Research

Formal reasoning on qualitative models of coinfection of HIV and Tuberculosis and HAART therapy

Anil Sorathiya*[1], Andrea Bracciali[2] and Pietro Liò[1]

Addresses: [1]Computer Laboratory, Cambridge University, William Gates Building,15 JJ Thomson Avenue, Cambridge CB3 0FD, UK and [2]Computer Science Department, University of Pisa, Largo Bruno Pontecorvo, 3, Pisa 56127, ITALY

E-mail: Anil Sorathiya* - anilkumar.sorathiya@cl.cam.ac.uk; Andrea Bracciali - braccia@di.unipi.it; Pietro Liò - pietro.lio@cl.cam.ac.uk
*Corresponding author

Abstract

Background: Several diseases, many of which nowadays pandemic, consist of multifactorial pathologies. Paradigmatic examples come from the immune response to pathogens, in which cases the effects of different infections combine together, yielding complex mutual feedback, often a positive one that boosts infection progression in a scenario that can easily become lethal. HIV is one such infection, which weakens the immune system favouring the insurgence of opportunistic infections, amongst which Tuberculosis (TB). The treatment with antiretroviral therapies has shown effective in reducing mortality.

An in-depth understanding of complex systems, like the one consisting of HIV, TB and related therapies, is an open great challenge, on the boundaries of bioinformatics, computational and systems biology.

Results: We present a simplified formalisation of the highly dynamic system consisting of HIV, TB and related therapies, at the cellular level. The progression of the disease (AIDS) depends hence on interactions between viruses, cells, chemokines, the high mutation rate of viruses, the immune response of individuals and the interaction between drugs and infection dynamics.

We first discuss a deterministic model of dual infection (HIV and TB) which is able to capture the long-term dynamics of CD4 T cells, viruses and Tumour Necrosis Factor (TNF). We contrast this model with a stochastic approach which captures intrinsic fluctuations of the biological processes. Furthermore, we also integrate automated reasoning techniques, i.e. probabilistic model checking, in our formal analysis. Beyond numerical simulations, model checking allows general properties (effectiveness of anti-HIV therapies) to be verified against the models by means of an automated procedure. Our work stresses the growing importance and flexibility of model checking techniques in bioinformatics.

In this paper we *i)* describe HIV as a complex case of infectious diseases; *ii)* provide a number of different formal descriptions that suitably account for aspects of interests; *iii)* suggest that the

integration of different models together with automated reasoning techniques can improve the understanding of infections and therapies through formal analysis methodologies.

Conclusion: We argue that the described methodology suitably supports the study of viral infections in a formal, automated and expressive manner. We envisage a long-term contribution of this kind of approaches to clinical Bioinformatics and Translational Medicine.

Background

Human diseases result from abnormalities in an extremely complex system of molecular processes that are often caused by viral or bacterial infections. In these pathological processes, virtually no molecular entity acts in isolation and complexity is caused by the vast amount of dependencies between molecular and phenotype features. The key player of the human survival is the immune system which is a complex system that can be described as a large network of dynamical agents (cells and signalling molecules). A great challenge for contemporary molecular medicine is the modelling, description and ultimately the comprehension of the multistep and multiscale nature of the immune response to pathogens. An even greater challenge is when viral and bacterial infections occur together in the same patient. TB, which is caused by Mycobacterium tuberculosis, is the most frequent co-infection in patients infected by the human immunodeficiency virus type 1 (HIV-1). TB makes more complicated the development of effective therapies.

Efforts for gaining further insights into the pathological mechanisms and novel therapeutic targets benefit from the integration of genomic, proteomic, metabolomic and environmental information. In this work we follow an integrated approach that relies on several sources, such as phylogeny information, traditionally explored within Bioinformatics, differential equations and stochastic modelling, both largely used in Systems Biology, and a formal reasoning technique, viz. Model Checking, developed in Theoretical Computer Science to assess properties of complex computational systems. Furthermore, the Bioinformatics approach in the study of viral dynamics has also focused on identifying variable regions in genomes and pathogenic islands in bacteria [1]. These integrated frameworks appear to be effective for the research in the biomedical problems.

Mathematical and computer science approaches have shown to be more effective in dissecting the network connectivity of cellular circuits and the corresponding dynamical characteristics. The mathematical description of the variation of biomolecular concentrations as a set of ODEs offers quantitative basis for predicting the behaviour and evolution of the system and for testing non-linearities. An alternative approach is to use stochastic simulation via the Gillespie algorithm which provides an exact algorithmic solution of a set of reactions and a meaningful way to consider the noise [2]. The use of mathematical models in immunology has been very successful and has represented an insightful and essential complement to in vivo and in vitro experimental design and interpretation. Kinetic Model suggests that HIV-1 in vivo is continuous and highly productive and that leads CD4 T cells count low [3]. Nowak and May has proposed a simple model on HIV dynamics by considering population of non-infected T cells, infected T cells and viral population [4]. Perelson has proposed mathematical models on interaction of HIV virus and T cells dynamic by considering one more variable (un-infectious viruses) [5]. The HIV quasi-species models have been inspired from the molecular quasi-species model in chemistry [6,7]. Indeed mathematical models of HIV dynamics have proven valuable in understanding the mechanisms of many of the observed features of the progression of the HIV infection, see for example [3,6-13]. In this paper we model coinfection of HIV and TB and the effect of HAART therapy on the model dynamics.

Recently, the observation that biological systems often exhibit interactive and concurrent behavior, similarly to computational concurrent systems, has led to the adoption of formal methods originally developed for the description and analysis of complex software systems in computer science. This abstraction "cell as computation", similar to the "DNA as string" and "protein as labeled graph" abstractions which have originated bioinformatics, has inspired the adoption of model checking methodologies to validate biological complex systems [14]. The growing success of model checking (see for instance [15-18]) relies in the specification of a biological property of interest which is expressed in a formal language, typically a formula of a suitable logic, and its verification is carried out by a fully automated procedure that returns either a positive response or a counter example [19].

The aim of this paper is to pipeline bioinformatics and quantitative models of infectious processes and anti-HIV therapies, and then show how model checking

techniques can contribute to the interpretation of the *in-silico* results obtained from quantitative models of infectious diseases dynamics. Given that viral infections and the functioning of therapies often present stochastic aspects, we introduce a succinct but descriptive stochastic model of HIV infection associated with TB opportunistic infection. Then we extend it by modelling (the effects of) HAART anti-HIV therapy. These models have then been implemented in PRISM, a state of the art probabilistic model checker supporting a logical language [20]. We illustrate, by means of two simple properties, the flavour of the verification made possible by these techniques and how this can contribute to a precise assessment of the information conveyed by the mathematical models.

### The complexity of HIV infection and anti-HIV therapies

Human immunodeficiency virus type 1 (HIV-1) infection is characterised by the progressive loss of CD4 T cells. Anti-HIV therapies act to eradicate or lower the concentration of the virus from the body and replenish the CD4 T cells reservoir. Infection by most strains of HIV requires interaction with CD4 T cells and a chemokine receptor, either CXCR4 or CCR5. Viral strains often use CCR5 during early stages of HIV-1 and then switches to CXCR4 to enter into the cells. This switch emerges in more than 50% of patients [21,22] and it has been linked with progression to AIDS because of an increased virulence through the formation of cell syncytia and the stimulation of the cellular factor called Tumour Necrosis Factor (TNF) which inhibits the replication of R5 HIV strains while has no effect on X4 HIV [23,24].

It takes on an average 10 years to get infected by AIDS after having HIV infection. Some patients died within 2 years after getting infected by HIV, while others remained free of AIDS for more than 15 years. The within-patient evolutionary process of viral sequence mutations during HIV infection has suggested improvements in anti-HIV therapies. Anti-HIV drugs are most effective when taken in a combination of three or more at the same time. This is called combination therapy or HAART (Highly Active Antiretroviral Therapy). Physicians recommend starting the therapies if you are ill because of HIV, or if your CD4 T cells count if low (below 200 cells per microL). HAART combinations usually include two drugs that are nucleoside analogues, and one protease inhibitor. The nucleoside analogues drugs result in targeting the viral reverse transcriptase which codes the viral RNA into the DNA that can be integrated into human cells, so transforming the cell into a factory for building blocks of the virus. The protease inhibitor acts as preventing an infected cell from producing new infectious virus particles.

Bioinformatic links between HIV and TB

Chemokines provide the key link between HIV and TB infection. Resistance to HIV infection has been found to be related to the following mutations. Delta32 CCR5, 190G CCR2 and 744A CX3CR1 and CCL3L1 [25]. The chemokine receptor CXCR3 can exhibit weak coreceptor function for several human immunodeficiency virus, both HIV-1 and HIV-2 strains and clinical isolates [26]. Gene expression data analysis of HIV infected macrophages showed large changes in beta-chemokines and RANTES (CCL5) [27]. The TB infection is known to produce larger effects than HIV infection on the chemokine networks. TB is associated with excess monocyte chemoattractant protein (MCP)-1 and Tumour Necrosis Factor (TNF)-alpha activity in situ, both are implicated in transcriptional activation of HIV-1 [28]. It is also important to mention significant elevation of the chemokines CCL3, CCL4 and CCL8 [29]. Gene expression data analysis of TB infected macrophages versus uninfected, showed upregulation of the following genes. interleukin-1 beta and interleukin-8, macrophage inflammatory protein-1 alpha, growth-related oncogene-beta, epithelial cell-derived neutrophil-activating peptide-78, macrophage-derived chemokine [30]. The most important mechanism through which TB enhances HIV-1 replication and the progression to AIDS in dually infected patients is the augmentation in expression of TNF-alpha and the HIV-1 noninhibitory beta-chemokine (MCP-1), low presence of HIV-1 inhibitory beta-chemokines (MIP-1 alpha, MIP-1 beta, and RANTES). We have collected a set of human aminoacid sequences of all the chemokine receptors known to chemokines involved in HIV and tuberculosis and a set of control chemokine receptors. We have built a phylogenetic tree that describes the statistical relationship between those chemokine receptors (see Figure 1). The tree was generated using maximum likelihood on an alignment of the sequences of the external regions of the protein loops. In violet the chemokines receptors disrupted by both TB and HIV; in red those disrupted by HIV. Note that the switch between CCR5 and CXCR4 may involve other chemokine receptors. This may suggest to concentrate some efforts on therapies by intervening by blocking other chemokine receptors disfunction.

The disruption caused by the dual infection (HIV and TB) focuses on RANTES which blocks CCR5 and whose expression is upregulated by TNF. Therefore we incorporated TNF in our deterministic model to predict the potential effect of HAART on coinfection of HIV and TB. That evidence would explain the increase in CXCR4 usage with the increase of TNF concentration [31]. The relationship between HIV and TB seems to be very profound since quasispecies of both HIV and TB influence each other differently [32,33]. Our
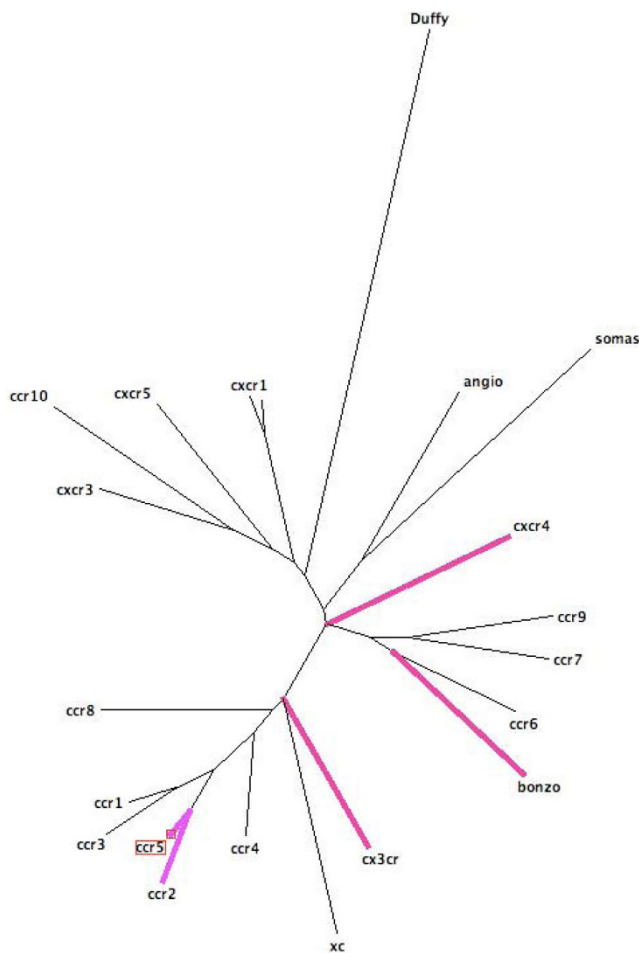
Figure 1
The maximum likelihood phylogeny under the JTT model of evolution for a set of chemokine receptors amino acid sequences.

phylogenetic results suggest that both diseases act synergically and not simply additively to alter the chemokine network.

Results and Discussion

Although differential equation models have long been used for the immune system and viral infection modelling, they focus on the average behavior of large populations of perfectly mixed, identical individuals. An improved realism is perhaps provided by stochastic simulations, which however are computationally intensive. Given that differential equations and stochastic descriptions have important pros and cons and a good degree of complementarity, a framework based on the implementation of both approaches, although time and resources expensive, appears advisable. Stochastic models, then, support the well developed probabilistic model checking analysis. We have extended a model firstly

presented in [7]. We have incorporated the Cytotoxic T Lymphocytes (CTLs) response and the dynamics of TB opportunistic infection so as to analyse the co-infection of HIV strains and another disease associated with it.

Furthermore, we have introduced an abstract representation of the HAART therapy treatment by altering the model's parameters that rule the dynamics of our model according to the known effects of the treatment. We are presenting results of the analysis of the HIV and the opportunistic TB infection dynamics and an associated therapy through differential equations, stochastic modelling and formal reasoning techniques.

HIV and the opportunistic TB infection
Experiments carried out by means the deterministic model are reported in Figure 2. Details about the construction of the model are reported in the Methods section. For the sake of space we do not show the sensitivity analysis of the deterministic model. Experiments with the model are able to faithfully reproduce well known curves of the virus and CD4 T cells dynamics in absence of TB. The introduction of TB provides a novel description of the dynamics occurring with this infection where viral load increases suddenly in the last stage of HIV-1 (AIDS) because of opportunistic diseases (TB) (see [33]). In particular, Figure 2(a) shows the behaviour of CD4 T cells during the switch between R5 and X4 strains. The general dynamics of CD4 T cells and viral load can be observed. In the latest stage of HIV-1 infection (opportunistic diseases stage) a sudden increase of the viral load can be observed, which lead to AIDS. The increase of TNF against the viral load time course is reported in Figure 2(b). This TNF dynamics provide a clear explanation of the quick growth of viruses when TB infection occurs. TB and X4 strain of HIV stimulate TNF production which, in turn, help them to progress at faster rate.

Modelling HAART therapy against HIV and TB
The model of HIV and TB co-infection can easily be extended to embrace the effect of a common anti-HIV treatment such as the HAART. Drug therapies such as HAART and Maraviroc, although with different mechanisms, result in decreasing the number of viruses and the death of infected cells. Exploiting the expressivity of the treatment, we model HAART in terms of its effects. For the sake of simplicity this can be done by changing the virus replication rate (π, see Methods), reflecting a decreased replication rate and budding from the infected cell. Simulation results are reported in Figure 3. Plots (a) and (b) show the effect of the HAART therapy delivered before and after the appearance of the X4 strain, respectively and before (c) and after (d) the TB infection, respectively. Clearly, both the appearance of X4 and the
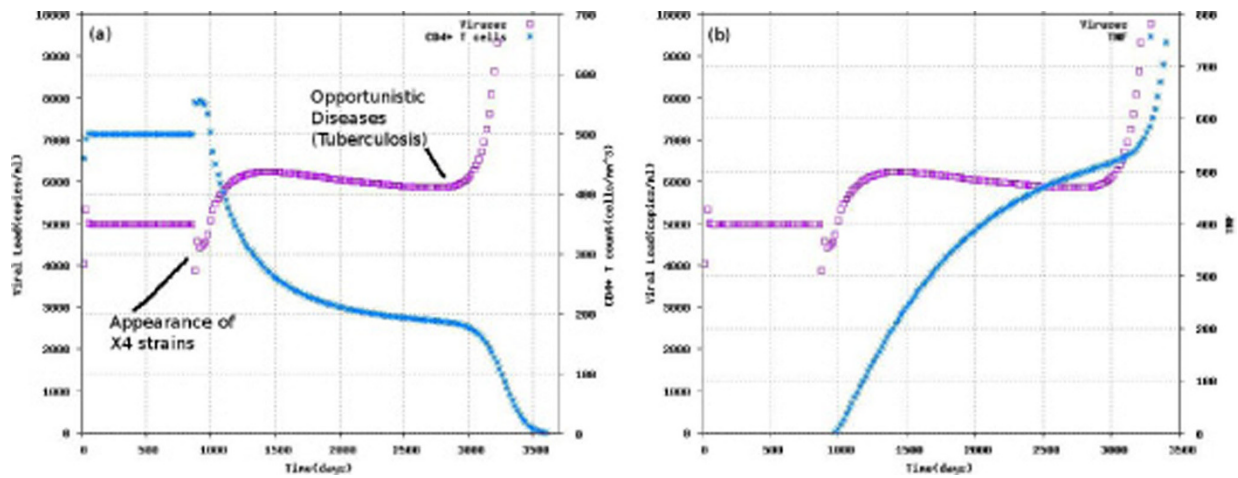
Figure 2
Viral load, CD4 T and TNF over the course of time. (a) Time evolution of viral load and CD4 where mutation of V5 leads to V4 at around day t = 900 and opportunistic diseases (TB) appears at around day t = 2900 (b) dynamics of viral load and TNF over time.
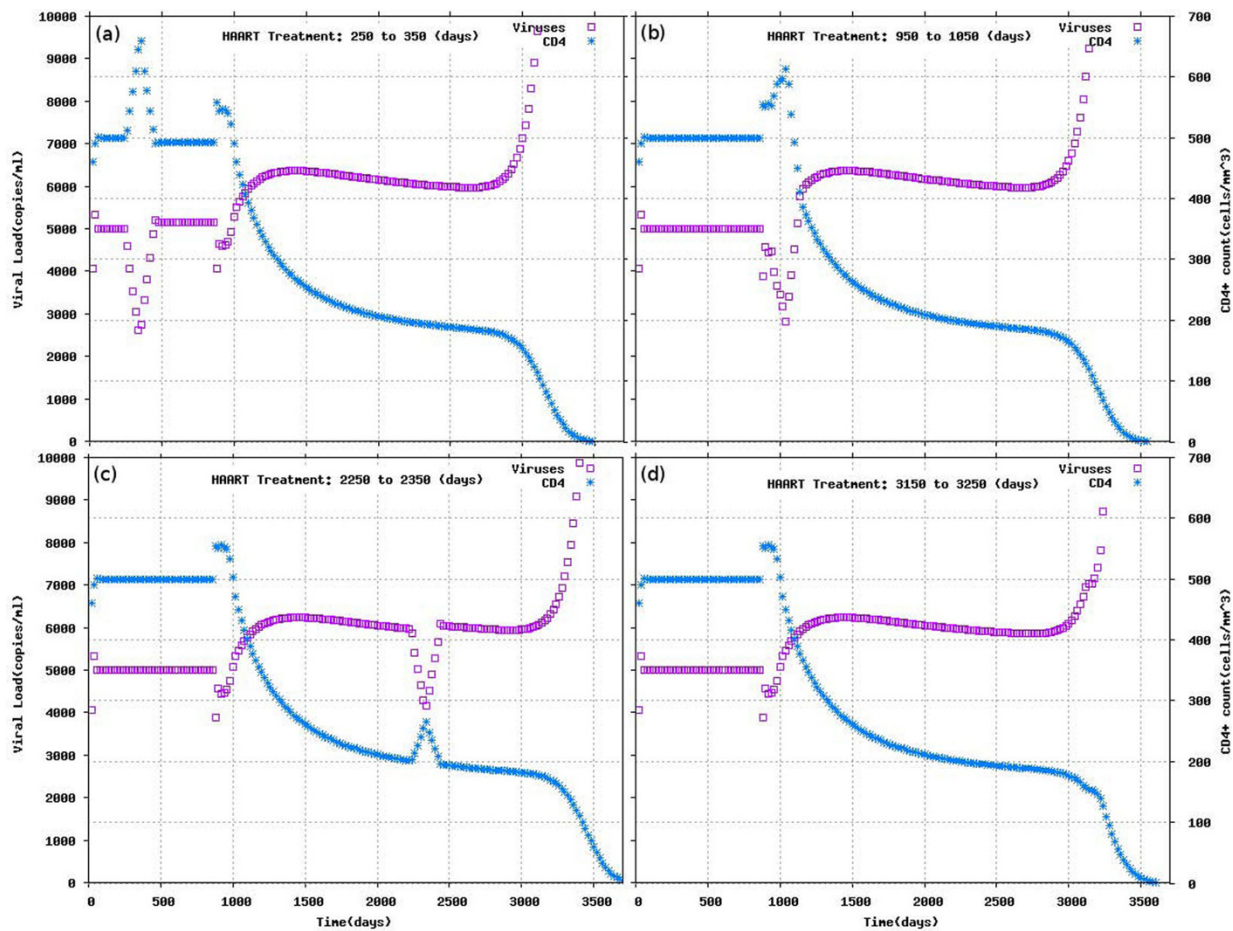
Figure 3
Experiments with HAART therapy. Therapy start from (a) day t = 250 to day t = 350 (b) day t = 950 to day t = 1050 (c) day t = 2250 to day t = 2350 and (d) day t = 3150 to day t = 3250.

occurrence of TB infection accelerate the decrease of CD4 T cells which the HAART is able to stop but not to completely reverse. An early start promulgated HAART treatment will change the probability of arising of X4 (this results was obtained by Sguanci et al). Here we show that short HAART treatments have small effect if administrated when CD4 T cells count is above 200 and they have even smaller effect when CD4 T cells counts are below 200.

A stochastic description of HAART, HIV and TB dynamics
Many natural and biological phenomena are intrinsically stochastic and discrete, and they can not always be properly described by means of a deterministic and continuous description. For instance, systemic emergent properties can be sensitive to the local presence of minimal (integer) quantities of molecules [34]. Roughly speaking, a stochastic model associates a probability to each state transition of the modelled system, expressed as a rate associated to the transition. Associated probability distributions are typically memoryless, and hence the overall system behaviour can be interpreted as a Markov Chain. Often, models are used to simulate possible evolutions of the system from initial conditions, [35].

Starting from the deterministic model of HIV, TB and HAART, we have determined a corresponding stochastic model. This has been done via standard transformation from deterministic rates into stochastic ones, according to a fixed reference volume of the model (see, e.g., [36]).

Simulation results are reported in Figure 4 about the dynamics observed for a specific possible evolution of

the system (chosen according to the probability distribution determined by the model). From Figure 4 (left) it emerges that CD4 T cells, representative of the immune system health, go quickly down in the massive presence of the X4 viral strain. As soon as CD4 T is below a threshold limit, 20 units on the scale of this model, TB develops and the viral load saturates. This reproduces, on a small scale, the general time course at individual level, like the one reported in Figure 2. In Figure 4 (right) a simulation under the same initial conditions is shown. In this case, the HAART treatment has been administered starting at a quite early time point (after few tenth of the initial second) and then stopped (at about 2.5 seconds). It can be observed that, even if HIV progression weakens the immune system, it never goes below the threshold, but rather the viral load seems quite well controlled. In the short interval, HAART prevented the development of TB once started before its insurgence. However the viral load upraises after the interruption of the therapy. The stochastic modeling approach shows important advantages, such as the possibility of pipelining with the probabilistic model checking and some limits in the computational intensive requirement. Work in progress will make use of grid computing facilities (EGEE and Cambridge) and cloud facilities.

Assessing HAART therapy against HIV and TB infection
Observing the quantitative results of the stochastic simulations in Figure 4, one can have an idea about the fact that the HAART therapy is somehow effective. However, we discuss now an example on how it is possible to move from a correct and informative, but somehow empirical, interpretation of data, as the idea
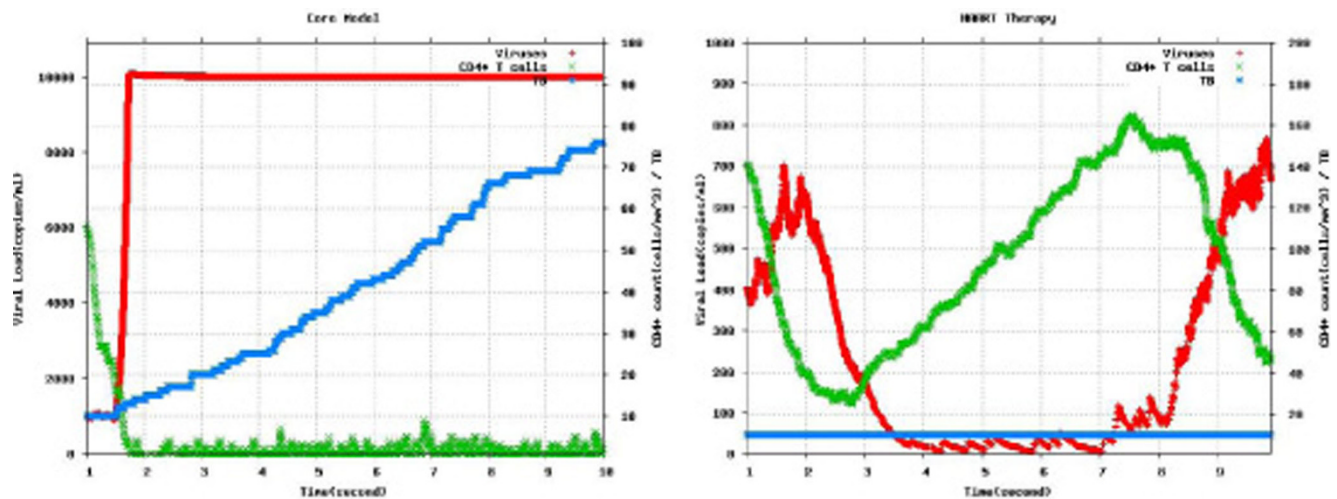
Figure 4
A limited fragment of the time course of Viruses, CD4 T cells and TB bacteria in the case of (a) no treatment and (b) treatment with HAART.

above, to a more precise and formal approach by specifying properties of interest as logical formulae. This has the advantage of requiring a definition of all the aspects of interest, of referring to a formal and unambiguous semantics, and of having an automated procedure for verifying formulae (as the one we adopted [20]). Differently, observations rely on the capability of properly interpreting the graphs, which sometimes may appear obscure and difficult to interpret. More details about the methodology can be found in the Methods section.

Therapy effects
We focus on two properties that can give a measure of the infection progression.

i) we consider the number of infected cells leading to virus replication as a measure of the spread of infection. We are interested in measuring how often this can occur. This can be expressed by a formula like

$$R\{"i\_to\_v"\} = ? \quad [C <= 5.0]$$

where $R\{"i\_to\_v"\}$ represents the number of occurrences of an event labelled with i_to_v, i.e. the transition from an infected CD4 T cell to virus replication. The question mark says we are interested in determining the number of such possible occurrences in the model dynamics. $C < 5.0$ stands for the time interval;

ii) we consider the amount of healthy CD4 T cells, represented by variable tc, as a measure of resistance to the virus attack. We are interested in the probability that tc is less than the threshold 20. Here ? requires the computation of a probability and $F [...] tc <= 20$ is a logic term expressing that the event $tc <= 20$ occurs within the interval [...]:

$$P = ? \quad [F[0.0, 5.0] \quad tc <= 20]$$

Experimental results
The analysis has been performed using the PRISM model checker (see Methods section). Properties can be validated either by constructing the complete Continuous-time Markov Chain relative to the given model model or by approximating verification through sampling a certain number of possible evolutions. The former can result costly or unfeasible for large models. Adopting the latter we obtained the results reported in Table 1.

Automated verification yields quantitative measures of the investigated properties. While the number of viral replications due to CD4 T cell infection is comparable (but without HAART many other infected cells contribute to virus replication), the probability of a failure

Table 1: Quantitative results from the automated verification of the effects of HAART therapy. While the number of viral replications due to CD4 T cell infection is comparable with and without HAART treatment for the time interval considered, a much stronger probability of a failure of the immune system, *close to twice as much as*, is observed without the HAART treatment

	HIV+TB	HIV+TB+HAART
$R\{"i\_to\_v"\} = ? [C <= 9.0]$	250	269
$P = ? [F[5.0, 8.0]tc <= 20]$	0.885	0.429

of the immune system is much stronger without HAART treatment, *closely to twice as much as*.

Conclusion
We have illustrated the potential benefits of formal methods and quantitative models when applied to the study of viral infections and therapy assessment within a computational bioinformatics approach. We have done this by presenting experiments on a proof-of-concept scenario regarding HIV infection and the relative TB opportunistic infection. We have adopted an integrated approach, combining deterministic and stochastic techniques and illustrating how properties of interests for the study of viral infections can be formalised in a general purpose logic, as the one supported by PRISM. Our work stresses the growing importance and flexibility of model checking techniques in bioinformatics. Noteworthily, the verification of these properties can precisely characterise the numerical results of simulations, and this can be helpful in comparing and assessing different antiviral therapies. In conclusion, the modelling of HIV infection has two important linked benefits. *i)* it has matured to the stage of allowing us to combine bioinformatics, computational modelling and formal reasoning techniques, as model checking, i.e. it provides a solid bridge between biological systems and the computational objects used to describe them; *ii)* these approaches together capture information that can be valuable in therapy validation suggesting the possibility of moving in the future to realistic cases, i.e. translational medicine and clinical bioinformatics.

Methods
In this section we provide further details about the construction of the used models and analysis methodologies.

A deterministic model of HIV strains and TB time evolution
Our work is based on a deterministic model of HIV-1 dynamics, firstly appeared in [7], which takes into account the models developed by Perelson and his followers [3,10,13,37]. These models are well presented

and take specific biological reality into account. Our initial model has been extended here by adding two more variables, Cytotoxic T Lymphocytes (CTLs) cells and TB, to capture dynamics of co-infection of HIV and TB.

As standard, we describe the variations of the quantities of the modelled entities as a set of differential equations. We start considering a pool of immature CD4 T cells, represented by the variable U, see equation (1). These cells are continuously produced by the thymus at a rate N_U, and evolve into differentiated, uninfected T-cells, T at a rate δ^U. Also, the TNF (F) contributes to the clearance of naive T-cells, via δ_F^U, and naive T cells are produced at fix rate k_Z due to CTLs (Z) response. T cells (T) are described by considering their different strains (T_i), which we do not detail here (see [7]). Beyond being produced, they can become infected (I_k) by interacting with the virus strains V_k at rate β_k, or die at rate δ^T, (equation (5)). Note that the infection parameter β is not constant over time, but depends on the distribution of the viral strains R5 and X4 [9]. Infected T cells are cleared out at a fixed rate, δ^I, and also due to the action of CTLs with rate δ_I^Z, (equation (2)). Equation (6) describes the budding of viruses, i.e. infected cells produce new viruses at rate π, and the fact that virus particles may be nonviable or being cleared out at rate c by immunoglobulin binding and subsequent engulfments by the macrophages. Next equation (3) describes latent TB (B) in the blood which propagates with the rate α when T cells goes below a given threshold representing the efficacy of the immune system. The equation (7) describes the TNF (F) dependence on X4 strains and on the presence of TB in the blood and the fact that the efficacy of such a factor naturally decay in time. Finally, the response of Cytotoxic T Lymphocyte cells Z is as in equation (4): CTLs response depends on number of infected cells (I) and is cleared out at fixed rate b. This model is general enough to be used as a framework for fitting real data and simulating superinfection and co-infection patterns.

$$\frac{dU}{dt} = N_U - \delta^U U - \delta_F^U UF + k_Z IZ \qquad (1)$$

$$\frac{dI_k}{dt} = \left(\sum_{k'} \mu_{kk'}\beta_{k'}V_{k'}\right)\left(\sum_i T_i\right) - \delta^I I - \delta_I^Z IZ \qquad (2)$$

$$\frac{dB}{dt} = \alpha B - \delta_Z^B BZ \qquad (3)$$

$$\frac{dZ}{dt} = \delta_I^Z IZ - bZ \qquad (4)$$

$$\frac{dT_i}{dt} = \delta^U U - (\sum_k \beta_k V_k)T_i - \delta^T T_i \qquad (5)$$

$$\frac{dV_k}{dt} = \pi I_k - cV_k \qquad (6)$$

$$\frac{dF}{dt} = K_F \sum_{k \in X4} V_k + K_B B - \delta^F F \qquad (7)$$

The stochastic model

The stochastic model has been derived from the deterministic one. As a simplification we consider only two viral strains: R5 viral strain ($V5$) and X4 viral strain ($V4$), and correspondingly only two kinds of infected cells (I4, I5 and I* for the union of the two). Analogously, quasispecies dynamics revert to $V5$ strains that transform into $V4$ strains (at rate β_{k54}). These two strains are the ones we focus on when observing the time course of infections. As for the deterministic model, we have included sufficient details of interaction among the species to express the properties of interest.

As mentioned, we needed to translate some parameters from deterministic to stochastic ones, accordingly to the reference volume of interest. Other parameters, particularly due to lack of existing data in literature, have been approximated by tuning the model on known macroscopic behaviour. We considered a volume of reaction large enough to contain a statistical reliable number of agents (viruses and cells) and values of extensive quantities are scaled according with the reaction volume considered.

We employ a *population-based* approach where the number of each type of species or cells are modelled, rather than the state of each individual component. Furthermore, also according to the PRISM modeling language we used, we rewrote the equations of the deterministic model in terms of a "reaction-centric" description. dynamics is represented by a set of reaction rules stating the involved entities and the results of interaction. These are reported in Table 2. Reaction (02) stands for the production of T cells from naive T cells, while (04) and (05) represent viral infection. Stochastic parameters are on the right part of the Table. Interestingly, parameters α and δ_Z^B vary to reflect the emergent behaviour of TB infection against the efficiency of the immune system. when below to a given threshold, the exponential growth of the HIV virus load appears, caused by the spreading of the infection to other cells and abstractly modelled here by such a set of ad-hoc rules. Initial population of cells in the model is: $U = 30$, $T = 50$, $I = 30$, $Z = 30$, $F = 0$, $V_4 = 50$ -V_5, $V_5 = 50$, $B = 10$, and the immune deficiency threshold is set to 20.

Table 2: Interactions in the stochastic model. The values of the parameters are from literature referred (for sake of clarity we skip the dimensionality which is the standard reported in literature), see also [42-44], or obtained by tuning of the model

(01)	$DU \xrightarrow{N_U} dU + U$	$N_U = 100$
(02)	$U \xrightarrow{\delta^U} T$	$\delta^U = 0.1$
(03)	$U + F \xrightarrow{\delta_F^U} F$	$\delta_F^U = 0.00001$
(04)	$T + V5 \xrightarrow{\beta_{k5}} I5$	$\beta_{k5} = 0.0025$
(05)	$T + V4 \xrightarrow{\beta_{k4}} I4$	$\beta_{k4} = 0.0025$
(06)	$V5 \xrightarrow{\beta_{k54}} V4$	$\beta_{k54} = 0.025$
(07)	$T \xrightarrow{\delta^T} 0$	$\delta_T = 0.1$
(08)	$I4 \xrightarrow{\delta^I} 0$	$\delta_I = 0.8$
(09)	$I5 \xrightarrow{\delta^I} 0$	$\delta_I = 0.8$
(10)	$I^* + Z \xrightarrow{k_Z} U + I^* + Z$	$k_Z = 0.01$
(11)	$I^* + Z \xrightarrow{\delta_I^Z} Z$	$\delta_I^Z = 0.004$
(12)	$I5 + \xrightarrow{\pi} 100 \times V5$	$\pi = 2.5$
(13)	$I4 + \xrightarrow{\pi} 200 \times V4$	$\pi = 2.5$
(14)	$V5 + \xrightarrow{c_5} 0$	$c_5 = 2.5$
(15)	$V4 + \xrightarrow{c} 0$	$c = 2.5$
(16)	$V4 + \xrightarrow{K_F} F + V4$	$K_F = 0.1$
(17)	$dB + \xrightarrow{\alpha} dB + B$	$\alpha = 2.5$ if $T < 20$ otherwise 0
(18)	$B + \xrightarrow{K_B} F + B$	$K_B = 10$
(19)	$B + Z \xrightarrow{\delta_Z^B} Z$	$\delta_Z^B = 0.00001$ if $T < 20$ otherwise 0
(20)	$F \xrightarrow{\delta^F} 0$	$\delta^F = 0.001$
(21)	$Z \xrightarrow{b} 0$	$b = 0.2$
		where I^* stands for $I4$ and $I5$ - two rules.

As far as (the effects of) HAART therapy is concerned, it has been modelled by means of a stochastic triggering event that activates and deactivates the treatment. Activation consists in the modification of interactions (12) and (13), whose rate is downgraded to 1, representing reduced morbidity of the virus, which is one of the main effects of HAART.

Probabilistic Model Checking
Being an introduction to probabilistic model checking beyond the scope of this paper, we refer the interested reader to the cited literature and references therein. Other recent works are [15,16,38]. We introduce informally the main concepts by examples.

The information encoded in a stochastic model describes, roughly speaking, the states in which the modelled system can find itself and the associated probability of being in a state at a certain time. For many natural phenomena that follow a memory-less probability distribution this amounts to a (Continuous Time) Markov Chain (CTMC). Furthermore, given a stochastic model, several algorithms to simulate the possible quantitative evolutions of the system have been defined, noticeably the Gillespie's algorithm [2]. State semantics, such as Transition Systems, have been extensively studied and used to precisely describe the behaviour of dynamical systems, like distributed and concurrent computational systems. State semantics may often become hardly manageable due to the exponential explosion of the number of states. However, several efficient and automated analysis techniques have been defined. Logical formulas can represent "homogeneous" sets of states. A formula like

$$\Diamond a.\phi$$

can be used to represent the set of states in which an "*a*-kind-of" transition can occur and the system evolves into a state fulfilling, in turn, the formula φ. In a sense, the states of the semantics become the model of the formula of interest. The problem of the verification of a formula against a state semantics is known as *Model Checking* [19] and yields either an affirmative answer or a counter example, e.g. a set of states not fulfilling the formula. Most of the times, the model checking procedure is fully automated. Logics [39] and model checking procedures [40] have been defined for probabilistic state semantics, that is *Probabilistic Model Checking*. A typical logic may express formulas like

$$P > p \, [\phi]$$

which, informally speaking, represents all the traces for which φ holds with probability bigger than p. Other examples are $F >= t\, x > 2$, the traces in which *eventually* $x > 2$ after t time units, and $G\, x < 10$, *globally*, for all the traces, the value of x is less than 10.

PRISM
The PRISM probabilistic model checker [20] is a tool for formal modelling and analysis of systems that exhibit random or probabilistic behaviour. Beyond CTMC, PRISM also supports DTCM and Markov decision processes [41]. PRISM has been chosen as it is one of the reference existing model checker, free and open source. However, its features are not determinant here, and other similar platforms could have been chosen. PRISM models are specified in a formal language that describes the entities present in the model, their

```
module Tcell
tc : [0..M] init Tcell;

 [o2]  true & tc < M   ->    1   : (tc' = tc+1);
 [o4b]   tc > 0        ->    tc  : (tc' = tc-1);
 [o5]    tc > 0        ->    tc  : (tc' = tc-1);
 ....

module Icell

 [o4b] true & ic < M   ->    Bk4   : (ic' = ic+1);
      ....

module Virus_x4

 [o4b] v4 > 0   ->    v4     : (v4' = v4-1);
 ....
```

Figure 5
The implementation of the stochastic model (Table 2) in the PRISM modelling language.

behaviour and their quantities. One or more entities can participate synchronously to an event that causes a state change. The state typically consists of the values of the variables of the model. As seen, T cell infection is an event to which virus strains and T cell participate. This event leads to an increased number of infected cells, and less T cells and viruses, which have been combined together. Participation to events has stochastic rates associated. The quantities of entities affect the stochastic dynamics as expected. Figure 5 shows the code for T cell. The variable *tc* keeps trace of the amount of T cells, then two of the possible interactions for T cells are reported, each consisting of a name, an integrity constraint, a rate and an effect: [*o2*] generation of a new T cell, and [*o5*] infection. The [*o5*] interaction is shared with *Virus X4* and *I cell* (see Table 2). As a consequence of the occurrence of such interaction, the former decreases its amount $v4' = v4 -1$, which is required to be non-negative, while the amount of the latter is instead increased. It is worth noting how rates are defined so as to respect stochastic semantics. Rates of synchronously executed actions are multiplied to determine the overall rate of the corresponding transition, i.e. the parameter of the negative exponential distribution associated. In the case of [*o5*], the overall rate is correctly determined by multiplying the amount of *T cell* and *Virus X4* together with the rate B_{k4}. A PRISM model can be used for simulation purposes, and for exact or approximated model checking as illustrated.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
AS and PL authors have studied the deterministic model of infection used in this paper. PL author has investigated the phylogenetic information. AS and AB authors have defined the stochastic version of the model and run in-silico experiments. All authors have worked on experiments results.

References
1. Lió P and Vannucci M: **Finding pathogenicity islands and gene transfer events in genome data.** *Bioinformatics* 2000, **16:**932–940.
2. Gillespie DT: **Exact Stochastic Simulation of Coupled Chemical Reactions.** *The Journal of Physical Chemistry* 1977, **81(25):** 2340–2361.
3. Ho D, Neumann AU, Perelson AS, Chen W, Leonard JM and M M: **Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection.** *Nature* 1995, **373:**123–126.
4. Nowak MA and May RM: *Virus dynamics. mathematical principles of immunology and virology* Oxford University Press; 2000.
5. Perelson AS: **Modelling viral and immune system dynamics.** *Nature Review Immunology* 2002, **2(1):**28–36.
6. Eigen M, McCaskill J and Schuster P: **The Molecular Quasi-Species.** *Advances in Chemical Physics* 1989, **75:**149–263.
7. Sguanci L, Bagnoli F and Li P: **Modeling HIV quasispecies evolutionary dynamics.** *BMC Evolutionary Biology* 2007, **7(2):**S5.
8. Ronga L and Perelson AS: **Modeling HIV persistence, the latent reservoir, and viral blips.** *Journal of Theoretical Biology* 2009.
9. Bagnoli F, Lio' P and Sguanci L: **Modeling viral coevolution. HIV multi-clonal persistence and competition dynamics.** *Physica A* 2005, **366:**333–346.
10. Chao L, Davenport MP, Forrest S and Perelson AS: **A stochastic model of cytotoxic T cell responses.** *Journal Theoretical Biology* 2004, **228:**227–240.
11. Celada F and Seiden PE: **Affinity maturation and hypermutation in a simulation of the humoral immune response.** *Journal of Theoretical Biology* 1996, **26:**1350–1358.
12. De Boer RJ and Perelson AS: **Towards a general function describing T cell proliferation.** *European Journal of Immunology* 1995, **175(4):**567–576.
13. Wiegel FW and Perelson AS: **Some scaling principles for the immune system.** *Immunology and Cell Biology* 2004, **82:**127–131.
14. Regev A and Shapiro E: **Cells as computation.** *Nature* 2002, **419:**343.
15. Mateus D, Gallois JP and Comet ea: **Symbolic modeling of genetic regulatory networks.** *Journal of Bioinformatics and Computational Biology* 2007, **5(2B):**627–640.
16. Fromentin J, Comet JP, Le Gall P and Roux O: **Analysing gene regulatory networks by both constraint programming and model-checking.** *Conference proceedings, IEEE Engineering in Medicine and Biology* 2007, 4595–4598.
17. Batt G, Yordanov B, Weiss R and Belta C: **Robustness analysis and tuning of synthetic gene networks.** *Bioinformatics* 2007, **23:**2415–2422.
18. Batt G, Ropers D, De Jong H, Geiselmann J, Mateescu R, Page M and Schneider D: **Validation of qualitative models of genetic regulatory networks by model checking. Analysis of the nutritional stress response in E. coli.** *Bioinformatics* 2005, **21(Suppl 1):**i19–i28.
19. Clarke EM, Grumberg O and Peled DA: **Model Checking.**The MIT Press; 1999.

20. **The PRISM model checker.** http://www.prismmodelchecker.org.
21. Karlsson I, Antonsson L, Shi Y, Oberg M, Karlsson A, Albert J, Olde B, Owman C, M J and Fenyo EM: **Coevolution of RANTES sensitivity and mode of CCR5 receptor use by human immunodeficiency virus type 1 of the R5 phenotype.** *Journal of Virology* 2004, **78:**11807–11815.
22. Gorry PR, Churchill M, Crowe SM, Cunningham AL and Gabuzda D: **Pathogenesis of macrophage tropic HIV.** *Current HIV Research* 2005, **3:**53–60.
23. Gray L, Sterjovski J, Churchill M, Ellery P, Nasr N, Lewin SR, Crowe SM, Wesselingh SL, Cunningham AL and Gorry PR: **Uncoupling coreceptor usage of human immunodeficiency virus type 1 (HIV-1) from macrophage tropism reveals biological properties of CCR5-restricted HIV-1 isolates from patients with acquired immunodeficiency syndrome.** *Virology* 2005, **337:**384–398.
24. Pavlakis GN, Valentin A, Morrow M and Yarchoan R: **Differential effects of TNF on HIV-1 expression. R5 inhibition and implications for viral evolution.** *Internation Conference AIDS* 2004, **15:**11–16.
25. Parczewski M, Leszczyszyn-Pynka M, M K, Adler G, Binczak-Kuleta A, Loniewska B, Boron-Kaczmarska A and Ciechanowicz A: **Sequence variants of chemokine receptor genes and susceptibility to HIV-1 infection.** *Journal of Applied Genetics* 2009, **50(2):**159–166.
26. Hatse S, Huskens D, Princen K, Vermeire K, Bridger GJ, De Clercq E, Rosenkilde MM, Schwartz TW and D S: **Modest human immunodeficiency virus coreceptor function of CXCR3 is strongly enhanced by mimicking the CXCR4 ligand binding pocket in the CXCR3 receptor.** *Journal of Virology* 2007, **81(7):**3632–3639.
27. Wen W, Chen S, Cao Y, Zhu Y and Yamamoto Y: **HIV-1 infection initiates changes in the expression of a wide array of genes in U937 promonocytes and HUT78 T cells.** *Virus Research* 2005, **113(1):**26–35.
28. Mayanja-Kizza H, Wu M, Aung H, Liu S, Luzze H, Hirsch C and Z T: **The interaction of monocyte chemoattractant protein-1 and tumour necrosis factor-alpha in Mycobacterium tuberculosis-induced HIV-1 replication at sites of active tuberculosis.** *Scandinavian Journal of Immunology* 2009, **69(6):**516–520.
29. Maddocks S, Scandurra GM, Nourse C, Bye C, Williams RB, Slobedman B, Cunningham AL and J BW: **Gene expression in HIV-1/Mycobacterium tuberculosis co-infected macrophages is dominated by M. tuberculosis.** *Tuberculosis (Edinb)* 2009, **89(4):**285–293.
30. Volpe E, Cappelli G, Grassi M, Martino A, Serafino A, Colizzi V, Sanarico N and Mariani F: **Gene expression profiling of human macrophages at late time of infection with Mycobacterium tuberculosis.** *Immunology* 2006, **118(4):**449–460.
31. Hirano F, Komura K, Fukawa E and Makino I: **Tumor necrosis factor alpha (TNF-alpha)-induced RANTES chemokine expression via activation of NF-kappaB and p38 MAP kinase. roles of TNF-alpha in alcoholic liver diseases.** *Journal of Hepatology* 2003, **38(4):**483–489.
32. Ranjbar S, Boshoff HI, Mulder A, Siddiqi N, Rubin EJ and Goldfeld AE: **HIV-1 replication is differentially regulated by distinct clinical strains of Mycobacterium tuberculosis.** *PLoS One* 2009, **4(7):**e6116.
33. Collins KR, Quinones-Mateu ME, Wu M, Luzze H, Johnson JL, Hirsch C, Toossi Z and Arts EJ: **Human immunodeficiency virus type 1 (HIV-1) quasispecies at the sites of Mycobacterium tuberculosis infection contribute to systemic HIV-1 heterogeneity.** *Journal of Virology* 2002, **76(4):**1697–1706.
34. Wilkinson D: **Stochastic Modelling for Systems Biology.**CRC press; 2006.
35. Van-Kampen NG: **Stochastic processes in physics and in chemistry.**Elsevier, Amsterdam; 1992.
36. Kierzek AM: **Stocks. Stochastic kinetic simulations of biochemical system with Gillespie algorithm.** *Bioinformatics* 2002, **18:**470–481.
37. Perelson AS, Neumann AU, Markowitz M, Leonard JM and Ho DD: **HIV-1 Dynamics in Vivo. Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time.** *Science* 1996, **271:**1582–1586.
38. Fisher J, Piterman N, Hajnal A and Henzinger TA: **Predictive modeling of signaling crosstalk during C. elegans vulval development.** *PLoS Computational Biology* 2007, **3(5):**e92.
39. Aziz A, Sanwal K, Singhal V and Brayton R: **Model checking continuous time Markov chains.** *ACM Transactions on Computational Logic* 2000, **1(1):**162–170.
40. Baier C, Haverkort B, Hermanns H and Katoen JP: **Model-checking algorithms for continuous-time Markov chains.** *IEEE Transactions on Software Engineering* 2003, **29(6):**524–541.
41. Kwiatkowska M, Norman G and Parker D: **Quantitative Verification Techniques for Biological Processes.**Algorithmic Bioprocesses, Springer: Condon A, Harel D, Kok J, Salomaa A, Winfree E 2009.
42. Bruch CL and Chao L: **Evolvability of an RNA virus determined by its mutational neighborhood.** *Nature* 2000, **406:**625–628.
43. Asquith B, Edwards CTT, Lipsitch M and McLean AR: **Inefficient Cytotoxic T Lymphocyte Mediated Killing of HIV-1 Infected Cells In Vivo.** *PLOS Biology* 2006, **4(4):**e90.
44. Althaus CL and Be Boer RJ: **Dynamics of Immune escape during HIV/SIV Infection.** *PLOS Computational Biology* 2008, **4(7):**e1000103.