A hybrid stochastic/deterministic model of intracellular HIV infection dynamics and estimation of viral production parameters

Samira Khalili and Antonios Armaou†, Senior Member, IEEE and AIChE

Abstract—This work is the initial step in development of an intracellular/extracellular model that describes the infection progression at the animal level, while having the required accuracy at the cellular level. Hence, a hybrid stochastic/deterministic intracellular model of HIV infection is developed to accurately model the events inside the infected cell. The model determines the probability of successful completion of intracellular infection cycle and estimates the number of produced virions per infected cell.

I. INTRODUCTION

Human Immunodeficiency Virus, HIV, is the causative agent of acquired immunodeficiency syndrome, AIDS, which has been one of the most destructive pandemics among humans and has been spread worldwide. There is no definitive cure for this disease and HAART (highly active antiretroviral therapy) is the current treatment for HIV infection.

In order to describe the dynamics of the disease, numerous mathematical models have been proposed in the open literature to capture different aspects of the disease progression. Previous work on modeling includes [22], in which an extracellular deterministic model was proposed to describe the dynamics of virus, T-cell, and infected T-cell of an average patient. Other research results include [32], [28], [21], [27] where various dynamic models of infection are proposed. In [23], an intracellular level model was developed to capture important steps of the intracellular HIV infection cycle. In [12], the dynamics of viral infection at both intracellular and extracellular infection level descriptions were investigated.

Such mathematical models [17], [26], [5] have been also employed to control disease progression and optimize medication schedules. In [33], a MPC based method for determining optimal treatment interruption of HAART was developed. A simulation model was developed in [30] to evaluate the clinical outcomes and life expectancy projections for three primary HIV infection treatment strategies. The majority of HIV models account for the effects of antiretroviral therapy by including drug efficacy as time-invariant constants [22], [7], [26], [6], [20]. Other studies such as [14], [29], [11] use empirical models to estimate the correlation between drug concentration and efficacy. However, no model has attempted to reflect the class-specific mechanisms by which the major types of antiretroviral drugs exert their inhibitory effect.

There are fewer intracellular models which describe the dynamics of viral components inside the infected host in comparison to extracellular models of virus-host interactions. The reason is the fact that the interaction dynamics of intracellular components are very complicated and quantitative information from experimental studies is not available for all parts of the cycle. These obstacles make developing intracellular mathematical models of HIV very difficult. In [23], a model for the entire intracellular infection cycle is developed. However, the model does not have the necessary level of detail. In other works such as [16], [18], [19], more precise models have been proposed, yet for a specific part of the cycle. The model constructed here includes the complete intracellular cycle and is based on reported physiological data collected through experiments.

In this work, a hybrid stochastic/deterministic model of intracellular events inside the infected cell is developed which considers the whole infection cycle from viral fusion to virion budding. Experimental data from various studies are used in calculations to create a more accurate description of the intracellular cycle. This model estimates the success rate of infection events, in other words the probability of virion production in a single infected cell. Also, the rate of production of HIV virions is calculated.

II. INFECTION CYCLE

Retroviruses is a large and diverse family of enveloped RNA viruses defined by their structure, composition, and replicative properties. Reverse transcription of the single stranded viral RNA to double stranded DNA and its integration into the genome of the host cell is the distinguishing feature of this family. HIV belongs to lentiviruses class of retroviruses. They have cone-shaped core, complex genome, remarkably complex interactions with the host, and a chronic course of disease [4].

Figure 1 shows the intracellular events of the infection cycle. HIV needs to attach to a host cell and enter its contents to the host to replicate. The process of viral entry into a target cell represents the first step in the viral infection cycle. It is characterized by a complex series of events that are initiated through the binding of the viral surface glycoproteins to specific receptor molecules on the host cell’s outer membrane. The primary receptor of the virus is the CD4 protein which is present on the surface of CD4+
T lymphocytes (which will be called T-cells afterwards), macrophage cells and some other cells [4]. T-cells, the main target of HIV infection, are specific type of white blood cells known as lymphocytes and are part of the immune system. Although HIV can bind to CD4 receptor on many cells, presence of a coreceptor is necessary for fusion of viral content into the host cell. CXCR4 and CCR5 are the two chemokine coreceptors of HIV [2].

Reverse transcription, the reverse flow of genetic information from viral RNA to DNA, is the hallmark of retroviral replication cycle. The reverse transcription protein (RT) uses viral RNA as a template to catalyze the polymerization of a double stranded DNA. This is a nonlinear process where up to three strands, a negative sense strand and two halves of the positive sense strand, may be transcribed at once. The next step in HIV growth cycle is integration. Soon after completion of viral DNA synthesis, while still in the cytoplasm, a viral enzyme called integrase, cleaves the viral DNA and provides the sites of attachment of the provirus to the host DNA. Then, the viral complex enters the nucleus of the host cell. Integration is an essential step in the life cycle of HIV. Integration contributes to viral replication in two important ways. First, since retroviral DNA is ordinarily unable to autonomously replicate, it depends on integration for stable maintenance in dividing cells. Once integrated, the provirus is replicated along with the host cell DNA and genetically transmitted as an integral element of the host genome. Unintegrated DNA is degradable and integration genetically transmitted as an integral element of the host genome. Unintegrated DNA is degradable and integration defines a turning point in the life cycle at which the virus can begin to multiply [4].

Translation of the viral genome results in different transcripts. After viral mRNA has left the nucleus, translation results in the synthesis of different polyproteins which will be cleaved into envelope and other proteins. Viral contents including two copies of the viral RNA genome gather at the surface of the cell and new viral particles are assembled at the plasma membrane and bud. Once formed, the new virus particle has to undergo a maturation step to become infectious. The protease enzyme becomes active and cleaves the gag-pol polyprotein into several smaller proteins. After this, the virus is able to infect new cells.

III. INTRACELLULAR MODEL DEVELOPMENT

Based on the explained mechanisms of the intracellular events (explained in section II), a stochastic model is developed that predicts the expected behavior of each infected cell as a function of time and the amount of available medication inside the cell. This molecular description of the events following the binding of the virus to the host cell provides the necessary information to develop a model of the intracellular HIV infection dynamics. Figure 2 shows the algorithm of the intracellular model. The model initiates when successful binding of the virus to the host occurs and the viral contents enter the host cytoplasm. The viral RNA is then reverse transcribed to create DNA. This process is explained in detail in our previous work [15]. In that paper, we developed a mechanistically informed model for the intracellular interaction of HIV-1 and nucleoside analogue reverse transcription inhibitors or NRTIs. The model could elucidate the efficacy profiles that might be expected for various intracellular conditions. We have linked efficacy with time-varying NRTI intracellular triphosphate concentration from physiologic data on plasma concentration maximums and half lives, such as that collected in [1] and [13].

As it is explained in [15], the probability of “successful inhibition” can be computed for a specific drug with known dosage as the sum of “dead-end formation” probability and “delay exceeding RNA lifespan” probability. Moreover, the probability of “no NRTI addition” and “delay not exceeding RNA lifespan” are available and the time distribution is also accessible for the latter case. The diamond labeled “reverse transcription” in figure 2 is actually a multi-step algorithm itself which is shown in figure 5 in [15]. Hence, when the “reverse transcription” step is executed in the simulation, a uniform random number is generated to determine the fate of the RNA among the three possible outcomes: successful inhibition, no NRTI addition, and time delay. If the first outcome is chosen, then the simulation ends in “Failure”. For the other two cases, the time needed to generate the complete genome is determined from the relevant time distributions (see figures 6 and 9 in [15]). When no drug is prescribed, the “no NRTI addition” case is always selected.

The produced DNA in cytoplasm is unstable and can be either degraded or transferred to the nucleus. The rate of degradation and transfer are reported in [23] and [16]. Then, based on Gillespie’s algorithm [10] a uniform random number is generated to determine if the DNA in cytoplasm is degraded (“failure”) or if it is transferred to the nucleus. Similarly, the fate of the DNA in the nucleus is determined which can be degradation (“failure”) or integration into the
host genome. After integration, the provirus becomes stable and deterministic description can be used.

Once integrated, the provirus is replicated along with the host cell DNA. During transcription, the genetic information contained within the DNA is copied to RNA. Still in the nucleus, the HIV transcripts undergo a series of post-transcriptional splicing events. This results in the generation of over 20 different mRNAs that can be grouped into 3 size classes [4]: unspliced or full length viral RNA (9kb) for synthesis of gag and gag-pol polyproteins, singly spliced viral RNA (4kb) for synthesis of Vif, Vpr, Vpu, and Env proteins, and fully (or doubly) spliced viral RNA (2kb) which encodes the regulatory proteins such as Tat, Rev, and Nef. The viral RNAs are produced in a sequential manner. The full length RNA is the initial product of transcription which is then translated into singly spliced which itself is spliced to generate fully spliced RNA. Among all the viral proteins, a few have significant role in the transcription of other proteins and the formation of new virus particles. These proteins are Tat, Rev, and Gag which are explicitly included in the model.

Rev, the regulator of viral gene expression, binds to the Rev response element (RRE). Rev’s function is to facilitate the export of full length and singly spliced viral RNA from the nucleus to the cytoplasm. RNA transcripts that are not fully spliced are generally retained inside the nucleus and only fully spliced RNA can be exported from nucleus to the cytoplasm where it is synthesized to Rev and Tat. Rev can be imported back to the nucleus where several copies of it associate on the RRE with the nuclear export protein. This association makes the export of unspliced and singly spliced RNAs possible. Rev levels are regulated by a feedback mechanism. The high level of Rev means high export levels of unspliced and singly spliced RNA. This means that the amount of RNA left for fully splicing is reduced and Rev expression is subsequently decreased.

Tat, the transactivator of transcription, is essential for HIV-1 replication. Tat binds to a short-stem loop structure, the transactivation response element (TAR). Binding of Tat to TAR activates transcription of HIV up to a thousand fold. Tat promotes completion of initiated transcriptional activity. In the absence of Tat only short transcripts (less than 100 nucleotides) are produced [4]. Figure 3 shows unspliced, singly spliced, fully spliced RNAs and regulatory proteins Rev and Tat, and structural proteins such as Gag.

The full length RNA in the nucleus is translated to several polyproteins, among them is Gag, the structural protein of HIV. Gag alone can form virus like particles [9]. It is determined that an average HIV particle contains ~ 5000 Gag proteins [3]. Other proteins as well as two copies of viral genome are also packed in the particle. Another important role of Gag is mediating interactions between Gag itself and the viral genome. In other words, Gag “recruits” viral RNA during assembly process [8]. Once the particle is formed, the new virus has to undergo a maturation step to become infectious. The protease enzyme becomes active and cleaves Gag and other polyproteins into several smaller proteins.

In [16], an intracellular model in the form of ODEs is proposed which focuses on the events after integration of provirus and describes the production and dynamics of full length, singly spliced, and fully spliced RNAs, Rev-bounded RNAs, and regulatory proteins Rev and Tat, in both nucleus and cytoplasm. This model is used in [31] where Gag is also considered in the model which makes the estimation of virion production possible. An ODE model is constructed based on
these two models. This model represents the “Transcription, Translation, Budding” block in figure 2.

IV. RESULTS AND DISCUSSIONS

By solving the ODE model of “Transcription, Translation, Budding” block, the number of virus particles budding from the infected cell can be calculated as a function of time. Figure 4 shows the number of produced virions after the integration of viral DNA.

![Fig. 4. Number of produced virions as a function of post integration time](image)

The life span of an infected cell is assumed to be two days [21]. The time needed for pre-integration steps should be considered along with Figure 4 to compute the number of produced virions. Figure 5 shows the distribution of pre-integration time for cases not ended in “failure”. The mean of the distribution is 12.04 hours with standard deviation of 0.09.

![Fig. 5. Distribution of pre-integration events time (no drug)](image)

For each case which successfully reaches the “Transcription, Translation, Budding” step, the pre-integration time is deducted from 48 hours (“available” time for virion production) to calculate the remaining time for post-integration events before the infected cell is cleared. Then, the corresponding number of virions is identified from Figure 4. The average number of produced virions is 313 in two days with standard deviation of 1.6. In other words, the full cycle completion time is 48 hours for all cases, but the number of virions varies for each run.

Since the infection has the potential to fail at three steps (the pink diamond blocks in Figure 2) during the cycle, the success rate or probability can be determined through a large number of simulations: the success rate will be the number of simulations concluded at “success” block divided by the total number of simulations. Hence, the outcome of the intracellular model is two important parameters: success rate and the rate of virion production. These parameters depend on the chosen treatment (types of drugs, dosage, etc) and are inputs to the multiscale model. The results shown in Figure 5 and the calculated value of 313 virions in two days are based on one million simulations with no treatment considered. The success rate and the rate of virion production are reported in Table I for no treatment and also under treatment cases. These values depend on drug dosage which is shown by \( x = \frac{C_{\text{max}}}{IC_{50}} \). Success rate of the intracellular cycle initiating by virus attachment and concluding by virion production is shown by \( \epsilon_c \).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>success rate ( \epsilon_c )%</th>
<th>number of virions</th>
<th>production rate ( n ) (virions/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No drug</td>
<td>51.18</td>
<td>313</td>
<td>156</td>
</tr>
<tr>
<td>AZT, ( x=1 )</td>
<td>30.97</td>
<td>234</td>
<td>117</td>
</tr>
<tr>
<td>3TC, ( x=0.5 )</td>
<td>30.82</td>
<td>292</td>
<td>146</td>
</tr>
<tr>
<td>3TC, ( x=1 )</td>
<td>18.56</td>
<td>271</td>
<td>135</td>
</tr>
<tr>
<td>AZT, ( x=1 ) &amp; 3TC, ( x=0.5 )</td>
<td>22.06</td>
<td>233</td>
<td>116</td>
</tr>
<tr>
<td>AZT, ( x=1 ) &amp; 3TC, ( x=1 )</td>
<td>13.10</td>
<td>223</td>
<td>113</td>
</tr>
</tbody>
</table>

It is interesting to investigate the reverse transcription time distribution under treatment. In figure 6 the time distribution under treatment with 3TC at \( x=1 \) and \( x=0.5 \) is shown along with the no drug distribution. It is important to emphasize here that only results from simulations in “success” are shown in the distribution. The black trajectory shows the no drug distribution and it can be observed that under treatment with 3TC at \( x=0.5 \), the distribution divides in two parts. The first peak represents the cases where no NRTI attachment has happened, whereas the wider peak shows the time delay caused by NRTI attachment and dissociation. It is observable that at higher concentration, 3TC at \( x=1 \), the first peak becomes smaller and the second peak shifts to larger times. This is obviously due to the fact that at higher concentrations, the probability of NRTI addition increases.

In figure 7 distributions under treatment with AZT at \( x=1 \), 3TC at \( x=1 \), and both AZT and 3TC at \( x=1 \) are shown. Since AZT has a higher affinity than 3TC (for detailed discussion

1525
The drawback of widely used HIV models is the phenomenological parameters which are determined by fitting available data to the proposed model. This goal is achieved by initially building a stochastic intracellular model. The efficacy model of our previous work [15] is part of this model where the fate of an infected cell is determined based on success or failure of intracellular events and the medication schedule. Hence, we have computed the success rate of infection events at the intracellular level as well as the number of produced virions. These parameters will be eventually incorporated in a multiscale model of HIV dynamics.

REFERENCES