basins VNP and VP, respectively.

In the dynamic equilibrium, the cellular fluxes from the basin VNP into VP are equal

$$N_{sVNP}(p_{VNP-VP}),= N_{sVP}(p_{VP-VNP}), \tag{52}$$

where N_{sVNP} and N_{sVP} denote the equilibrium cell counts in the states VNP and VP, respectively.

Substituting the escape rates, the ratio of cells in the VNP and VP state was expressed as follows:

$$\frac{N_{sVNP}}{N_{sVP}} = \frac{p_{VP-VNP}}{p_{VNP-VP}} = \frac{\exp^{-\left(\frac{\Delta U_{VP}}{\eta}\right)}}{\exp^{-\left(\frac{\Delta U_{VNP}}{\eta}\right)}}. \tag{53}$$

Finally, the probability of obtaining cells in the VNP state was defined as

$$P(VNP) = N_{sVNP} / (N_{sVNP} + N_{sVP}), \tag{54}$$

and the probability of obtaining cells in the VP state was defined as:

$$P(VP) = 1 - P(VNP). \tag{55}$$

It is important to mention that similar calculations were made using the model given by equations (41) and (42). In those calculations the potential functions were calculated along the nullclines of the system and the procedure was followed until determine the probability of obtaining cells in the VP and VNP state. The results obtained with those calculations are equivalent to those obtained with the simplified version of the model.

6.1.2 Analytical solution of the deterministic model

Considering the limit $(\theta = 0)$ the system given in equation **(46)** can be rewritten in the form:

$$\frac{d}{dt}V = k_{BV} + k_V\left(V(t)/\left(V(t)+k_M\right)\right) - k_{dV}V(t). \tag{56}$$

Redefining the parameters in the dimensionless form: $a_{bv} = k_{BV}/k_{dV}k_M$ and $a_v = k_V/k_{dV}k_M$, and redefining the variables in the dimensionless form: $v = V/k_M$ and $\vartheta = t \cdot k_{dV}$. The following dimensionless model is obtained:

$$\frac{d}{d\vartheta}v = a_{bv} + a_v\left(v(\vartheta)/\left(v(\vartheta)+1\right)\right) - v(\vartheta). \tag{57}$$

Defining $dv/d\vartheta = 0$ and solving for *v*, two possible solutions are obtained:

$$v_1^* = 1/2\left(a_{bv} + a_v - 1 + \sqrt{4a_{bv} + \left(a_{bv} + a_v - 1\right)^2}\right), \tag{58}$$

$$v_2^* = 1/2\left(a_{bv} + a_v - 1 - \sqrt{4a_{bv} + \left(a_{bv} + a_v - 1\right)^2}\right). \tag{59}$$

Defining positive values for the parameters: $(a_v > 0, a_{bv} > 0)$, and using symbolic methods in Matlab it can be proved that: $v_1^* > 0$ and $v_2^* < 0$.

So, v_1^* is the only positive and biological significant solution. In other words, for all positive parameter values the system is monostable.

6.2 Pseudocodes

6.2.1 The SSA

This algorithm was implemented as suggested by (Gillespie 1976).

Input: stoichiometry; reaction rates; initial state;

Definitions: T is the simulation time; M chemical reactions; $X(t)$ is the vector of chemical species; $a_\mu(X)$ is a propensity function of reaction μ; τ is the time step for the occurrence of the next reaction; φ is a stoichiometric vector.

Output: state dynamics.

Begin

While $t < T$ do

Generate u_1 and u_2 as $u(0,1)$ random variables

$$a_0 = \sum_{\mu=1}^{M} a_\mu\left(X(t)\right), \tag{60}$$

$$\tau = \left(1/a_0\left(X(t)\right)\right)\ln\left(1/u_1\right). \tag{61}$$

Select the reaction μ such that

$$\sum_{k=1}^{\mu-1} a_i\left(X(t)\right) < u_2 a_0\left(X(t)\right) \le \sum_{k=1}^{\mu} a_i\left(X(t)\right). \tag{62}$$

Record time as:

$$t = t + \tau, \tag{63}$$

and update the system count as:

$$X(t+\tau) = X(t) + \varphi_{\mu i}. \tag{64}$$

End

6.2.2 The DSSA

This algorithm was implemented as suggested by (Barrio et al. 2006).

Input: stoichiometry; reaction rates; initial state;

Definitions: T is the simulation time; θ is the delay size; M chemical reactions; $X(t)$ is the vector of chemical species; $a_\mu(X)$ is a propensity function of reaction μ; τ is the time step for the occurrence of the next reaction; φ is a stoichiometric vector.

Output: state dynamics.

Begin

While $t < T$ do

Generate u_1 and u_2 as $u(0,1)$ random variables

$$a_0 = \sum_{\mu=1}^{M} a_\mu\left(X(t)\right), \tag{65}$$

$$\tau = \left(1/a_0\left(X(t)\right)\right)\ln\left(1/u_1\right). \tag{66}$$

Select j such that

$$\sum_{k=1}^{\mu-1} a_k\left(X(t)\right) < u_2 a_0\left(X(t)\right) \le \sum_{k=1}^{\mu} a_k\left(X(t)\right). \tag{67}$$

If delayed reactions are scheduled within $(t, t+\tau]$ then let k be the delayed reaction scheduled next at time $t+\theta$

$$X(t+\theta) = X(t) + \varphi_k, \tag{68}$$

$$t = t + \theta. \tag{69}$$

Else

If μ is not a delayed reaction then

$$X(t+\tau) = X(t) + \varphi_\mu. \tag{70}$$

Else

Record time, $t+\tau+\theta$, for delayed reaction

$$t = t + \theta. \tag{71}$$

- *End*

6.2.3 Algorithm for the calibration of the model

Generate a matrix with the number of combination of random parameter sets to test

Load this matrix of parameters

For i=1:Nd (where N_d is the total data set)

Run the DSSA for 23 days.

Compare simulation results with experimental data using the objective function:

$$In\mathrm{L}(N_d) = \sum_{i=1}^{n}\left(\left(ed_i - s_i(N_d)\right)^2 / \sigma_i^2\right), \tag{72}$$

where e*d* is the experimental data in a given time, and *s* is the simulation data in a given time, *n* is the total number of data points, σ^2 is the variance in the data in a given time.

Save vector $L(N_d)$

End

Choose min $L(N_d)$

Save parameters values associated to *min* $L(N_d)$.

6.2.4 Algorithm used to determine robustness in the parameter values

Generate a matrix with the number of combination of random parameter sets to test

Load this matrix of parameters

For j=1:D (where D=[0s, 300s, 600s, 900s, 1200s], used time delays)

For i=1:N_d (where N_d is the total data set)

Run the DSSA for 3 days.

Determine if bimodal dynamics are obtained at the end of the simulation

Save vector *Bim(N_d) containing parameters with bimodal dynamics*

End

End

Determine the percentage of parameters with bimodal dynamics.

6.2.5 Algorithm used to test genotypic variability in Tat protein

Load parameters values associated to *min* $L(N_d)$, generated in Section 6.2.3.

Begin

Generate the vector kv_s=[k_V*0.2, k_V*0.4, k_V*0.6, k_V*0.8, k_V*1]

For i=1: N_2 (where N_2=5, the length of kv_s)

Run the DSSA for 23 days, using for each iteration an element of vector *kv_s* as the value of parameter k_V.

Save percentage of infected cells during the simulated time in vector *G(i)*

End

6.2.6 Algorithm for cell population dynamics

Load parameters values associated with the *minimum value in* $L(N_d)$, generated in Section 6.2.3.

Use the following four assumptions:

1) M_0 initial cells are given at t_0.

2) At t_0 M_0 cells are infected and in latent state.

3) Cell division is not considered in the system.

4) Virus-productive cells die after two days of infection.

Begin

Generate a circular area of radius *r*.

Generate M_0 random points inside radius r.

Associate each point to a trajectory in the stochastic simulation.

While $t < T$ do

Run the DSSA for 23 days.

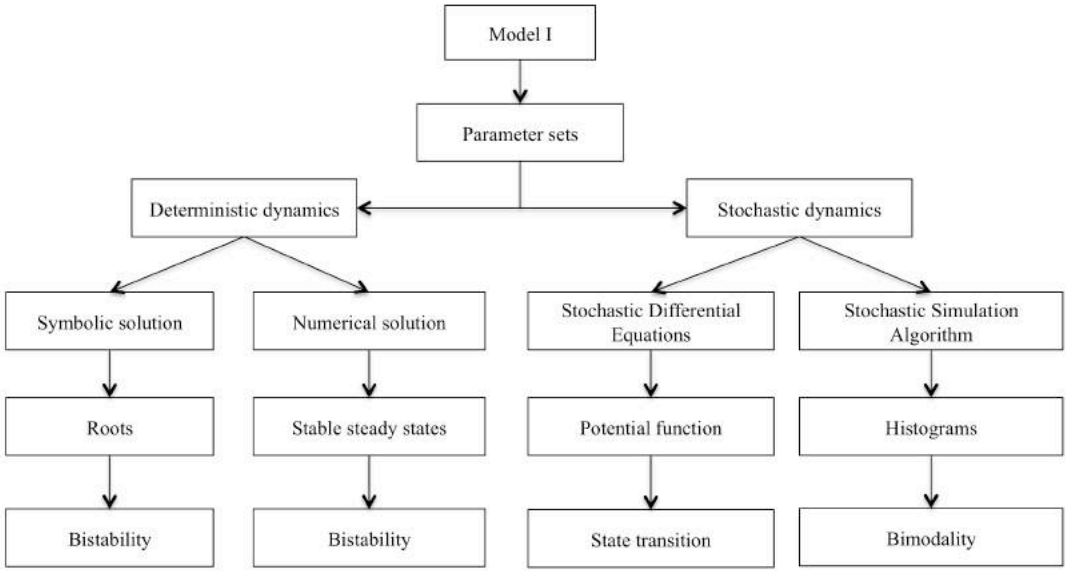
When *V>1000 Molecules/Cell*, infected cells in latent state become into virus-productive.

Remove form simulation virus-productive cells after two days of infection.

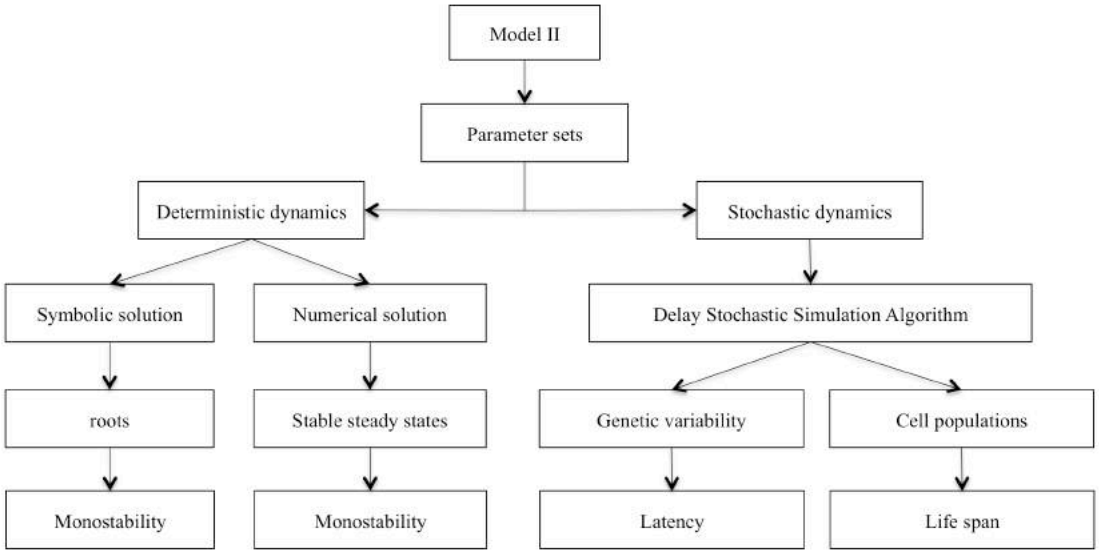
End

6.3 Flow diagrams

6.3.1 Flow diagram for the model of the virus and host interactions



6.3.2 Flow diagram for the model of the variability in Tat protein



6.4 Experimental data fitting

Experimental data was kindly provided by Aaron D. Donahue from Luis Pasteur Institute, France. This data describes time measurements of cultures of lymphocytes infected with HIV (Donahue et al. 2012). The total number of infected cells (TIC) in the culture is the sum of two cell subpopulations: infected cells in the virus-productive state (*PS*) and infected cells in the latent state (*LS*),

$$TIC = PS + LS. \tag{S1}$$

According to the experimental setup three forces affected TIC. The first one is cell replication, the second one is cell death given by cytopathic effects of the virus replication (lytic effect), and the last one is given by the dilution in the culture medium (this is the result of replacing old culture medium with fresh medium). The equation that describes the changes in the cell culture size is given by:

$$\frac{\Delta TIC}{\Delta t} = replication - lytic - dilution \tag{S2}$$

Given the experimental manipulation during the first 23 days, cell replication and cell dilution have approximately the same effect in the cell culture. For this reason, in this part of the data is reasonable to assume an initial fixed number of infected cells and associate the observed changes in TIC to the lytic effect. Assuming neither replication nor dilution and normalizing to the initial TIC gave us reasonable agreement between experimental data and simulation results (see Figure 31). Nevertheless, even after consider a large number of possible parameters sets it was not possible to improve the parameter fit.

To consider the additional forces in the evolution of the cell culture and improve the fit between simulation results and experimental data we developed a model that considers HIV intracellular dynamics coupled with cell population dynamics. Here we used the model given in equation (46) to describe HIV intracellular dynamics and we modeled the cell population dynamics using the following two differential equations:

$$\frac{dL}{dt} = k_L - k_P LV, \tag{S3}$$

$$\frac{dP}{dt} = k_P LV - k_{dP} P, \tag{S4}$$

that account for the temporal evolution of the infected cells in latent state (L) and for the temporal evolution of infected cells in a virus-productive state (P). The meaning of the

parameters is as follows: parameter k_L is the replication rate of L, k_{dP} is the transition rate of L into P and parameter k_{dP} considers lytic and dilution effects and represents the decay rate of P.

Maintaining the same nominal values given in Table 5 for the intracellular dynamics, keeping the limit $\tau = 0$ and using COPASI for estimating the rest of the parameters we properly reproduce the experimental data. See Figure 35.

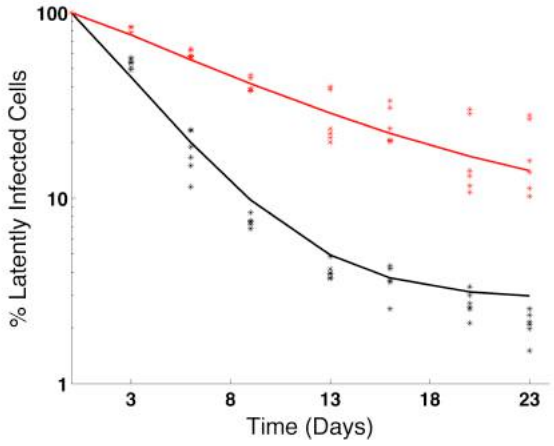


Figure 35 Cell Population Fitting.

Mathematical model that describes HIV intracellular dynamics coupled with cell population dynamics using the following initial conditions: *V(0)=1 Molecules/Cell, L(0)=* 1×10^6 Cells, *P(0)=* 0 Cells and using the parameter values for the cell population dynamics (k_L =0.1 Cells/s, k_P =5.2 $\times10^{-9}$ Cells/s*Molecules, k_{dP} = 5.2 $\times10^{-6}$ 1/s). Data points for 23 days of experimental and simulation time. Black points represent experimental data for cells infected with *wild-type* viruses. Red points represent experimental data for cells infected with mutant viruses. Black line represents simulation results for infected cells with *wild-type* viruses. Red line represents simulation results for infected cells with mutant viruses.

Considering a system that models cell population dynamics coupled with viral intracellular dynamics do have an effect in the system. This effect determines more complex evolution of the cell population that improves fitting between experimental data and simulations. However, a reasonable good fitting was obtained neglecting the effects of cell replication and dilution. In the main document we pursuit such simple intracellular model because here it is where our scientific question resides.

7 References

Adrews, S.S., Dinh, T. & Arkin, A.P., 2009. Stochastic Models of Biological Processes. *Encyclopedia of Complexity and Systems Science*, pp.8730–8749.

Althaus, C.L. & De Boer, R.J., 2010. Intracellular transactivation of HIV can account for the decelerating decay of virus load during drug therapy. *Molecular Systems Biology*, 6(348), p.348.

Artyomov, M.N. et al., 2007. Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities. *Proc. Nat. Acad. Sci. USA*, 104(48), pp.18958–18963.

Barr, S.D., 2010. Cellular HIV-1 restriction factors: a new avenue for AIDS therapy? *Future Virology*, 5, pp.417–433.

Barr, S.D., Smiley, J.R. & Bushman, F.D., 2008. The interferon response inhibits HIV particle production by induction of TRIM22. *PLoS pathogens*, 4(2), p.e1000007.

Barrio, M. et al., 2006. Oscillatory regulation of Hes1: Discrete stochastic delay modelling and simulation. *PLoS computational biology*, 2(9), p.e117.

Bohan, C. A., F. Kashanchi, B. Ensoli, L. Buonaguro, K. A. BorisLawrie, and J.N.B., 1992. Analysis of Tat transactivation of human immunodeficiency virus transcription in vitro. *Gene Expr.*, 2, pp.391–407.

Burnett, J.C. et al., 2009. Control of stochastic gene expression by host factors at the HIV promoter. *PLoS pathogens*, 5(1), p.e1000260.

Carter, J.B. & Saunders, V.A., 2007. Virology: principles and applications. *John Wiley and Sons, 2007*.

Chomont, N. et al., 2009. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature medicine*, 15(8), pp.893–900.

Choudhary, S.K. & Margolis, D.M., 2011. Curing HIV: Pharmacologic approaches to target HIV-1 latency. *Annual Review of Pharmacology and Toxicology*, 51, pp.397–418.

Chun, T.-W. et al., 1997. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*, 387(6629), pp.183–188.

Coiras, M. et al., 2009. Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. *Nature reviews. Microbiology*, 7(11), pp.798–812.

Cojo, M.S. et al., 2011. Mechanisms of RNA Interference in the HIV-1-Host Cell Interplay. *AIDS Rev*, 13, pp.149–160.

Craciun, G., Tang, Y. & Feinberg, M., 2006. Understanding bistability in complex enzyme-driven reaction networks. *Proceedings of the National Academy of Sciences*, 103(23), pp.8697–8702.

Davey Jr., R.T. et al., 1999. HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc Natl Acad Sci U S A*, 96(26), pp.15109–15114.

Van Dijk, D. et al., 2010. Identifying potential survival strategies of HIV-1 through virus-host protein interaction networks. *BMC Syst Biol*, 4, p.96.

Dimmock, N.J., Andrew, J.E. & Leppard, K., 2007. Introduction to Modern Virology. *John Wiley & Sons, 2007*, 6.

Donahue, D.A. et al., 2012. The viral protein Tat can inhibit the establishment of HIV-1 latency. *Journal of virology*, 86(6), pp.3253–63.

Donahue, D.A. & Wainberg, M.A., 2013. Cellular and molecular mechanisms involved in the establishment of HIV-1 latency. *Retrovirology*, 10(1), p.11.

Duggal, N.K. & Emerman, M., 2012. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nature reviews. Immunology*, 12(10), pp.687–95.

Eriksson, S. et al., 2013. Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. *PLoS pathogens*, 9(2), p.e1003174.

Espert, L. et al., 2005. Interferon-induced exonuclease ISG20 exhibits an antiviral activity against human immunodeficiency virus type 1. *The Journal of general virology*, 86(Pt 8), pp.2221–9.

Finzi, D. et al., 1997. Identification of a Reservoir for HIV-1 in Patients on Highly Active Antiretroviral Therapy. *Science*, 278(5341), pp.1295–1300.

Finzi, D. et al., 1999. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med*, 5(5), pp.512–517.

Ganesh, L. et al., 2003. The gene product Murr1 restricts HIV-1 replication in resting CD4 1 lymphocytes. *Nature*, 426, pp.853–857.

Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4), pp.403–434.

Gillespie, D.T., 2000. Chemical Langevin equation. *Journal of Chemical Physics*, 113(2000), pp.297–306.

Gillespie, D.T., 2007. Stochastic simulation of chemical kinetics. *Annual review of physical chemistry*, 58, pp.35–55.

Haaland, R.E., Yu, W. & Rice, A.P., 2005. Identification of LKLF-regulated genes in quiescent CD4+ T lymphocytes. *Mol Immunol*, 42(5), pp.627–641.

Havlir, D. V et al., 2005. Predictors of residual viremia in HIV-infected patients successfully treated with efavirenz and lamivudine plus either tenofovir or stavudine. *J Infect Dis*, 191(7), pp.1164–1168.

Ho, Y.C. et al., 2013. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell*, 155(3), pp.540–551.

Honda, K. et al., 2005. Regulation of the type I IFN induction: a current view. *International immunology*, 17(11), pp.1367–78.

Honda, K. & Taniguchi, T., 2006. IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nature reviews. Immunology*, 6(9), pp.644–58.

Hoops, S. et al., 2006. COPASI--a COmplex PAthway SImulator. *Bioinformatics (Oxford, England)*, 22(24), pp.3067–74.

Horner, S.M. & Gale, M., 2013. Regulation of hepatic innate immunity by hepatitis C virus. *Nature medicine*, 19(7), pp.879–88.

Hsu, S., 1981. On a Resource Based Ecological Competition Model with Interference. *Journal of mathematical biology*, 52, pp.45–52.

Huang, J. et al., 2007. Cellular microRNAs contribute to HIV-1 latency in resting primary CD4+ T lymphocytes. *Nature medicine*, 13(10), pp.1241–7.

Ivashkiv, L.B. & Donlin, L.T., 2014. Regulation of type I interferon responses. *Nature reviews. Immunology*, 14(1), pp.36–49.

Keizer, J., 1987. *Statistical Thermodynamics of Nonequilibrium Processes* 1st ed., Springer.

Kim, H. & Yin, J., 2005. Robust Growth of Human Immunodeficiency Virus Type 1 (HIV-1). *Biophysical Journal*, 89(4), pp.2210–2221.

Klipp, E. et al., 2011. *Systems Biology*, Wiley-Blackwell.

Knipe, D.M. et al., 2007. *Fields' Virology*, Wolters Kluwer Health/Lippincott Williams & Wilkins.

Konkoli, Z. & Jesorka, A., 2013. Fluctuations in Tat copy number when it counts the most: a possible mechanism to battle the HIV latency. *Theoretical biology & medical modelling*, 10, p.16.

Kramers, H.A., 1940. Brownian motion in a field of force and the diffusion model. *Physica*, (4), pp.284–304.

Laidler, K.J., 1987. *Chemical Kinetics*, the University of Michigan: Harper & Row.

Li, L. et al., 2012. Impact of Tat Genetic Variation on HIV-1 Disease. *Advances in virology*, 2012(Figure 1), p.123605.

Madigan, M.T. et al., 2010. Brock Biology of Microorganisms . *Benjamin Cummings*, 13.

Maitra, R.K. & Silverman, R.H., 1998. Regulation of Human Immunodeficiency Virus Regulation of Human Immunodeficiency Virus Replication by 2J , 5J-Oligoadenylate-Dependent RNase L. *Journal of virology*, 72(2).

Maldarelli, F. et al., 2007. ART suppresses plasma HIV-1 RNA to a stable set point predicted by pretherapy viremia. *Plos Pathogens*, 3(4), p.e46.

Marcello, A., 2006. Latency: the hidden HIV-1 challenge. *Retrovirology*, 3, p.7.

Markevich, N.I., Hoek, J.B. & Kholodenko, B.N., 2004. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *The Journal of cell biology*, 164(3), pp.353–9.

Van Marle, G. & Power, C., 2005. Human immunodeficiency virus type 1 genetic diversity in the nervous system: evolutionary epiphenomenon or disease determinant? *Journal of neurovirology*, 11(2), pp.107–28.

Mayne, M. et al., 1998. HIV-1 tat molecular diversity and induction of TNF-alpha: implications for HIV-induced neurological disease. *Neuroimmunomodulation*, 5(3-4), pp.184–92.

Mcnaught, A.D. & Wilkinson, A., 1997. *IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book")*. 2nd ed. O. Blackwell Scientific Publications, ed., WileyBlackwell; 2nd Revised edition edition.

Murray, J.D., 2002. *Mathematical Biology: I. An Introduction*, Springer.

Nabel, G. & Baltimore, D., 1987. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature*, 326(711-713).

Ning, S., Huye, L.E. & Pagano, J.S., 2005. Regulation of the transcriptional activity of the IRF7 promoter by a pathway independent of interferon signaling. *The Journal of biological chemistry*, 280(13), pp.12262–70.

O'Connell, K.A., Bailey, J.R. & Blankson, J.N., 2009. Elucidating the elite: mechanisms of control in HIV-1 infection. *Trends Pharmacol Sci*, 30(12), pp.631–637.

Okumura, A. et al., 2006. Innate antiviral response targets HIV-1 release by the induction of ubiquitin-like protein ISG15. *Proc. Nat. Acad. Sci. USA*, 103(5), pp.1440–5.

Pahle, J., 2009. Biochemical simulations: Stochastic, approximate stochastic and hybrid approaches. *Briefings in Bioinformatics*, 10(1), pp.53–64.

Palmer, S. et al., 2008. Low-level viremia persists for at least 7 years in patients on suppressive antiretroviral therapy. *Proc Natl Acad Sci U S A*, 105(10), pp.3879–3884.

Palmer, S., Josefsson, L. & Coffin, J.M., 2011. HIV reservoirs and the possibility of a cure for HIV infection. *J Intern Med*, 270(6), pp.550–560.

Perelson, A.S. et al., 1997. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*, 387(6629), pp.188–191.

Perelson AS, Essunger P, Cao Y, Vesanen M, Hurley A, Saksela K, Markowitz M, H.D., 1997. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature.*, 387(6629), pp.188–91.

Perez-Caballero, D. et al., 2009. Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell*, 139(3), pp.499–511.

Phillips, R. et al., 2010. Physical Biology of the Cell. *American Journal of Physics*, 78(11), p.1230.

Qian, H. & Bishop, L.M., 2010. The chemical master equation approach to nonequilibrium steady-state of open biochemical systems: linear single-molecule enzyme kinetics and nonlinear biochemical reaction networks. *International journal of molecular sciences*, 11(9), pp.3472–500.

Rand, U. et al., 2012. Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Molecular systems biology*, 8(584), p.584.

Razooky, B.S. & Weinberger, L.S., 2011. Mapping the architecture of the HIV-1 Tat circuit: A decision-making circuit that lacks bistability and exploits stochastic noise. *Methods (San Diego, Calif.)*, 53(1), pp.68–77.

Reddy, B. & Jhon, Y., 1999. Quantitative Intracellular Kinetics of HIV Type 1. *Aids Research and Human Retroviruses*, 15(3), pp.273–283.

Rong, L. & Perelson, A.S., 2009. Asymmetric division of activated latently infected cells may explain the decay kinetics of the HIV-1 latent reservoir and intermittent viral blips. *Mathematical biosciences*, 217(1), pp.77–87.

Sadler, A.J. et al., 2009. An antiviral response directed by PKR phosphorylation of the RNA helicase A. *PLoS pathogens*, 5(2), p.e1000311.

Sanghvi, V.R. & Steel, L.F., 2012. RNA silencing as a cellular defense against HIV-1 infection: progress and issues. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 26(10), pp.3937–45.

Di Santo, J.P., 2001. Lung Krupple-like factor: a quintessential player in T cell quiescence. *Nat Immunol*, 2(8), pp.667–668.

Schröder, H.C. et al., 1990. Binding of Tat protein to TAR region of human immunodeficiency virus type 1 blocks TAR-mediated activation of (2′-5′) oligoadenylate synthetase. *AIDS Research and Human retroviruses*, 6(5), pp.659–672.

Schwanhäusser, B. et al., 2011. Global quantification of mammalian gene expression control. *Nature*, 473(7347), pp.337–42.

Sisan, D.R. et al., 2012. Predicting rates of cell state change caused by stochastic fluctuations using a data-driven landscape model. *Proc. Nat. Acad. Sci. USA*.

Slice, L.W. et al., 1992. Characterization of recombinant HIV-1 Tat and its interaction with TAR RNA. *Biochemistry*, 31(48), pp.12062–12068.

Strebel, K., Luban, J. & Jeang, K.T., 2009. Human cellular restriction factors that target HIV-1 replication. *Bmc Medicine*, 7.

Strogatz, S.H., 2001. *Nonlinear dynamics and chaos: with applications to physics, biology and chemistry*, Perseus publishing.

Strogatz, S.H., 1994. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, Perseus Books Group.

Takeuchi, O. & Akira, S., 2009. Innate immunity to virus infection. *Immunological reviews*, 227(1), pp.75–86.

Tan, J. et al., 2012. Modeling and dynamical analysis of virus-triggered innate immune signaling pathways. *PloS one*, 7(10), p.e48114.

The Mathworks Inc., 2010. MATLAB Version 7.10.0 (R2010a).

Ullah, M. & Wolkenhauer, O., 2011. *Stochastic approaches for systems biology*, Springer.

Vajpayee, M. & Mohan, T., 2011. Current practices in laboratory monitoring of HIV infection. , (December), pp.801–822.

Wasserstein-Robbins, F., 2010. A mathematical model of HIV infection: Simulating T4, T8, macrophages, antibody, and virus via specific anti-HIV response in the presence of adaptation and tropism. *Bull Math Biol*, 72(5), pp.1208–1253.

Weinberger, A.D. & Weinberger, L.S., 2013. Stochastic fate selection in HIV-infected patients. *Cell*, 155(3), pp.497–9.

Weinberger, L.S. et al., 2005. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell*, 122(2), pp.169–82.

Weinberger, L.S., Dar, R.D. & Simpson, M.L., 2008. Transient-mediated fate determination in a transcriptional circuit of HIV. *Nature genetics*, 40(4), pp.466–70.

Weinberger, L.S. & Shenk, T., 2007. An HIV feedback resistor: auto-regulatory circuit deactivator and noise buffer. *PLoS biology*, 5(1), p.e9.

Wu, L., 2012. SAMHD1: a new contributor to HIV-1 restriction in resting CD4+ T-cells. *Retrovirology*, 9, p.88.

Wu, Y., 2004. HIV-1 gene expression: lessons from provirus and non-integrated DNA. *Retrovirology*, 1, p.13.

Zhou, H.-X., 2010. Rate theories for biologists. *Quarterly reviews of biophysics*, 43(2), pp.219–93.

Zhu, T. et al., 1993. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science (New York, N.Y.)*, 261(5125), pp.1179–81.