Article

Markov Chain-Based Stochastic Modelling of HIV-1 Life Cycle in a CD4 T Cell

Igor Sazonov [1,†] , Dmitry Grebennikov [2,3,4,†] , Andreas Meyerhans [5,6,†] and Gennady Bocharov [2,3,7,*,†]

1   College of Engineering, Swansea University, Bay Campus, Fabian Way SA1 8EN, UK; i.sazonov@swansea.ac.uk
2   Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences (INM RAS), 119333 Moscow, Russia; dmitry.ew@gmail.com
3   Moscow Center for Fundamental and Applied Mathematics at INM RAS, 119333 Moscow, Russia
4   World-Class Research Center "Digital Biodesign and Personalized Healthcare", Sechenov First Moscow State Medical University, 119991 Moscow, Russia
5   ICREA, Pg. Lluis Companys 23, 08010 Barcelona, Spain; Andreas.Meyerhans@upf.edu
6   Infection Biology Laboratory, Universitat Pompeu Fabra, 08003 Barcelona, Spain
7   Institute of Computer Science and Mathematical Modelling, Sechenov First Moscow State Medical University, 119991 Moscow, Russia
*   Correspondence: g.bocharov@inm.ras.ru
†   These authors contributed equally to this work.

Abstract: Replication of Human Immunodeficiency Virus type 1 (HIV) in infected $CD4^+$ T cells represents a key driver of HIV infection. The HIV life cycle is characterised by the heterogeneity of infected cells with respect to multiplicity of infection and the variability in viral progeny. This heterogeneity can result from the phenotypic diversity of infected cells as well as from random effects and fluctuations in the kinetics of biochemical reactions underlying the virus replication cycle. To quantify the contribution of stochastic effects to the variability of HIV life cycle kinetics, we propose a high-resolution mathematical model formulated as a Markov chain jump process. The model is applied to generate the statistical characteristics of the (i) cell infection multiplicity, (ii) cooperative nature of viral replication, and (iii) variability in virus secretion by phenotypically identical cells. We show that the infection with a fixed number of viruses per $CD4^+$ T cell leads to some heterogeneity of infected cells with respect to the number of integrated proviral genomes. The bottleneck factors in the virus production are identified, including the Gag-Pol proteins. Sensitivity analysis enables ranking of the model parameters with respect to the strength of their impact on the size of viral progeny. The first three globally influential parameters are the transport of genomic mRNA to membrane, the tolerance of transcription activation to Tat-mediated regulation, and the degradation of free and mature virions. These can be considered as potential therapeutical targets.

Keywords: HIV life cycle; mathematical model; stochastic processes; Markov chain; heterogeneity; sensitivity analysis

1. Introduction

Infection with the Human Immunodeficiency Virus type-1 (HIV) remains a global problem of public health concern, with more than 70 million people infected since the early 1980s [1]. Unfortunately, there are no efficient vaccines against HIV [2]. The within-host infection dynamics is characterized by (i) a variability of the disease course in infected individuals [3], (ii) the ability of the virus to rapidly establish a large population of latently infected cells [4], (iii) an increasing genetic diversity of the virus quasi-species [5–7], (iv) heterogeneity of the infected cells with respect to the multiplicity of infection [8–11], and (v) the resulting variability in viral progeny [12]. The multi-physics nature of the HIV infection calls for the development of computational models integrating the infection-driven processes across the systemic physiological-, organ and tissue-, and single-cell

levels of resolution. The description of within-cell virus replication is a cornerstone of the development of multiscale mathematical models of HIV infection [13,14].

The virus life cycle in permissive cells represents the primary driver of HIV infection. So far, a number of mathematical models have been developed to simulate the replication of HIV in an infected cell at different levels of details (see for a list [15]). Importantly, the high-resolution models represent a deterministic description of the biochemical reaction steps underlying the HIV life cycle. However, many of the replication stages are characterized by low numbers of reactants, and hence there is a strong impact of random effects and fluctuations in the reaction kinetics [16]. A systematic analysis of the contribution of stochastic processes to HIV-replication is still missing.

We have recently developed a detailed in silico model of the HIV life cycle in productively infected CD4 T cells [15] formulated with a system of 24 ordinary differential equations (ODE). This deterministic model enables estimation of biochemical parameters of HIV replication. The resulting mechanistic description of the HIV life cycle established a mechanistic basis for the development of a stochastic description of the process kinetics using the Monte Carlo framework.

In this paper, we formulate the stochastic Markov chain-type model of HIV replication in productively infected cells in the form of Gillespie's algorithm [17]. We propose a modification of this algorithm that accelerates the computation process by a factor of 200 compared with the classical Gillespie's algorithm. This is essential, because to obtain the comprehensive statistical data, a very large number of realisations of the stochastic process should be computed.

We apply the stochastic model to examine the statistical characteristics of key aspects of single-cell infection:

- Cell-to-cell variability in HIV progeny production;
- Multiplicity of single-cell infection;
- Global sensitivity of specific reaction steps on net virus production.

Our study delineates the contribution of random effects to (i) the multiplicity of infection, (ii) the cooperative nature of viral replication, and (iii) the variability in the scale of virus secretion by phenotypically identical cells.

2. Results

2.1. Governing Deterministic Equations

The HIV-1 replication cycle presented in Figure 1 consists of multiple stages described below. The specific biochemical reactions and model parameters are from our previous work [15].
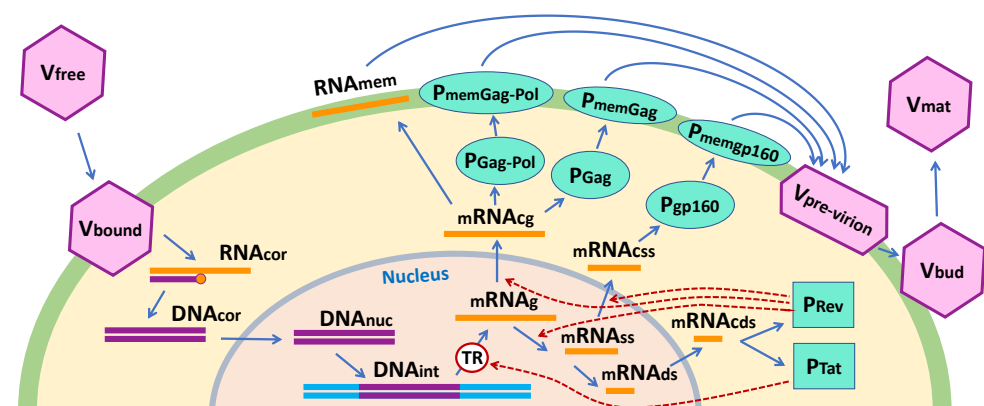

Figure 1. Biochemical scheme of the HIV-1 replication cycle.

2.1.1. Virus Entry

The entry stage is split into three steps (see also [15]) represented in Figure 1:

1. Virion binding to CD4 receptors (the viral glycoprotein gp120 binds to CD4 receptors on the T cell surface);
2. Binding to the co-receptor CCR5 or CXCR4;
3. Virus membrane and cell membrane fusion, i.e., the nucleocapsid is uncoated and the viral RNA is injected into the cell.

The above is described by the following equations:

$$\dot{x}_1 = -(k_{\text{bound}} + d)x_1 \tag{1}$$

$$\dot{x}_2 = k_{\text{bound}}x_1 - (k_{\text{fuse}} + d_{\text{bound}})x_2 \tag{2}$$

where:

$x_1 = [V_{\text{free}}]$ is the number of free virions outside the cell;

$x_2 = [V_{\text{bound}}]$ is the number of virions bound to CD4 and the co-receptor.

The model parameters

$k_{\text{bound}} = 3.1 \, \text{h}^{-1}$; $k_{\text{fuse}} = 0.7 \, \text{h}^{-1}$;

$d = 0.38 \, \text{h}^{-1}$; $d_{\text{bound}} = 0.0008 \, \text{h}^{-1}$;

represent the rate of virion binding to the CD4+ T cell membrane, the rate of virion fusion with the cell, the clearance rate of free mature virions, and the degradation rate of bound virions, respectively. Their reference values and admissible ranges are specified in [15].

2.1.2. Reverse Transcription

Bound virions fuse with the host cell membrane. This step results in the release of the virion content into the cell cytoplasm setting up the stage of reverse transcription [18]. The reverse transcription creates a double-stranded proviral DNA from two single-stranded RNA genomes.

The reverse transcription reaction consists of three sequential processes [15]:

1. Synthesis of minus-strand DNA from viral RNA;
2. Synthesis of plus-strand DNA;
3. Double-strand DNA formation.

This stage is modelled by the following equations:

$$\dot{x}_3 = k_{\text{fuse}}x_2 - (k_{\text{RT}} + d_{\text{RNA}_{\text{cor}}})x_3 \tag{3}$$

$$\dot{x}_4 = k_{\text{RT}}x_3 - (k_{\text{DNA}_t} + d_{\text{DNA}_{\text{cor}}})x_4 \tag{4}$$

where:

$x_3 = [\text{RNA}_{\text{cor}}]$ is the number of genomic RNA molecules in the cytoplasm;

$x_4 = [\text{DNA}_{\text{cor}}]$ is the number of proviral DNA molecules synthesized by reverse transcription.

The model parameters

$k_{\text{RT}} = 0.43 \, \text{h}^{-1}$; $k_{\text{DNA}_t} = 0.12 \, \text{h}^{-1}$;

$d_{\text{RNA}_{\text{cor}}} = 0.21 \, \text{h}^{-1}$; $d_{\text{DNA}_{\text{cor}}} = 0.03 \, \text{h}^{-1}$; represent the reverse transcription rate, the transport rate of DNA from cytoplasm to nucleus, the degradation rate of RNA in the cytoplasm and the degradation rate of DNA in the cytoplasm, respectively. Their reference values and admissible ranges are specified in [15].

2.1.3. Integration

The synthesized proviral DNA associates with virus-encoded integrase (IN) and other proteins. This forms a high-molecular-weight nucleoprotein complex (pre-integration complex, PIC). The PIC is transported into the nucleus for subsequent integration of proviral DNA into the host cell chromosomal DNA [19]. The change in the amount of

proviral DNA in the nucleus and the amount of integrated DNA are modelled with the following equations [15]:

$$\dot{x}_5 = k_{\text{DNA}_t} x_4 - (k_{\text{int}} + d_{\text{DNA}_{\text{nuc}}}) x_5 \tag{5}$$

$$\dot{x}_6 = k_{\text{int}} x_5 - d_{\text{DNA}_{\text{int}}} x_6 \tag{6}$$

where:

$x_5 = [\text{DNA}_{\text{nuc}}]$ is the number of DNA molecules in the nucleus;

$x_6 = [\text{DNA}_{\text{int}}]$ is the number of integrated DNA.

The model parameters

$k_{\text{int}} = 0.14\ \text{h}^{-1}$; $d_{\text{DNA}_{\text{nuc}}} = 0.001\ \text{h}^{-1}$; $d_{\text{DNA}_{\text{int}}} = 0.00002\ \text{h}^{-1}$; represent the integration rate, the degradation rate of DNA in the nucleus and the degradation rate of DNA integrated into the chromosome, respectively. Their reference values and admissible ranges are specified in [15].

2.1.4. Transcription

The transcription of HIV starts after the infected cell receives external activation signals. It results in the synthesis of three types of messenger RNA (mRNA), i.e., the full-length (around 9 kb), singly-spliced (around 4 kb) and doubly-spliced (around 2 kb) [20]. The transcription stage is followed by the transport of mRNAs to the cell cytoplasm. The viral Tat and Rev proteins regulate the transcription and mRNA distribution. These stages are described by the following equations considering the specific parameterizations of the feedback regulation similar to those presented in [21,22]:

$$\dot{x}_7 = f_{\text{TR}}\, x_6 - (k_{\text{ssRNA}_g} g_{\text{Rev}} + k_{\text{eRNA}_g} f_{\text{Rev}} + d_{\text{RNA}_g}) x_7 \tag{7}$$

$$\dot{x}_8 = k_{\text{ssRNA}_g} g_{\text{Rev}}\, x_7 - (k_{\text{dsRNA}_{ss}} g_{\text{Rev}} + k_{\text{eRNA}_{ss}} f_{\text{Rev}} + d_{\text{RNA}_{ss}}) x_8 \tag{8}$$

$$\dot{x}_9 = k_{\text{dsRNA}_{ss}} g_{\text{Rev}}\, x_8 - (k_{\text{eRNA}_{ds}} + d_{\text{RNA}_{ds}}) x_9 \tag{9}$$

$$\dot{x}_{10} = k_{\text{eRNA}_g} f_{\text{Rev}}\, x_7 - (k_{\text{tp,RNA}} + d_{\text{RNA}_g}) x_{10} \tag{10}$$

$$\dot{x}_{11} = k_{\text{eRNA}_{ss}} f_{\text{Rev}}\, x_8 - d_{\text{RNA}_{ss}} x_{11} \tag{11}$$

$$\dot{x}_{12} = k_{\text{eRNA}_{ds}}\, x_9 - d_{\text{RNA}_{ds}} x_{12} \tag{12}$$

where:

$x_7 = [\text{mRNA}_g]$ is the number of HIV mRNA molecules in the nucleus: g for genomic or full-length;

$x_8 = [\text{mRNA}_{ss}]$ is the number of HIV singly spliced (ss) mRNA molecules in the nucleus;

$x_9 = [\text{mRNA}_{ds}]$ is the number of HIV doubly spliced (ds) mRNA molecules in the nucleus;

$x_{10} = [\text{mRNAc}_g]$ is the number of HIV mRNA molecules in the cytoplasm: g for genomic or full-length;

$x_{11} = [\text{mRNAc}_{ss}]$ is the number of HIV singly spliced (ss) mRNA molecules in the cytoplasm;

$x_{12} = [\text{mRNAc}_{ds}]$ is the number of HIV doubly spliced (ds) mRNA molecules in the cytoplasm.

$f_{\text{TR}} = f_{\text{TR}}(x_{16}) = \text{TR}_{\text{cell}} + f_{\text{Tat}}(x_{16}) \text{TR}_{\text{Tat}}$; $f_{\text{Tat}} = f_{\text{Tat}}(x_{16}) = x_{16}/(\theta_{\text{Tat}} + x_{16})$;

$f_{\text{Rev}} = f_{\text{Rev}}(x_{17}) = x_{17}/(\theta_{\text{Rev}} + x_{17})$;

$g_{\text{Rev}} = g_{\text{Rev}}(x_{17}) = 1 - \beta f_{\text{Rev}}$.

The model parameters

$\text{TR}_{\text{cell}} = 15\ \text{h}^{-1}$; $\text{TR}_{\text{Tat}} = 1500\ \text{h}^{-1}$;

$\theta_{\text{Tat}} = 10^3$ molec.; $\theta_{\text{Rev}} = 7.7 \times 10^4$ molec.; $\beta = 0.9$;

$k_{\text{ssRNA}_g} = 2.4\ \text{h}^{-1}$; $k_{\text{eRNA}_g} = 2.3\ \text{h}^{-1}$; $k_{\text{eRNA}_{ss}} = 2.3\ \text{h}^{-1}$;

$k_{\text{dsRNA}_{ss}} = 2.4\ \text{h}^{-1}$; $k_{\text{eRNA}_{ds}} = 4.6\ \text{h}^{-1}$; $k_{\text{tp,RNA}} = 2.8\ \text{h}^{-1}$;

$d_{\text{RNA}_g} = d_{\text{RNA}_{ss}} = d_{\text{RNAc}_{ds}} = 0.12\ \text{h}^{-1}$; represent the cell intrinsic rate of basal transcription, the level of transcription induced by Tat transactivation, the inhibitory effect of Rev on the splicing rates implying their $1/(1-\beta)$-fold reduction at the saturation level of Rev, the rate of splicing for full-length virus RNA, the rate of $[\text{mRNA}_g]$ export from the nucleus, the rate of $[\text{mRNA}_{ss}]$ export from the nucleus, the rate of splicing for singly spliced virus

RNA, the rate of $[\mathrm{mRNA_{ds}}]$ export from the nucleus, the transport rate of $[\mathrm{mRNA_g}]$ to the cell membrane, and the degradation rates of $[\mathrm{mRNA_i}]$, $i \in \{\mathrm{g, ss, ds}\}$, respectively. Their reference values and admissible ranges are specified in [15].

2.1.5. Translation

Ribosomes function to translate the viral mRNA into specific proteins. The Gag and Gag-Pol proteins are coded by the full-length mRNA. The gp160, Vif, Vpu, and Vpr proteins are coded by singly spliced mRNAs. Finally, the Tat, Rev, and Nef proteins are coded by doubly spliced mRNAs. The synthesized proteins then fold into active proteins. In our deterministic model we consider the kinetics of the Gag-Pol, Gag, gp160, Tat, and Rev proteins [15]. The following set of differential equations is used:

$$\dot{x}_{13} = k_{\mathrm{trans}} f_{\mathrm{g,Gag\text{-}Pol}} x_{10} - (k_{\mathrm{tp,Gag\text{-}Pol}} + d_{\mathrm{p,Gag\text{-}Pol}}) x_{13} \tag{13}$$

$$\dot{x}_{14} = k_{\mathrm{trans}} f_{\mathrm{g,Gag}} x_{10} - (k_{\mathrm{tp,Gag}} + d_{\mathrm{p,Gag}}) x_{14} \tag{14}$$

$$\dot{x}_{15} = k_{\mathrm{trans}} f_{\mathrm{ss,gp160}} x_{11} - (k_{\mathrm{tp,gp160}} + d_{\mathrm{p,gp160}}) x_{15} \tag{15}$$

$$\dot{x}_{16} = k_{\mathrm{trans}} f_{\mathrm{ds,Tat}} x_{12} - d_{\mathrm{p,Tat}} x_{16} \tag{16}$$

$$\dot{x}_{17} = k_{\mathrm{trans}} f_{\mathrm{ds,Rev}} x_{12} - d_{\mathrm{p,Rev}} x_{17} \tag{17}$$

where:
$x_{13} = [\mathrm{P_{Gag\text{-}Pol}}]$ is the number of protein molecules: Gag-Pol;
$x_{14} = [\mathrm{P_{Gag}}]$ is the number of protein molecules: Gag;
$x_{15} = [\mathrm{P_{gp160}}]$ is the number of protein molecules: gp160;
$x_{16} = [\mathrm{P_{Tat}}]$ is the number of protein molecules: Tat;
$x_{17} = [\mathrm{P_{Rev}}]$ is the number of protein molecules: Rev.

The model parameters
$k_{\mathrm{trans}} = 524 \, \mathrm{h}^{-1}$;
$f_{\mathrm{g,Gag\text{-}Pol}} = 0.05$; $f_{\mathrm{g,Gag}} = 0.95$; $f_{\mathrm{ss,gp160}} = 0.64$; $f_{\mathrm{ds,Tat}} = 0.025$; $f_{\mathrm{ds,Rev}} = 0.2$;
$k_{\mathrm{tp,Gag\text{-}Pol}} = 2.8 \, \mathrm{h}^{-1}$; $k_{\mathrm{tp,Gag}} = 2.8 \, \mathrm{h}^{-1}$; $k_{\mathrm{tp,gp160}} = 2.8 \, \mathrm{h}^{-1}$;
$d_{\mathrm{p,Gag\text{-}Pol}} = 0.09 \, \mathrm{h}^{-1}$; $d_{\mathrm{p,Gag}} = 0.09 \, \mathrm{h}^{-1}$; $d_{\mathrm{p,gp160}} = 0.02 \, \mathrm{h}^{-1}$;
$d_{\mathrm{p,Tat}} = 0.04 \, \mathrm{h}^{-1}$; $d_{\mathrm{p,Rev}} = 0.07 \, \mathrm{h}^{-1}$; represent the rate of mRNA to proteins translation, f_{ij} stand for the fraction of $[\mathrm{mRNA_i}]$ coding $[\mathrm{P_j}]$, $i \in \{\mathrm{g, ss, ds}\}$, $j \in \{\mathrm{Gag\text{-}Pol, Gag, gp160, Tat, Rev}\}$. Following them, the parameters define the rates of protein $[\mathrm{P_j}]$ transport to membrane, $j \in \{\mathrm{Gag\text{-}Pol, Gag, gp160}\}$, and the degradation rates of proteins Gag-Pol, Gag, gp160, Tat and Rev, respectively. Their reference values and admissible ranges are specified in [15].

2.1.6. Assembly, Budding and Maturation

The last stages of the HIV-1 replication cycle consist of the transport of the proteins (both regulatory and accessory), viral glycoproteins and the full-length mRNA molecules to the cell plasma membrane. The viral proteins (Gag-Pol, Gag, gp160) undergo a number of post-translational modifications. These include folding, oligomerization, glycosylation, and phosphorylation [23]. The final step is the assembly of Gag and Gag-Pol proteins at the cell membrane followed by encapsidation of the viral RNA genomes. The budding of the synthesized virions ends with their maturation [24]. The respective equations of the deterministic model read:

$$\dot{x}_{18} = k_{\mathrm{tp,RNA}} x_{10} - k_{\mathrm{comb}} N_{\mathrm{RNA}} f_{\mathrm{c}} - d_{\mathrm{RNA_g}} x_{18} \tag{18}$$

$$\dot{x}_{19} = k_{\mathrm{tp,Gag\text{-}Pol}} x_{13} - k_{\mathrm{comb}} N_{\mathrm{Gag\text{-}Pol}} f_{\mathrm{c}} - d_{\mathrm{mem,Gag\text{-}Pol}} x_{19} \tag{19}$$

$$\dot{x}_{20} = k_{\mathrm{tp,Gag}} x_{14} - k_{\mathrm{comb}} N_{\mathrm{Gag}} f_{\mathrm{c}} - d_{\mathrm{mem,Gag}} x_{20} \tag{20}$$

$$\dot{x}_{21} = k_{\mathrm{tp,gp160}} x_{15} - k_{\mathrm{comb}} N_{\mathrm{gp160}} f_{\mathrm{c}} - d_{\mathrm{mem,gp160}} x_{21} \tag{21}$$

where: $x_{18} = [\text{RNA}_{\text{mem}}]$, $x_{19} = [\text{P}_{\text{mem,Gag-Pol}}]$, $x_{20} = [\text{P}_{\text{mem,Gag}}]$, $x_{21} = [\text{P}_{\text{mem,gp160}}]$ are, respectively, the number of viral RNA molecules and the Gag-Pol, Gag, and gp160 viral protein molecules at the membrane.

The model parameters

$k_{\text{tp,RNA}} = 2.8\ \text{h}^{-1}$; $k_{\text{tp,Gag-Pol}} = 2.8\ \text{h}^{-1}$; $k_{\text{tp,Gag}} = 2.8\ \text{h}^{-1}$; $k_{\text{tp,gp160}} = 2.8\ \text{h}^{-1}$;

$k_{\text{comb}} = 8.0\ \text{h}^{-1}$; $N_{\text{RNA}} = 2$; $N_{\text{Gag-Pol}} = 250$; $N_{\text{Gag}} = 5000$; $N_{\text{gp160}} = 24$; represent the rates of RNA and protein $[\text{P}_j]$ transport to the membrane, $j \in \{\text{Gag-Pol, Gag, gp160}\}$, the incorporation rate of molecules into pre-virion complexes, the number of viral RNA transcripts in a new virion, the number of Gag-Pol molecules in a new virion, the number of Gag molecules in a new virion, and the number of gp160 molecules in a new virion, respectively. Their reference values and admissible ranges are specified in [15].

The model parameter $K_{\text{V}_{\text{rel}}}$ entering the virion complex assembly function,

$$f_{\text{c}} = f_{\text{c}}(x_{18}, \ldots, x_{21}) = x_{18} \cdot \frac{x_{19}}{x_{19} + K_{\text{V}_{\text{rel}}} N_{\text{Gag-Pol}}} \cdot \frac{x_{20}}{x_{20} + K_{\text{V}_{\text{rel}}} N_{\text{Gag}}} \cdot \frac{x_{21}}{x_{21} + K_{\text{V}_{\text{rel}}} N_{\text{gp160}}},$$

$K_{\text{V}_{\text{rel}}} = 10^3$ refers to the characteristic scale of the viral progeny per replication cycle [13]. The model parameters

$d_{\text{RNA}_{\text{g}}} = 0.12\ \text{h}^{-1}$; $d_{\text{mem,Gag-Pol}} = 0.004\ \text{h}^{-1}$;

$d_{\text{mem,Gag}} = 0.004\ \text{h}^{-1}$; $d_{\text{mem,gp160}} = 0.014\ \text{h}^{-1}$; represent the degradation rates of $[\text{RNA}_{\text{mem}}]$, the membrane-anchored protein Gag-Pol, the membrane-anchored protein Gag, and the membrane-associated gp160, respectively. Their reference values and admissible ranges are specified in [15].

In the current model of HIV replication, the Michaelis–Menten-type parametrization is used to limit the rate of assembling of progeny virions by the least abundant protein component out of the three considered $[\text{P}_{\text{mem,Gag-Pol}}]$, $[\text{P}_{\text{mem,Gag}}]$, $[\text{P}_{\text{mem,gp160}}]$. The parameters $K_{\text{V}_{\text{rel}}}$, $N_{\text{Gag-Pol}}$, N_{Gag}, N_{gp160} specify the reference amount of the viral progeny per cell and the number of protein molecules Gag-Pol, Gag, and gp160 required for each virion, respectively.

Note that the above function $f_{\text{c}}(x_{18}, \ldots, x_{21})$ is the only difference from the similar equations presented in [15], where the corresponding function reads as $f_{\text{c}} = x_{18} x_{19} x_{20} x_{21}$. This modification was proposed in a model of influenza A virus replication [25,26].

At the late phase of virus replication, the viral RNA and viral proteins at the membrane associate with the pre-virion complex and then combine to generate a new virion, as shown in Figure 1

$$\dot{x}_{22} = k_{\text{comb}} f_{\text{c}} - (k_{\text{bud}} + d_{\text{comb}}) x_{22} \tag{22}$$

$$\dot{x}_{23} = k_{\text{bud}} x_{22} - (k_{\text{mat}} + d_{\text{bud}}) x_{23} \tag{23}$$

$$\dot{x}_{24} = k_{\text{mat}} x_{23} - d x_{24} \tag{24}$$

where:

$x_{22} = [\text{V}_{\text{pre-virion}}]$ is the number of virions on the membrane;

$x_{23} = [\text{V}_{\text{bud}}]$ is the number of free viruses after budding from the cell;

$x_{24} = [\text{V}_{\text{mat}}]$ is the number of mature virions outside the cell.

The model parameters

$k_{\text{comb}} = 8\ \text{h}^{-1}$; $k_{\text{bud}} = 2.0\ \text{h}^{-1}$; $k_{\text{mat}} = 2.4\ \text{h}^{-1}$;

$d_{\text{comb}} = 0.52\ \text{h}^{-1}$; $d_{\text{bud}} = 0.38\ \text{h}^{-1}$; $d = 0.38\ \text{h}^{-1}$; represent the incorporation rate of molecules into pre-virion complexes, the budding rate of new virions, the maturation rate, the degradation rate of the assembled pre-virion complex, the degradation rate for budded immature viral-like particles, and the clearance rate of free mature virions, respectively. Their reference values and admissible ranges are specified in [15].

The solution of the deterministic system (1)–(24) with initial condition $[\text{V}_{\text{free}}](0) = 4$ is presented in Figure 2.

2.2. Stochastic Markov Chain Modelling

The use of ordinary differential equations for the biochemical species concentrations assumes that they vary continuously and in a deterministic manner. However, some of the HIV-1 replication steps are characterized by low numbers of the reactants and the greater impact of the random fluctuations in the reaction rates, thus invalidating to a certain degree the deterministic approach. The stochastic re-formulation of the model allows one to overcome the limitations of the deterministic framework. A general approach to stochastic formulation is based on considering the reactions as Markov processes, i.e., random walks in the state space of the system, implemented numerically using Monte Carlo techniques. Markov chain model considers the exact number of species with a discrete change in their numbers according to the probabilities of the individual reaction to happen per unit of time. The later are defined by the reaction rates and the abundance of the chemical species of the underlying deterministic model listed in Table 1. We follow the Gillespie approach [17] as detailed below. The full details of the Markov chain are presented in Table 1. Here, superscript $++$ in x_n^{++} means that this component is increased by one, and $--$ in x_n^{--} means that this component is decreased by one.
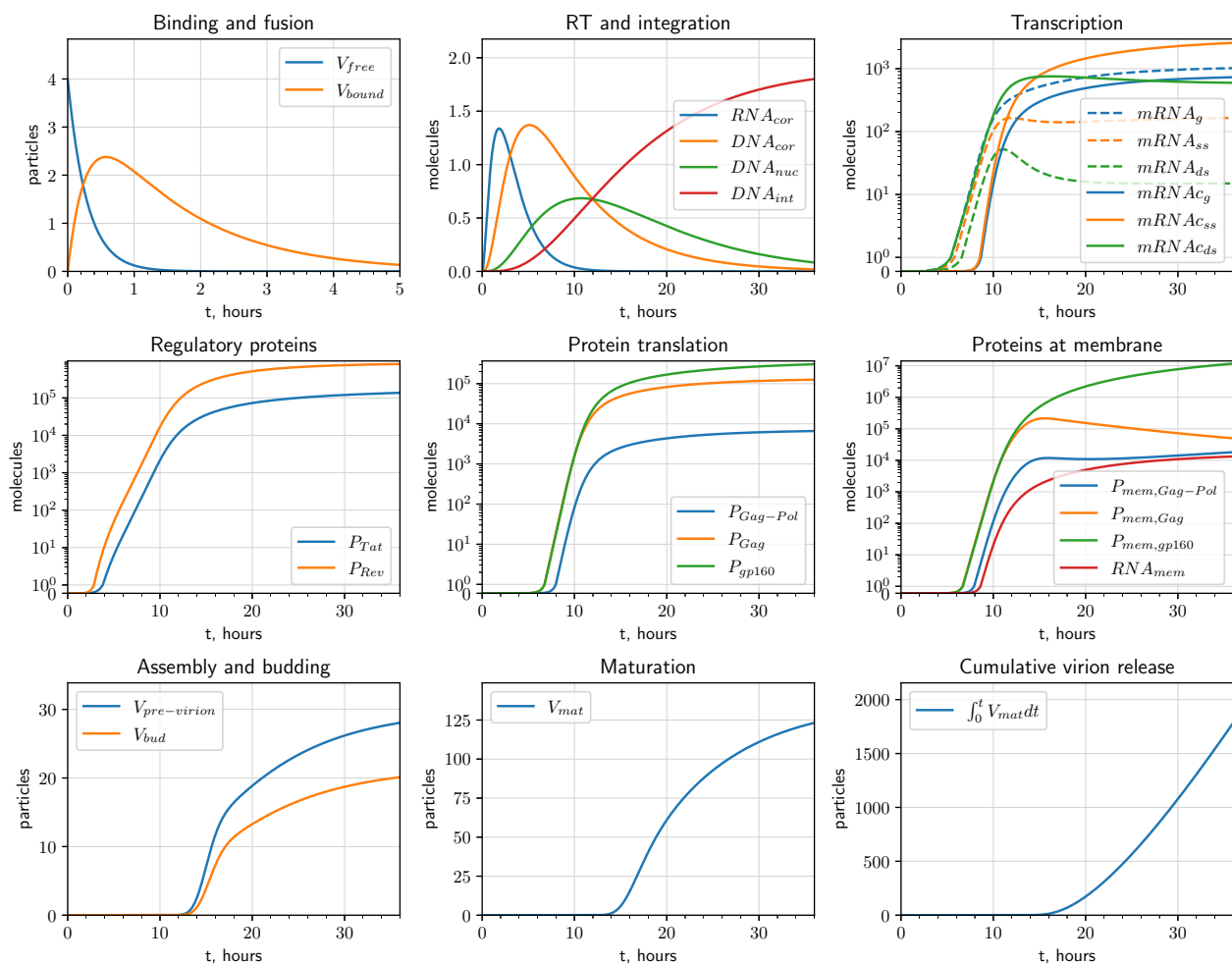

Figure 2. Numerical solution of the deterministic model of HIV-1 replication.

2.2.1. Algorithm

The realizations of the stochastic Markov chain model are inherently variable in their dynamics. To robustly estimate the mean trajectories and uncertainty in the dynamics, an averaging is required over a large ensemble of realizations ($\sim 10^5$). This might appear to be computationally demanding and depends on the simulation time for a single run of the model. Therefore, a method for computer modelling of the Markov chain (1) should be fast

and accurate. To model a Markov chain numerically, several methods have been proposed; the most often used is Gillespie's direct method [17,27,28].

During the process, some components can reach large numbers. This essentially slows down the computation process, as a high population number causes extremely short time intervals between events in the Markov chain. At the same time, the evolution of highly populated components can be modelled accurately enough by a deterministic process described by an appropriate ODE. To overcome this issue, a hybrid approach has been proposed in [29,30] in which the simulation process is switched from stochastic to deterministic for all components simultaneously as soon as the component with the smallest population size exceeds a certain threshold (see also [28]). In [31,32], a method was applied in which the deterministic dynamics for highly populated components and the stochastic dynamics for smaller populated components are coupled and computed in parallel.

Table 1. Markov chain: the list of individual reactions m, the corresponding state transitions and the propensities of reactions which define the probabilities of the individual reactions to occur over a small time interval δt.

m	Transition	Propensity, a_m	m	Transition	Propensity, a_m
1	x_1^{--}, x_2^{++}	$k_{\text{bound}} x_1$	27	x_{13}^{--}	$d_{\text{p,Gag-Pol}} x_{13}$
2	x_1^{--}	$d x_1$	28	x_{14}^{++}	$k_{\text{trans}} f_{\text{g,Gag}} x_{10}$
3	x_2^{--}, x_3^{++}	$k_{\text{fuse}} x_2$	29	x_{14}^{--}, x_{20}^{++}	$k_{\text{tp,Gag}} x_{14}$
4	x_2^{--}	$d_{\text{bound}} x_2$	30	x_{14}^{--}	$d_{\text{p,Gag}} x_{14}$
5	x_3^{--}, x_4^{++}	$k_{\text{RT}} x_3$	31	x_{15}^{++}	$k_{\text{trans}} f_{\text{ss,gp160}} x_{11}$
6	x_3^{--}	$d_{\text{RNA}_{\text{cor}}} x_3$	32	x_{15}^{--}, x_{21}^{++}	$k_{\text{tp,gp160}} x_{15}$
7	x_4^{--}, x_5^{++}	$k_{\text{DNA}_t} x_4$	33	x_{15}^{--}	$d_{\text{p,gp160}} x_{15}$
8	x_4^{--}	$d_{\text{DNA}_{\text{cor}}} x_4$	34	x_{16}^{++}	$k_{\text{trans}} f_{\text{ds,Tat}} x_{12}$
9	x_5^{--}, x_6^{++}	$k_{\text{int}} x_5$	35	x_{16}^{--}	$d_{\text{p,Tat}} x_{16}$
10	x_5^{--}	$d_{\text{DNA}_{\text{nuc}}} x_5$	36	x_{17}^{++}	$k_{\text{trans}} f_{\text{ds,Rev}} x_{12}$
11	x_6^{--}	$d_{\text{DNA}_{\text{int}}} x_6$	37	x_{17}^{--}	$d_{\text{p,Rev}} x_{17}$
12	x_7^{++}	$f_{\text{TR}}(x_{16}) x_6$	38	x_{18}^{--}	$k_{\text{comb}} N_{\text{RNA}} f_c(x_{18,...,21})$
13	x_7^{--}, x_8^{++}	$k_{\text{ssRNA}_g} g_{\text{Rev}}(x_{17}) x_7$	39	x_{18}^{--}	$d_{\text{RNA}_g} x_{18}$
14	x_7^{--}, x_{10}^{++}	$k_{\text{eRNA}_g} f_{\text{Rev}}(x_{17}) x_7$	40	x_{19}^{--}	$k_{\text{comb}} N_{\text{Gag-Pol}} f_c(x_{18,...,21})$
15	x_7^{--}	$d_{\text{RNA}_g} x_7$	41	x_{19}^{--}	$d_{\text{mem,Gag-Pol}} x_{19}$
16	x_8^{--}, x_9^{++}	$k_{\text{dsRNA}_{\text{ss}}} g_{\text{Rev}}(x_{17}) x_8$	42	x_{20}^{--}	$k_{\text{comb}} N_{\text{Gag}} f_c(x_{18,...,21})$
17	x_8^{--}, x_{11}^{++}	$k_{\text{eRNA}_{\text{ss}}} f_{\text{Rev}}(x_{17}) x_8$	43	x_{20}^{--}	$d_{\text{mem,Gag}} x_{20}$
18	x_8^{--}	$d_{\text{RNA}_{\text{ss}}} x_8$	44	x_{21}^{--}	$k_{\text{comb}} N_{\text{gp160}} f_c(x_{18,...,21})$
19	x_9^{--}, x_{12}^{++}	$k_{\text{eRNA}_{\text{ds}}} x_9$	45	x_{21}^{--}	$d_{\text{mem,gp160}} x_{21}$
20	x_9^{--}	$d_{\text{RNA}_{\text{ds}}} x_9$	46	x_{22}^{++}	$k_{\text{comb}} f_c(x_{18,...,21})$
21	x_{10}^{--}, x_{18}^{++}	$k_{\text{tp,RNA}} x_{10}$	47	x_{22}^{--}, x_{23}^{++}	$k_{\text{bud}} x_{22}$
22	x_{10}^{--}	$d_{\text{RNA}_g} x_{10}$	48	x_{22}^{--}	$d_{\text{comb}} x_{22}$
23	x_{11}^{--}	$d_{\text{RNA}_{\text{ss}}} x_{11}$	49	x_{23}^{--}, x_{24}^{++}	$k_{\text{mat}} x_{23}$
24	x_{12}^{--}	$d_{\text{RNA}_{\text{ds}}} x_{12}$	50	x_{23}^{--}	$d_{\text{bud}} x_{23}$
25	x_{13}^{++}	$k_{\text{trans}} f_{\text{g,Gag-Pol}} x_{10}$	51	x_{24}^{--}	$d x_{24}$
26	x_{13}^{--}, x_{19}^{++}	$k_{\text{tp,Gag-Pol}} x_{13}$			

Here we further develop this approach and propose an algorithm with coupled stochastic and deterministic processes with capability to automatically switch the dynamics of any component, x_n, from stochastic to deterministic and back as soon as this component exceeds a predefined threshold $\bar{X}$ or, respectively, becomes below it. Therefore, at any time interval between the transitions, all components are divided into two time-varying sets: $\mathcal{S}_X = \{n : x_n \leq \bar{X}\}$ and $\overline{\mathcal{S}_X} = \{n : x_n > \bar{X}\}$. Components, $x_n \in \mathcal{S}_X$, currently having a relatively small number of particles are modelled stochastically by the Markov chain described in Table 1. Other components, $x_n \in \overline{\mathcal{S}_X}$, with a large population size are modelled by the corresponding deterministic Equations (1)–(24). With the change of population, a component, x_n, can be moved automatically from one set to another.

We also divide the reactions into the sets of stochastic ones, $\mathcal{S}_T$, for which the transition in Table 1 contains at least one stochastically modelling component, and deterministic ones, $\overline{\mathcal{S}_T}$, with transitions without such a component at the current time interval. The stochastic reactions $m \in \mathcal{S}_T$ are modelled by Gillespie's next reaction method: at every step two uniformly distributed random numbers r_1, r_2 on $(0,1)$ are generated. The first number gives the next time step $\Delta t = -(\ln r_1)/A$ where $A = \sum_{m \in \mathcal{S}_T} a_m$. The second random number determines the next reaction index $p \in \mathcal{S}_T$: the smallest integer satisfying $A_p > Ar_2$ where $A_p = \sum_{m=1, m \in \mathcal{S}_T}^{p} a_m$, i.e., the summations are made on stochastic reactions only. As we have to search between up to 51 reactions, the binary search is used to accelerated finding the reaction p. Additionally, to accelerate the computation, the propensity is updated only for reactions in which a_m depends on changed components in the given step. For this purpose, a special array is prepared in which propensities to be updated are indicated for a given component x_n and another similar array for every reaction m. As for deterministically varying components, $x_n, n \in \overline{\mathcal{S}_X}$, the corresponding equations from the ODE system (1)–(24) are integrated employing the predictor–corrector method, as described in [32]. To provide a proper accuracy, the time step is restricted by the value $\Delta t_{\max}$.

The algorithm is implemented in C++. Here arrays of pointers to functions are actively used to directly call functions that should be calculated for a given reaction without spending time on other reactions; this also accelerates the computation. We compute functions of propensities a_m for stochastic reactions or the right-hand sides of ODEs (1)–(24) for deterministic reactions.

The hybrid code is flexible: for a negative threshold $\bar{X} < 0$, only deterministic equations are integrated. Here the solution coincides with that shown in Figure 2 obtained by direct integration of ODEs (1)–(24) with the use of high order adaptive step size methods. If the threshold is set extremely high, say, 10^{10}, such that the probability to reach this value for any component is infinitesimal, then all components are treated as stochastic, modelled by Markov chain Table 1 and computed by Gillespie's method. This allows the estimation of the accuracy of the hybrid algorithm by comparing histograms for the hybrid and the fully stochastic models. We set the threshold $\bar{X} = 2000$ for our computations. Our study showed that the computed statistical characteristics coincide very closely with those obtained by Gillespie's method.

The computations were run on an Intel Xeon E3-1220 v5 CPU 3GHz$\times$4. One realization of the full model with $[V_{\text{free}}](0) = 4$ required about 14 s of CPU time, while the hybrid model was computed in 60 milliseconds on average. Thus, the use of the modified hybrid scheme proposed here accelerates the computation by a factor of 200. We computed 10^5 realizations for every value of an initial number of free virions $[V_{\text{free}}](0)$ to obtain the statistic characteristics described in the next section.

2.2.2. Stochastic Modelling Results

The stochastic modelling enables computation of the probability of a cell to be infected in a certain time depending on the initial number of free virions attaching to the cell. Experimental data presented in [10] indicate that the number of provirus copies in splenic cells of SIVmac251-infected rhesus macaques ranged from one to three. A number of virions attaching to the target cell equal to four ($[V_{\text{free}}](0) = 4$) enables the integration of a similar number of proviral DNA, thus approximating the scenario of in vivo infection. The corresponding plots are shown in Figure 3A. There is a smooth monotonically growing dependence on $[V_{\text{free}}](0)$ tending to one with the growth of this value. Furthermore, the probability of successful integration increases with time.
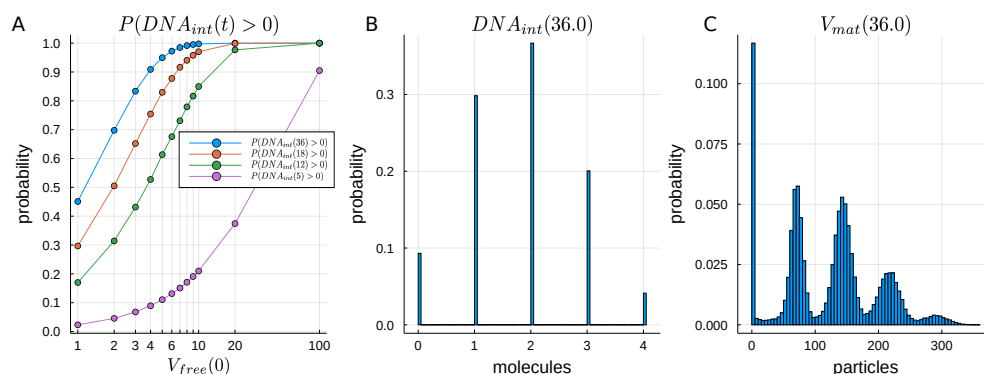
Figure 3. Probability of cell infection (i.e., integration of at least one proviral DNA molecule $[\text{DNA}_{\text{int}}]$) at a certain time as a function of the initial number of free virions $[\text{V}_{\text{free}}](0)$ (**A**). Histograms of number of integrated proviral DNA molecules $[\text{DNA}_{\text{int}}]$ (**B**), number of released matured virions $[\text{V}_{\text{mat}}]$ (**C**) at $t = 36$ h for initial condition $[\text{V}_{\text{free}}](0) = 4$.

The random effects in reaction kinetics result in a heterogeneous distribution of cells with respect to the amount of integrated proviral DNA (Figure 3B). The simulation results indicate that for the parameter settings specified in Table 1, infection with four virus particles leads to a distribution of integrated DNAs ranging from zero to four with the median value of two. The distribution of $[\text{DNA}_{\text{int}}]$ defines the phenotypic diversity of secreted virions as shown in Figure 3C. Note that the above computation did not consider intracellular restriction mechanisms that would modify these values.

The evolution of the histograms of the stochastic model variables for initial condition $[\text{V}_{\text{free}}](0) = 4$, i.e., MOI = 4, is presented in Figure 4. They show that the abundance of the biochemical components underlying the virus replication is not homogeneous but displays multi-hump patterns indicated by strips with darker colours. The number of humps depends on the species and ranges from two to five. This is clearly seen in Figure 5, where the histograms of all components are presented at time $t = 36$ h for initial condition $[\text{V}_{\text{free}}](0) = 4$.

The variance in the intensities of individual reaction stages of the life cycle is shown in Figure 6. They are consistent with the transcription and translation data presented in [33].

2.3. Sensitivity Analysis

In a previous work [15], the local sensitivity of the deterministic model was computed using the adjoint method to identify the potential targets for antiviral therapy. As we have modified the model by considering the features of the assembly in more detail, we conducted the sensitivity analysis for the extended model. We considered the global variance-based method of sensitivity analysis, which is used to determine the contributions of individual parameter variations to the changes in the variance of the model output while allowing simultaneous variations across the whole input space of all model parameters. Let $y = f(\mathbf{p})$ be the model output where $\mathbf{p} = [p_1, \dots, p_L]$ is a vector of inputs, i.e., model parameters, which are treated as independently distributed random variables. The variance-based methods of global sensitivity analysis rely on the following decomposition of the output variance [34,35]:

$$\text{Var}(y) = \sum_{i=1}^{L} V_i + \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} V_{ij} + \dots + V_{1\,2\dots L}, \tag{25}$$

which includes the variance V_i caused by varying p_i alone, the variance of second-order interactions V_{ij} caused by varying p_i and p_j simultaneously (additionally to individual V_i and V_j) and of the other higher-order interactions, up to the variance $V_{1\,2\dots L}$ caused by varying all components.
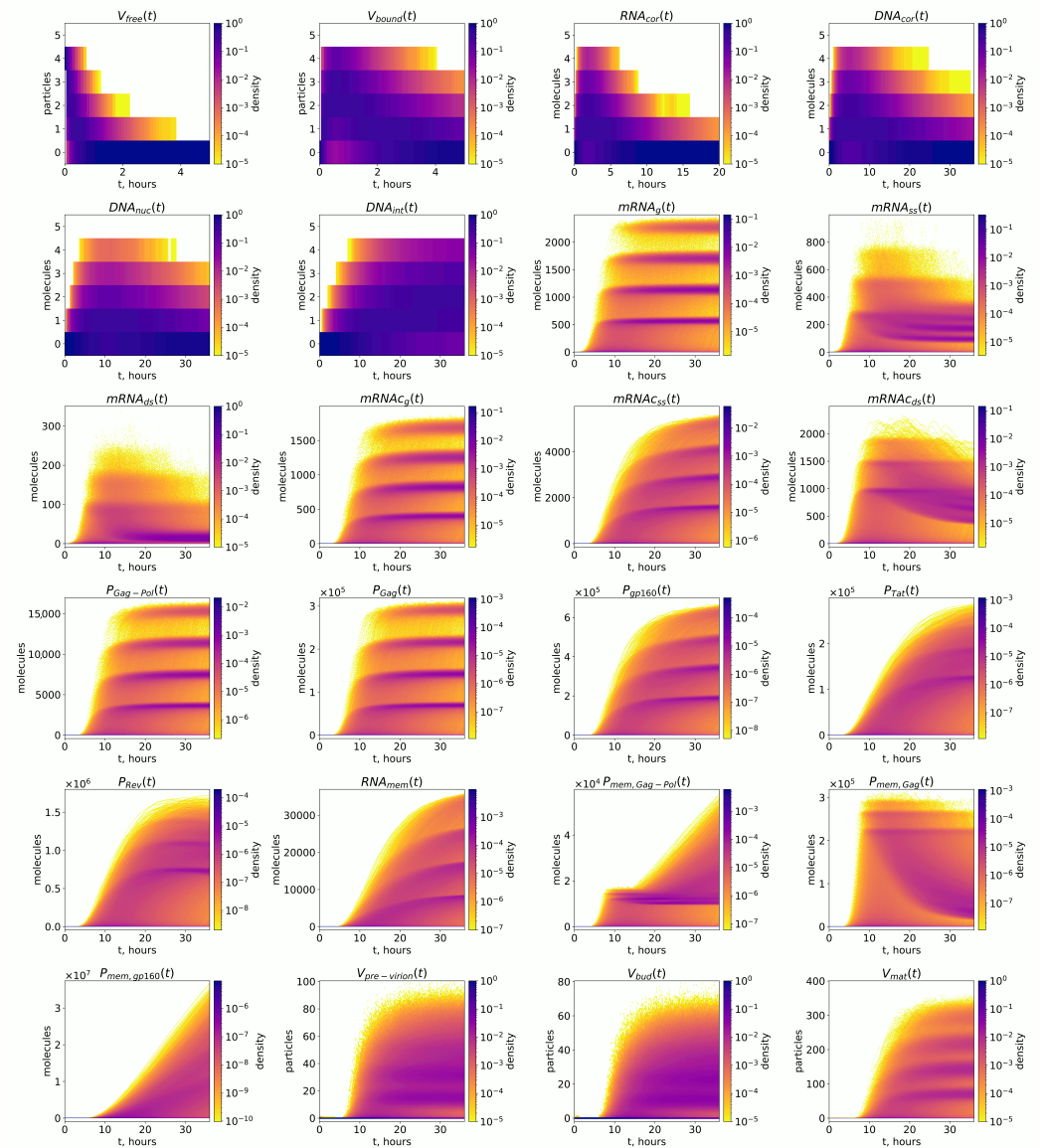
Figure 4. Temporal evolution of the histograms of the stochastic model variables for the initial condition $[V_{free}](0) = 4$.
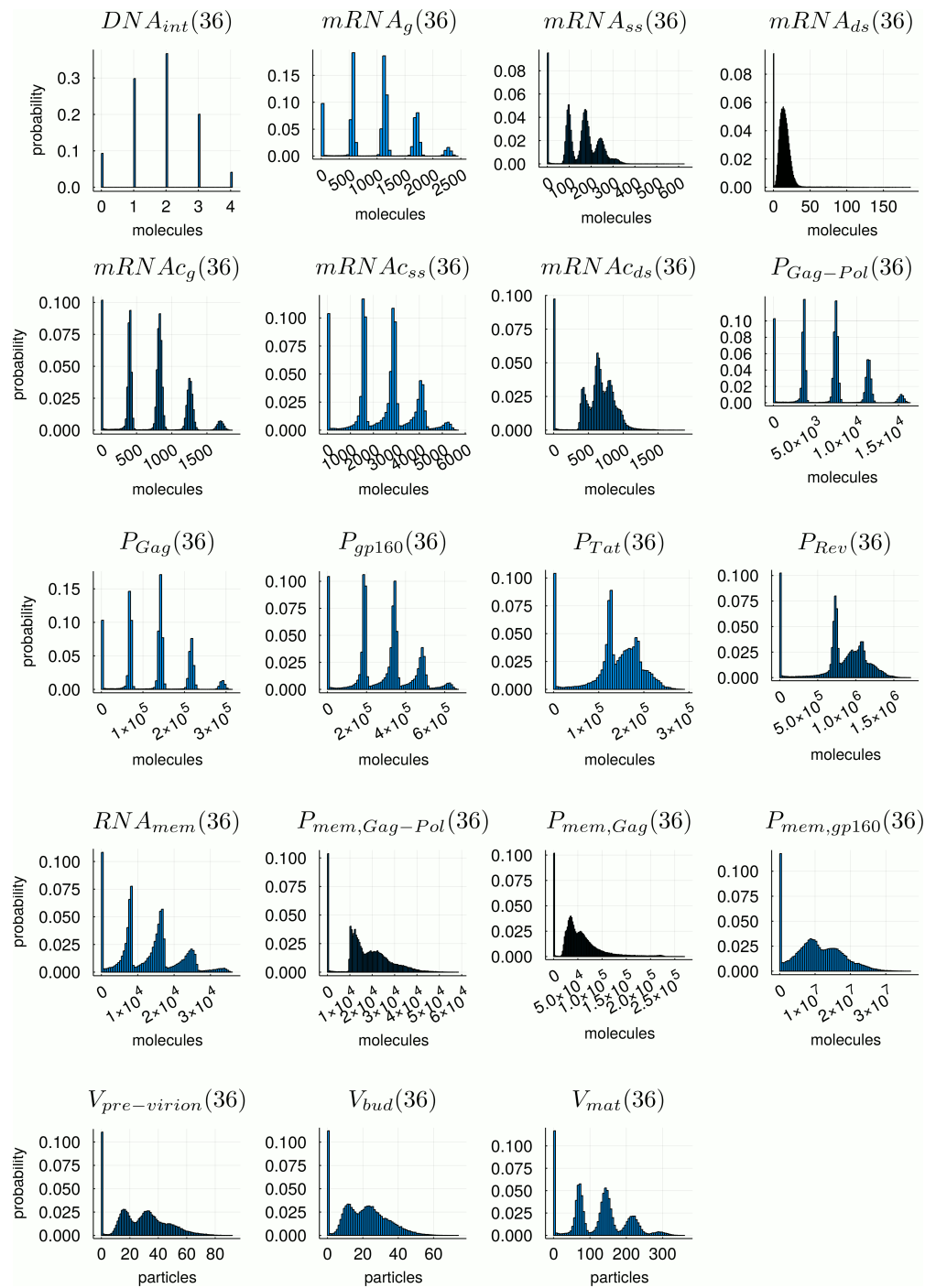
Figure 5. Histograms of the stochastic model variables at $t = 36$ h for the initial condition $[\mathrm{V_{free}}](0) = 4$.
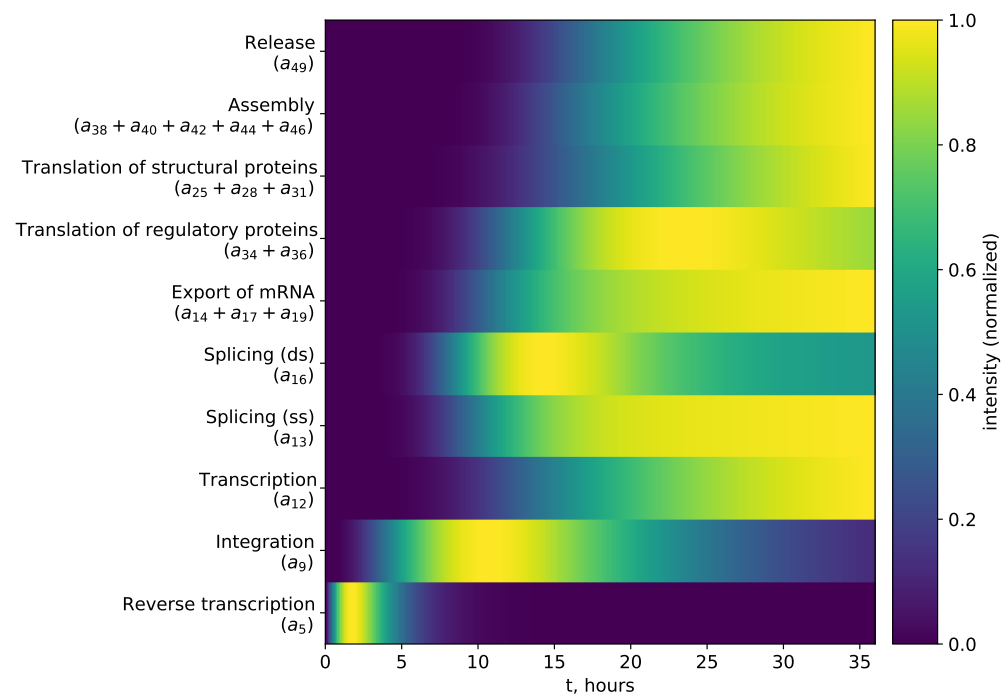
Figure 6. Temporal dynamics of the intensities of stages of the HIV-1 replication life cycle. The propensities that constitute each stage intensity are indicated on the vertical axis. The propensities represent the average values over stochastic realizations normalized by their maximum values.

The first-order S_i and total-order S_i^{total} sensitivity indices are defined as

$$S_i = \frac{V_i}{\text{Var}(y)} = \frac{\text{Var}_{p_i}(E_{\mathbf{p}_{\sim i}}(y|p_i))}{\text{Var}(y)}, \tag{26}$$

$$S_i^{\text{total}} = 1 - \frac{\text{Var}_{\mathbf{p}_{\sim i}}(E_{p_i}(y|\mathbf{p}_{\sim i}))}{\text{Var}(y)}, \tag{27}$$

where $\mathbf{p}_{\sim i} = \{p_l\}_{l \in \overline{1,L}}^{l \neq i}$ is a set of all parameters except p_i. The sensitivity index S_i measures the effect of varying parameter p_i alone, averaged over variations in other parameters $\mathbf{p}_{\sim i}$, while S_i^{total} approximates the effects of variations of p_i including all of the variance caused by its interactions with other parameters.

There are two well-established methods to compute the sensitivity indices specified in Equations (26) and (27). Sobol's method employs low-discrepancy sequences and Monte-Carlo integration methods to explore the input space of parameter variations. The extended Fourier amplitude sensitivity testing (eFAST) method makes use of Fourier decomposition to search in the frequency space along mono-dimensional search curves for each parameter, which makes it possible to obtain both first- and total-order indices with a lower number of model evaluations than with Sobol's method [36].

2.3.1. Sensitivity Analysis of the Deterministic Model towards the Cumulative Virion Release

To study the sensitivity of the processes towards the cumulative number of released virions throughout 36 h (the area under the curve-type of the metric for released virions $J(\boldsymbol{y}) = \int_0^T [\text{V}_{\text{mat}}] dt$, which characterises the availability of the virus for infection of other cells, or the total virus exposure), we applied the eFAST method to the deterministic model with $[\text{V}_{\text{free}}](0) = 4$ virions. We assumed the uniform distribution of model parameters in the ranges specified in [15]. For the newly introduced parameter $K_{\text{V}_{\text{rel}}} = 10^3$ we assumed the range from 10^2 to 10^4 virions. As proposed in [37], we included in the input variables a

"dummy" input, which is uniformly distributed in the unit segment and is not presented in the model. The total sensitivity index of dummy input $S_{\text{dummy}}^{\text{total}}$ should have a low value that represents the variance from interactions between other parameters but not with the parameter of interest, which is included in the definition of the total order index (27) [37]. The dummy input can be regarded as a negative control to determine significant sensitivity indices relative to the level of variance $S_{\text{dummy}}^{\text{total}}$.

The first- and total-order sensitivity indices to the cumulative number of released virions throughout 36 h computed with the eFAST method are presented in Figure 7. We used $N_r = 20$ resamples of the method to obtain the means and standard errors of the means of sensitivity indices, and to test for their significance with Welch's t-test. The sample size for each search curve was set to $N_s = 10^4$. The sensitivity indices of the parameters that have the significant effects are summarized in Table 2.

Table 2. Sensitivity indices of the model parameters that have the biggest influence on cumulative virion release. The parameters are sorted in descending order by total order sensitivity indices. The means $\pm$ SEMs obtained on $N_r = 20$ resamples are indicated.

Parameter	Description	Sensitivity Indices	
		First Order	Total Order
$k_{\text{tp,RNA}}$	Transport of genomic mRNA to membrane	0.099 ± 0.006	0.669 ± 0.004
θ_{Tat}	Tolerance of transcription activation to Tat-mediated regulation	0.089 ± 0.009	0.450 ± 0.009
$k_{\text{trans}} f_{\text{ds,Tat}}$	Translation of Tat molecules	0.019 ± 0.001	0.235 ± 0.001
d	Degradation of free and mature virions	0.022 ± 0.002	0.226 ± 0.004
$k_{\text{trans}} f_{\text{g,Gag-Pol}}$	Translation of Gag-Pol molecules	0.0155 ± 0.0009	0.201 ± 0.002
$k_{\text{trans}} f_{\text{g,Gag}}$	Translation of Gag molecules	0.0153 ± 0.0006	0.177 ± 0.006
$d_{\text{DNA}_{\text{cor}}}$	Degradation of DNA during RT	0.019 ± 0.002	0.158 ± 0.005
k_{comb}	Assembly of pre-virion complexes	0.0028 ± 0.0001	0.068 ± 0.002
TR_{Tat}	Tat-induced transcription rate	0.0049 ± 0.0004	0.066 ± 0.003
$k_{\text{comb}} N_{\text{Gag}}$	Gag contribution to pre-virion assembly	0.0014 ± 0.0001	0.063 ± 0.002
$k_{\text{comb}} N_{\text{Gag-Pol}}$	Gag-Pol contribution to pre-virion assembly	0.00125 ± 0.00005	0.063 ± 0.002
$d_{\text{RNA}_{\text{ds}}}$	Degradation of doubly-spliced mRNA	0.0027 ± 0.0003	0.058 ± 0.004
$K_{\text{V}_{\text{rel}}} N_{\text{Gag}}$	Tolerance of pre-virion assembly to Gag availability on membrane	0.0036 ± 0.0006	0.051 ± 0.003

2.3.2. Sensitivity Analysis of Early Stages of Cell Infection Using the Stochastic Model

An important aspect of HIV-1 infection is the establishment of a pool of latently infected cells. After the proviral genome integration into the chromosome of the host cell [DNA$_{\text{int}}$] the provirus may remain silent, i.e., transcription of viral genes can be suppressed by the infected cell. Although our model does not describe the mechanisms by which this suppression is regulated, we can investigate the impact of variations of the early stage processes on the distribution of the number of integrated proviral DNA molecules [DNA$_{\text{int}}$](36) using the stochastic model.

To perform a sensitivity analysis on the stochastic model, one needs to define the model output that can be compared between the ensembles of stochastic realizations with perturbed parameters. Given the ensemble of model runs, histograms and empirical cdfs approximate the pdf and cdf of the variables at certain times, and various distance functions defined on them can be used. In [38], the sensitivity analysis of stochastic Markov Chain models was performed using the histogram distance between the unperturbed model output and the output with perturbed parameters. The histogram distance between sets $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$D(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{K} \left| \frac{\sum_{j=1}^{n_x} \chi(x_j, I_k)}{n_x} - \frac{\sum_{j=1}^{n_y} \chi(y_j, I_k)}{n_y} \right| \tag{28}$$

where n_x and n_y are the number of elements in **x** and **y**, respectively, K is the number of histogram bins dividing the range from $\min\{x_{\min}, y_{\min}\}$ to $\max\{x_{\max}, y_{\max}\}$ in I_k intervals, and $\chi(x_j, I_k)$ is a characteristic function that is equal to 1 if x_j lies in I_k and 0 otherwise.
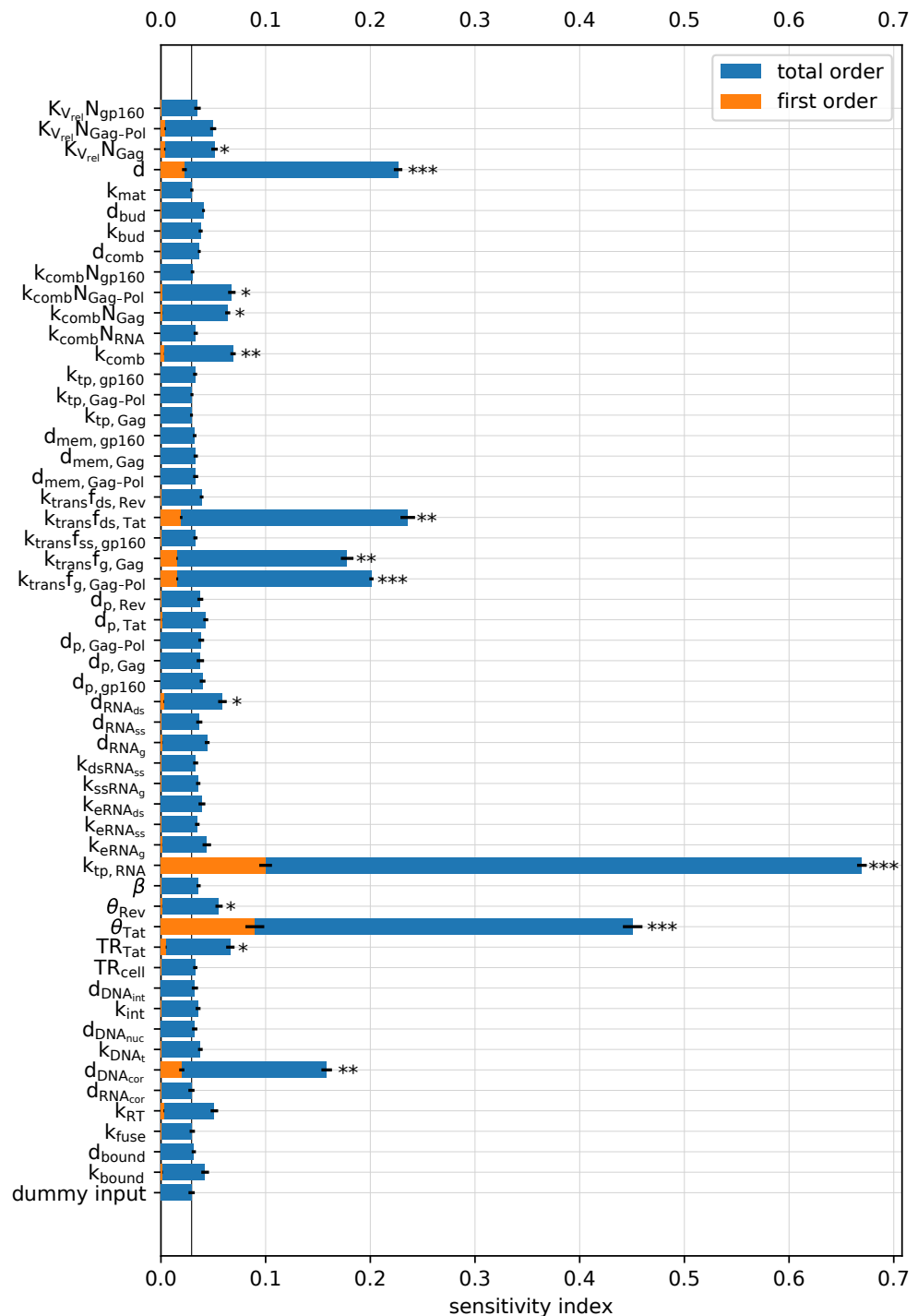

Figure 7. Sensitivity indices for the cumulative number of released virions up to 36 h computed with the deterministic model. The bars and the error bars correspond to the means $\pm$ SEMs obtained on $N_r = 20$ resamples. The significance levels of Welch's t-test for comparing total order indices against $S_{\text{dummy}}^{\text{total}}$ are $p < 10^{-20}$ (***), $p < 10^{-10}$ (**), $p < 10^{-5}$ (*).

The local sensitivity indices can be defined as

$$S_i = \frac{D(\mathbf{x_0}, \mathbf{x_{p_i}})}{\Delta p_i}, \quad \hat{S}_i = S_i p_i \tag{29}$$

where $\mathbf{x_0}$ is the set of model outputs with baseline parameters, $\mathbf{x_{p_i}}$, with the increased parameter p_i on a small value Δp_i. The indices $\hat{S}_i$ were scaled on baseline parameter values to make their ranking possible.

Here, we only implement the local sensitivity approach due to the computational costs needed to obtain the ensemble of stochastic realizations with a fixed set of parameters. However, the low-cost screening methods of global sensitivity analysis, such as the Morris method [39,40], can also be effectively implemented for stochastic models [38].

The model output is the number of integrated proviral genomes $x_6 = [\text{DNA}_{\text{int}}]$ at $t = 36$ h. We obtained an ensemble of 10^5 model runs for each parameter set. For this number of runs, the measured baseline self-distance $D(\mathbf{x_0}, \mathbf{x_0})$ equals zero. The baseline parameter values are perturbed by 1%. In all stochastic realizations, the possible values for model output $\{0, 1, 2, 3, 4\}$ are the same. Hence, $K = 5$. The sensitivity indices $\hat{S}_i$ are reported in Table 3.

Table 3. Local sensitivity indices $\hat{S}_i$ of the stochastic model parameters having the most influence on the number of integrated proviruses $[\text{DNA}_{\text{int}}](36)$.

Parameter	Description	Sensitivity
k_{bound}	Binding rate of free virions to CD4+ T cell membrane	0.498
d	Degradation rate of HIV particles	0.752
k_{fuse}	Rate of virion fusion into the cell	0.988
d_{bound}	Degradation rate of virions bound to membrane	1.080
k_{RT}	Rate of reverse transcription	0.804
$d_{\text{RNA}_{\text{cor}}}$	Degradation rate of viral RNA in cytoplasm	0.712
k_{DNA_t}	Rate of viral DNA transfer to nucleus	0.738
$d_{\text{DNA}_{\text{cor}}}$	Degradation rate of viral DNA in cytoplasm	0.464
k_{int}	Rate of viral DNA integration into host chromosome	0.446
$d_{\text{DNA}_{\text{nuc}}}$	Degradation rate of viral DNA in nucleus	0.346
$d_{\text{DNA}_{\text{int}}}$	Degradation rate of DNA integrated into chromosome	0.582

3. Discussion

Our study provides a high-resolution model of the stochastic dynamics of HIV replication in productively infected cells. The model is built as a Markov chain and implemented by a Gillespie-type algorithm [17]. In contrast to the deterministic prototype model [15], the stochastic version considers a Mechaelis–Menten type description of the engagement of Gag and Gag-Pol polyproteins and gp160 glycoproteins into an assembly of nascent virions forming around dimeric viral RNA. The numerical implementation of the model was characterised by using (i) a hybrid method to adaptively manage the low and high abundance of molecular components, (ii) binary search, and (iii) a list of pointers of active reaction components at each step. These make single runs fast and computationally less demanding. A global sensitivity analysis method, i.e., the extended Fourier amplitude sensitivity testing (eFAST) [36,37], was implemented. Note that the representation of the Mechaelis–Menten kinetics in the stochastic algorithm requires further systematic analysis and justification.

Extensive simulations with the model led to a number of biologically relevant insights concerning the heterogeneity of cell infection and viral replication. First, the infection with a fixed number of viruses per CD4 T cell ($V_0 = [\text{V}_{\text{free}}](0)$) leads to a heterogeneity of infected cells with respect to the number of integrated proviral genomes ranging from zero to V_0 (see Figure 3). This finding suggest that in the analysis of the mechanisms behind multiply infected cell populations, not only phenotypic differences in cell susceptibility need to be considered [11] but also the inherent randomness and the discrete nature

of the biochemical stages of HIV infection. The multiplicity of infection translates into histograms of the viral progeny showing multimodal Gaussian-type distributions with clear peaks at some regularly spaced values with increasing variance from left to the right. Second, the computed dynamics of the distributions of the 24 HIV life cycle components allows one to identify the bottleneck factors in virus production (see Figure 5). For the reference set of the model parameters, these are Gag-Pol proteins. The stochastic model provides a direct insight into the intensities (fluxes) of the reaction steps through the life cycle of the virus (see Figure 6), which are comparable with experimentally measured transcription and translation events [33]. Third, local and global sensitivity analysis of the stochastic model provided a ranking of model parameters that strongly affect the size of viral progeny, and thus can be considered as potential therapeutic targets. In contrast to the local sensitivity features of the deterministic model [15], the first three globally influential parameters are:

- Transport of genomic mRNA to membranes;
- Tolerance of transcription activation to Tat-mediated regulation;
- Degradation of free and mature virions.

From a biological point of view, an extension of the model is required to consider antiviral defence mechanisms activated in the infected cells (e.g., type I interferon, SAMHD1, Tetherin) and the counteraction of the viral proteins (Vpu, Vif) inhibiting them. These mechanisms will impact the proviral copy numbers and virion burst size. In addition, the transport kinetics of the virus life cycle constituents requires a proper incorporation into the model [41]. These will be subject of our future work. Overall, our results provide a new tool to simulate the HIV life cycle in infected cells and suggest that stochastic effects inherent in the HIV replication cycle must be considered among the relevant mechanisms contributing to the phenotypic diversity and variability of dynamics of HIV infection. A clear distinction between deterministic and stochastic components of HIV-infected host cell interactions will provide a better understanding of the origins of heterogeneity of the viral replication. The knowledge could be further utilized in the analyses of variability of other virus infections, such as HBV, HCV, SARS-CoV-2.

Abbreviations

The following abbreviations are used in this manuscript:

RT   Reverse Transcription
pdf   probability density function
cdf   cumulative distribution function

References

1. Gao, Y.; McKay, P.; Mann, J. Advances in HIV-1 vaccine development. *Viruses* **2018**, *10*, 167. [CrossRef] [PubMed]
2. Iwasaki, A.; Omer, S.B. Why and how vaccines work. *Cell* **2020**, *183*, 290–295. [CrossRef]
3. Gurdasani, D.; Iles, L.; Dillon, D.G.; Young, E.H.; Olson, A.D.; Naranbhai, V.; Fidler, S.; Gkrania-Klotsas, E.; Post, F.A.; Kellam, P.; et al. A systematic review of definitions of extreme phenotypes of HIV control and progression. *AIDS* **2014**, *28*, 149–162. [CrossRef] [PubMed]
4. Siliciano, R.F.; Greene, W.C. HIV latency. *Cold Spring Harb. Perspect. Med.* **2011**, *1*, a007096. [CrossRef] [PubMed]

5.   Hahn, B.H.; Shaw, G.M.; Taylor, M.E.; Redfield, R.R.; Markham, P.D.; Salahuddin, S.Z.; Wong-Staal, F.; Gallo, R.C.; Parks, E.S.; Parks, W.P. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* **1986**, *232*, 1548–1553. [CrossRef] [PubMed]

6.   Meyerhans, A.; Cheynier, R.; Albert, J.; Seth, M.; Kwok, S.; Sninsky, J.; Morfeldt-Månson, L.; Asjö, B.; Wain-Hobson, S. Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. *Cell* **1989**, *58*, 901–910. [CrossRef]

7.   Zanini, F.; Brodin, J.; Thebo, L.; Lanz, C.; Bratt, G.; Albert, J.; Neher, R.A. Population genomics of intrapatient HIV-1 evolution. *eLife* **2015**, *4*, e11282. [CrossRef]

8.   Josefsson, L.; King, M.S.; Makitalo, B.; Brännström, J.; Shao, W.; Maldarelli, F.; Kearney, M.F.; Hu, W.S.; Chen, J.; Gaines, H.; et al. Majority of CD4+ T cells from peripheral blood of HIV-1–infected individuals contain only one HIV DNA molecule. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11199–11204. [CrossRef]

9.   Jung, A.; Maier, R.; Vartanian, J.P.; Bocharov, G.; Jung, V.; Fischer, U.; Meese, E.; Wain-Hobson, S.; Meyerhans, A. Multiply infected spleen cells in HIV patients. *Nature* **2002**, *418*, 144. [CrossRef]

10.   Schultz, A.; Sopper, S.; Sauermann, U.; Meyerhans, A.; Suspène, R. Stable multi-infection of splenocytes during SIV infection—The basis for continuous recombination. *Retrovirology* **2012**, *9*, 31. [CrossRef]

11.   Ito, Y.; Remion, A.; Tauzin, A.; Ejima, K.; Nakaoka, S.; Iwasa, Y.; Iwami, S.; Mammano, F. Number of infection events per cell during HIV-1 cell-free infection. *Sci. Rep.* **2017**, *7*, 6559. [CrossRef]

12.   Andreu-Moreno, I.; Bou, J.V.; Sanjuán, R. Cooperative nature of viral replication. *Sci. Adv.* **2020**, *6*, eabd4942. [CrossRef] [PubMed]

13.   Bocharov, G.; Chereshnev, V.; Gainova, I.; Bazhan, S.; Bachmetyev, B.; Argilaguet, J.; Martinez, J.; Meyerhans, A. Human immunodeficiency virus infection: From biological observations to mechanistic mathematical modelling. *Math. Model. Nat. Phenom.* **2012**, *7*, 78–104. [CrossRef]

14.   Bouchnita, A.; Bocharov, G.; Meyerhans, A.; Volpert, V. Towards a multiscale model of acute HIV infection. *Computation* **2017**, *5*, 6. [CrossRef]

15.   Shcherbatova, O.; Grebennikov, D.; Sazonov, I.; Meyerhans, A.; Bocharov, G. Modeling of the HIV-1 life cycle in productively infected cells to predict novel therapeutic targets. *Pathogens* **2020**, *9*, 255. [CrossRef]

16.   Weinberger, L.S.; Burnett, J.C.; Toettcher, J.E.; Arkin, A.P.; Schaffer, D.V. Stochastic gene expression in a Lentiviral positive-feedback Loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* **2005**, *122*, 169–182. [CrossRef]

17.   Gillespie, D.T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **1977**, *81*, 2340–2361. [CrossRef]

18.   Hu, W.S.; Hughes, S. HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med* **2012**, *2*, a006882. [CrossRef] [PubMed]

19.   Craigie, R.; Bushman, F. HIV DNA integration. *Cold Spring Harb. Perspect. Med.* **2012**, *2*, a006890. [CrossRef]

20.   Chereshnev, V.; Bocharov, G.; Bazhan, S.; Bachmetyev, B.; Gainova, I.; Likhoshvai, V.; Argilaguet, J.; Martinez, J.; Rump, J.; Mothe, B.; et al. Pathogenesis and treatment of HIV infection: The cellular, the immune system and the neuroendocrine systems perspective. *Int. Rev. Immunol.* **2013**, *32*, 1–25. [CrossRef] [PubMed]

21.   Kim, H.; Yin, J. Robust growth of human immunodeficiency virus type 1 (HIV-1). *Biophys. J.* **2005**, *89*, 2210–2221. [CrossRef] [PubMed]

22.   Likhoshvai, V.; Khlebodarova, T.; Bazhan, S.; Gainova, I.; Chereshnev, V.; Bocharov, G. Mathematical model of the Tat-Rev regulation of HIV-1 replication in an activated cell predicts the existence of oscillatory dynamics in the synthesis of viral components. *BMC Genom.* **2014**, *15* (Suppl. 12), S1. [CrossRef] [PubMed]

23.   Reddy, B.; Yin, J. Quantitative intracellular kinetics of HIV type 1. *AIDS Res. Hum. Retroviruses* **1999**, *15*, 273–283. [CrossRef]

24.   Freed, E. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* **2015**, *13*, 484–496. [CrossRef] [PubMed]

25.   Heldt, F.S.; Frensing, T.; Reichl, U. Modeling the intracellular dynamics of influenza virus replication to understand the control of viral RNA synthesis. *J. Virol.* **2012**, *86*, 7806–7817. [CrossRef]

26.   Heldt, F.S.; Kupke, S.Y.; Dorl, S.; Reichl, U.; Frensing, T. Single-cell analysis and stochastic modelling unveil large cell-to-cell variability in influenza A virus infection. *Nat. Commun.* **2015**, *6*, 8938. [CrossRef] [PubMed]

27.   Gibson, M.A.; Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* **2000**, *104*, 1876–1889. [CrossRef]

28.   Marchetti, L.; Priami, C.; Thanh, V.H. *Simulation Algorithms for Computational Systems Biology*; Texts in Theoretical Computer Science. An EATCS Series; Springer: Berlin/Heidelberg, Germany, 2017. [CrossRef]

29.   Sazonov, I.; Kelbert, M.; Gravenor, M.B. A two-stage model for the SIR outbreak: Accounting for the discrete and stochastic nature of the epidemic at the initial contamination stage. *Math. Biosci.* **2011**, *234*, 108–117. [CrossRef]

30.   Safta, C.; Sargsyan, K.; Debusschere, B.; Najm, H.N. Hybrid discrete/continuum algorithms for stochastic reaction networks. *J. Comput. Phys.* **2015**, *281*, 177–198. [CrossRef]

31.   Sazonov, I.; Grebennikov, D.; Kelbert, M.; Bocharov, G. Modelling stochastic and deterministic behaviours in virus infection dynamics. *Math. Model. Nat. Phenom.* **2017**, *12*, 63–77. [CrossRef]

32.   Sazonov, I.; Grebennikov, D.; Kelbert, M.; Meyerhans, A.; Bocharov, G. Viral infection dynamics model based on a Markov process with time delay between cell infection and progeny production. *Mathematics* **2020**, *8*, 1207. [CrossRef]

33.   Mohammadi, P.; Desfarges, S.; Bartha, I.; Joos, B.; Zangger, N.; Muñoz, M.; Günthard, H.F.; Beerenwinkel, N.; Telenti, A.; Ciuffi, A. 24 h in the life of HIV-1 in a T cell line. *PLoS Pathog.* **2013**, *9*, e1003161. [CrossRef] [PubMed]

34.   Sobol, I.M. Sensitivity analysis for non-linear mathematical models. *Math. Model. Comput. Exp.* **1993**, *1*, 407–414.

35. Sobol, I. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **2001**, *55*, 271–280. [CrossRef]

36. Saltelli, A.; Tarantola, S.; Chan, K.P.S. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* **1999**, *41*, 39–56. [CrossRef]

37. Marino, S.; Hogue, I.B.; Ray, C.J.; Kirschner, D.E. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *J. Theor. Biol.* **2008**, *254*, 178–196. [CrossRef]

38. Degasperi, A.; Gilmore, S. Sensitivity analysis of stochastic models of bistable biochemical reactions. In *Formal Methods for Computational Systems Biology*; Lecture Notes in Computer Science; Bernardo, M., Degano, P., Zavattaro, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–20. [CrossRef]

39. Morris, M.D. Factorial sampling plans for preliminary computational experiments. *Technometrics* **1991**, *33*, 161–174. [CrossRef]

40. Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global Sensitivity Analysis. The Primer*; Google-Books-ID: WAssmt2vumgC; John Wiley & Sons: Hoboken, NJ, USA, 2008; [CrossRef]

41. Cheng, Z.; Hoffmann, A. A stochastic spatio-temporal (SST) model to study cell-to-cell variability in HIV-1 infection. *J. Theor. Biol.* **2016**, *395*, 87. [CrossRef]