



Published in final edited form as:

Proc Am Control Conf. 2017 May ; 2017: 122–127. doi:10.23919/ACC.2017.7962941.

Experiment Design for Early Molecular Events in HIV Infection

Aditya Jagarapu,

Dept. of Biomedical Engineering, University of Delaware

LaMont Cannon, and

Dept. of Biomedical Engineering, University of Delaware

Ryan Zurakowski [Senior Member, IEEE]

Dept. of Biomedical Engineering, University of Delaware

Abstract

The recent introduction of integrase inhibitors to the HIV antiviral repertoire permits us to create in vitro experiments that reliably terminate HIV infection at the point of chromosomal integration. This allows us to isolate the dynamics of a single round of infection, without needing to account for the influence of multiple overlapping rounds of infection. By measuring the various nucleic acid concentrations in a population of infected target cells at multiple time points, we can infer the rates of these molecular events with great accuracy, which allows us to compare the rates between target cells with different functional phenotypes. This information will help in understanding the behavior of the various populations of reservoir cells such as active and quiescent T-cells which maintain HIV infection in treated patients. In this paper, we introduce a family of models of the early molecular events in HIV infection, with either linear dynamics or age-structured delays at each step. We introduce an experimental design metric based on the delta AIC (Akaike Information Criteria) between a model fit for simulated data from a matching model vs a mismatched model, which allows us to determine a candidate experiment design's ability to discriminate between models. Using parameters values drawn from experimentally-derived priors corrupted with appropriate measurement noise, we confirm that a proposed sampling schedule at different time points allows us to consistently discriminate between candidate models.

I. Introduction

Human Immunodeficiency virus (HIV) affects specific cells of the immune system called CD4⁺ T cells. Uncontrolled replication of HIV will lead to Acquired Immunodeficiency Syndrome (AIDS). An estimated 36.7 million people are living with HIV as of 2015[1]. Advances in anti-retroviral therapy have dramatically improved our ability to control this infection. Several techniques are used to control viral replication by targeting different steps in the life cycle of the virus, using inhibitors that act either alone or in combinations. Even though treatment with antiretroviral therapy may reduce the plasma viral load to concentrations below the limit of detection, it cannot eliminate the virus completely. Rapid rebound of viremia is observed during treatment interruption. This may be due to activation of the stable viral reservoirs or ongoing rounds of successful infection of CD4⁺ T cells [2, 3, 4, 5, 6]. These processes are not fully understood; investigating the kinetics and mechanisms of active versus latent infection is an important step in developing a cure for HIV.

Several mechanisms have been proposed to explain the reasons behind viral rebound that takes place during the early stage of infection. Pre-integration factors such as incomplete reverse transcription reaction due to low levels of nucleotide concentrations in the cytoplasm, failure to import the ribosome-sized pre-integration complex into the nucleus, and rapid degradation of viral RNA and unintegrated DNA limit progression of viral replication [7, 8]. Understanding dynamic rates of key intracellular molecular events during the life cycle of HIV will help us in identifying the rate limiting steps that are responsible for the differences in observed phenotypic behavior between quiescent and active T cells. Introduction of integrase inhibitor during the viral replication of HIV will assist in conducting reliable in-vitro experiments that will guide us in determining the rate parameters during the early HIV infection. Concentrations of various nucleic acids collected at different time points from integrase inhibitor cultured cell types of different activation levels, interpreted through dynamic models, will allow us to determine the rate constants of each major molecular step during early infection. This will provide us with data to narrow the viable hypotheses concerning the critical differences between quiescent and active T-cells during early HIV infection, which may in turn suggest new molecular targets and therapeutic approaches to controlling the HIV reservoir. A family of mathematical models describing the early intracellular molecular events following viral entry into the host cell in the presence of integrase inhibitors have been developed in this study. An experimental design is introduced suggesting an appropriate sampling schedule for various nucleic acid concentrations that will give us sufficient information to estimate the parameters from the models. A metric is introduced to evaluate the ability of an experiment design to distinguish between candidate models of a family in our study. Several approaches have also been applied by others in order to study model distinguishability based on different metrics [21, 22]. One study used maximizing distance (L-2 norm) between the outputs of rival models by varying initial conditions, external inputs and structural changes [23]. Experiment design for our problem is tested using simulated data to verify our ability to distinguish between source models. The overall goal of this study is to propose an experimental design that will help us in discriminating possible candidate models by selecting an appropriate model using experimental data.

II. Model

A. Life Cycle of HIV and the role of Integrase Inhibitor

HIV attacks CD4 T cells and attaches itself to the surface receptors CD4 and CCR5 or CXCR4 co-receptor. After fusion with the host cytoplasm the viral core has the capability of converting its RNA genome into unintegrated DNA in the cytoplasm (DNA_{u}) using reverse transcriptase [9, 10]. The resulting unintegrated DNA is transported to the nucleus, where it integrates with the host genome [11]. Transcription of the integrated viral genome produces full length mRNA that are further spliced into singly and doubly mRNA transcripts [12]. Translation of full length mRNA produces necessary structural and enzymatic proteins that will help in the development of the progeny of the virus [13]. When sufficient quantities of viral material accumulate these proteins are transported to the plasma membrane and start budding. Integrase inhibitors prevent the integration of the linear viral DNA virus into the host genome. As result 2-LTR (Long terminal repeats) circles are formed from the un-

integrated nuclear DNA, which will be used as markers for cellular replication [14]. A compartment model for the life cycle of HIV is shown in Figure 1 indicating different stages in a host cells such as viral entry, reverse transcription (RNA to DNA), integration (nuclear DNA to integrated DNA), translation (DNA to mRNA) and other nuclear transport steps such as transfer of DNA from cytoplasm to nucleus, release of viral particles.

Three different deterministic models; linear, single delay and double delay dynamics; are considered to model these intracellular molecular events. These models track the populations of HIV RNA, linear unintegrated DNA and 2-LTR episomes with respect to time. Measurement units of each state are in copies/cell.

B. Linear Model (M1)

Our first model assumes linear transition rates between each of the reverse transcription and integration steps. Viral RNA (R) is converted to linear unintegrated DNA (DNA_u) at a linear rate k_1 (days^{-1}) and decays at linear rate k_3 (days^{-1}). Similarly linear unintegrated DNA is converted into 2-LTR (L) circles at a linear rate k_2 (days^{-1}) and decays at a linear rate k_4 (days^{-1}). Solid lines indicate the transition rates and the dashed lines indicate the degradation rates in figures 2, 3, and 4.

The model equations for the linear kinetic model M1 are

$$\frac{d[R]}{dt} = -k_1[R] - k_3[R] \quad (1)$$

$$\frac{d[\text{DNA}_u]}{dt} = k_1[R] - k_2[\text{DNA}_u] - k_4[\text{DNA}_u] \quad (2)$$

$$\frac{d[L]}{dt} = k_2[\text{DNA}_u] \quad (3)$$

C. Single Delay Model (M2)

The second model assumes a time delay that dominates the conversion of the single stranded RNA into unintegrated double stranded DNA, which will be a substrate for the integration step. Reverse Transcription involves the creation of a DNA complementary copy of the parent RNA strand, through the sequential aggregation of nucleotides. The DNA copy is 9748 nucleotides long, which are added sequentially [10]. Queued processes of this type result in approximate delay kinetics. The single delay model is similar to the linear kinetic model as discussed above with the exception that the viral RNA (R) is converted to linear unintegrated DNA (DNA_u) at a constant rate delayed by time τ_1 (days). We introduce this delay using age structured time delay compartment model by introducing 10 (n) stages

between RNA and unintegrated DNA compartments. Each stage representing the addition of certain number of nucleotides to the RNA genome.

$$\frac{d[R_1]}{dt} = -\frac{n}{\tau_1}[R_1] - k_3[R_1] \quad (4)$$

$$\frac{d[R_i]}{dt} = \frac{n}{\tau_1}[R_{i-1}] - \frac{n}{\tau_1}[R_i] - k_3[R_i], i=1, 2, 3, \dots 10 \quad (5)$$

$$\frac{d[\text{DNA}_u]}{dt} = \frac{n}{\tau_1}[R_{10}] - k_2[\text{DNA}_u] - k_4[\text{DNA}_u] \quad (6)$$

$$\frac{d[L]}{dt} = k_2[\text{DNA}_u] \quad (7)$$

D. Double Delay Model (M3)

In addition to the delay described during reverse transcription in the previous model we also assume a delay in the integration step of unintegrated DNA into host genome. After reverse transcription the unintegrated DNA is transported into the nucleus, its terminal ends are cleaved, and the linear viral DNA is inserted into a chromosome, resulting in pro-viral DNA. Transport within cells is accomplished by actin-myosin motion along cytoskeletal filaments, which may be well modeled as a queueing process that introduces a delay; we model the transport kinetics of DNA_u into the nucleus with a delay (τ_2 , days) using similar age structured modeling with 10 stages (n) as mentioned previously.

$$\frac{d[R_1]}{dt} = -\frac{n}{\tau_1}[R_1] - k_3[R_1] \quad (8)$$

$$\frac{d[R_i]}{dt} = \frac{n}{\tau_1}[R_{i-1}] - \frac{n}{\tau_1}[R_i] - k_3[R_i], i=2, 3, \dots 10 \quad (9)$$

$$\frac{d[\text{DNA}_{u,1}]}{dt} = \frac{n}{\tau_1}[R_{10}] - \frac{n}{\tau_2}[\text{DNA}_{u,1}] - k_4[\text{DNA}_{u,1}] \quad (10)$$

$$\frac{d[\text{DNA}_{u,j}]}{dt} = \frac{n}{\tau_2} [\text{DNA}_{u,j}] - \frac{n}{\tau_2} [\text{DNA}_{u,j}] - k_4 [\text{DNA}_{u,j}] \quad j=2,3,\dots,10 \quad (11)$$

$$\frac{d[L]}{dt} = \frac{n}{\tau_2} [\text{DNA}_{u,10}] \quad (12)$$

III. Preliminary Data

With all the three models in place we examined the model fit with experimental data obtained from Pace et al. [15]. Infecting CD4+ T cells with HIV in the presence of integrase inhibitor raltegravir generated the experimental data. Measurements for total HIV DNA (un-integrated DNA and 2-LTR) and 2-LTR circles were made between 0 and 10 days post infection as shown in figure 5.

The parameters for all the three models were extracted from the five experimental data points using Bayesian Markov Chain Monte Carlo (MCMC) methods with Gibbs sampling [16-18] using non-informative priors. Posterior distributions were generated using likelihood function for qPCR measurement assay with 50 copies as the censoring limit and lognormal noise with a density-dependent standard deviation, $\sigma(c)$:

$$\sigma(c) = 10^{-0.21 - 0.24 \log(C)} \quad (13)$$

where 'c' is the expected number of copies in the sample predicted by the model. Given the model limit of detection and log normal standard deviation, the likelihood function for a measurement 'm' given a model concentration 'c' as follows:

$$L((c|m) = \begin{cases} f_{\text{LN}}(m, c, \sigma(c)), & m > 50 \\ F_{\text{LN}}(50, c, \sigma(50)), & m = 50 \end{cases} \quad (14)$$

Where, f_{LN} is the lognormal probability distribution and F_{LN} is the lognormal cumulative distribution function [19]. Figure 5 shows model fit with experimental data using parameter values obtained from maximum likelihood estimates.

All the three models fit the experimental data with similar likelihood values as shown in table 1 and cannot be distinguished from one another with the limited set of experimental data available. In order to identify the true biological phenomena we need to have a better experimental design which will help us in distinguishing between the above mathematical models and also help in furnishing better parameter estimates.

IV. Experimental Design

Failure to identify the appropriate model given limited experimental data from Pace et al. forces us to come up with an experimental design that can distinguish all the three models. Based on input from our experimental collaborator, we propose in-vitro experiments on CD4+ T cells from a normal donor inoculated with HIV and cultured in IL-7 (Interleukin) and raltegravir. RNA should be harvested from the cells every two hours for the first 12 hours; DNA should be harvested from the cells every two hours for two days and then every 8 hours till four days post infection. At each time point, approximately 3 million cells will be harvested. The RNA, 2-LTR and total HIV DNA levels should be measured at each time point using qPCR techniques.

In order to verify whether our experimental design can help us in distinguishing models, we generated simulated data from each of the three models M1 (Linear), M2 (Single Delay) and M3 (Double Delay). Simulated data for all the time points were generated according to the sampling schedule as mentioned above. However the simulated data is noiseless and the process of measurement using qPCR technique produces some noise. Noise resulting from the Polymerase Chain Reaction (PCR) is lognormal with a standard deviation given by equation 13 and was applied to the noise free simulated data, c , generated from the model. Each set of simulated data is tested against all three models in order to identify whether the model that generated the data can be determined from the data.

In order to test the model compatibility we use MCMC analysis to compute the maximum likelihood for each of the models against the simulated data. Later using this maximum likelihood value, we calculated Akaike Information Criteria (AIC) for each of the model. The AIC, represented in equation 15 estimates the relative goodness-of-fit of a model by approximating how much information is lost when a given model is used to represent a process [20]. Table II lists all the maximum likelihood values generated from each model (rows) with simulated data from a specific model (columns).

With the AIC values in place for each model (AIC_M), we compare it with the best AIC (i.e. AIC_{min}) for the respective simulated data set in order to establish whether a given model matched with the same model from which the data is generated. The difference between the best and a particular model AIC values is termed as ΔAIC according to equation 16. Table III lists AIC values for all the models with respect to the simulated data,

$$AIC = 2 * k - 2 * \ln(L) \quad (15)$$

$$\Delta AIC = AIC_{min} - AIC_M \quad (16)$$

Where, 'k' is the number of free parameters in the model and 'L' is the maximum likelihood value.

Table IV displays AIC values across all the models generated from equation 16. Our results show that each model matches to the simulated data generated from the same model (i.e. $AIC = 0$) and hence distinguishability between models with respect to data has been attained.

Table V displays the range obtained for all the parameters obtained from simulated data of M1 (1), M2 (2), M3 (3) compared against parameters obtained from experimental data Pace et al. (P) [15]. Simulated data from the suggested experimental design gave us parameters with much tighter estimates when compared to the parameters obtained from Pace et al.

A. Simulated Data generated from Model 1 (Linear Dynamics) vs Simulation results from all Models

Figure 6 shows simulation results across all the models with maximum likelihood parameter estimates obtained from MCMC techniques with simulated data from Model 1. It is shown that model 1 (M1) ideally fits the simulated data while other models fail reproducing them. This also matches with AIC values as we can see from table IV that the corresponding value for simulated data from M1 and model simulation from M1 is 0.

B. Simulated Data generated from Model 2 (Single Delay Dynamics) vs Simulation results from all models

Figure 7 shows simulation results across all the models with maximum likelihood parameter estimates obtained from MCMC techniques with simulated data from Model 2. It is also shown that model 2 (M2) ideally fits the simulated data while other models fail reproducing them. This also matches with AIC values as we can see from table IV that the corresponding value for simulated data from M2 and model simulation from M2 is 0.

C. Simulated Data generated from Model 3 (Double Delay Dynamics)

Figure 8 shows simulation results across all the models with maximum likelihood parameter estimates obtained from MCMC techniques with simulated data from Model 3. It is also shown that model 3 (M3) ideally fits the simulated data while other models fail reproducing them. This also matches with AIC values as we can see from table IV that the corresponding value for simulated data from M3 and model simulation from M3 is 0.

V. Conclusion

Three different models viz. linear dynamics, single delay and double delay dynamics have been proposed to understand intracellular molecular events of HIV in the presence of integrase inhibitors. Distinguishability between models with respect to experimental data has also been demonstrated with simulated data from the proposed experimental design. Our results illustrate that the proposed experimental design will help in understanding the biological phenomena underlying during viral replication between quiescent and active T cells.

The proposed experiment design was suggested by our collaborators, and represents perhaps the highest possible achievable sampling. Now that feasibility of the approach is shown,

future work will focus on optimizing the trade-off between model fit accuracy and experimental costs, including the total number of samples, the cells used per sample, and the frequency of sampling.

Acknowledgments

We would like to thank Dr. Una O'Doherty for introducing us to this problem and supplying the data used in this research.

¹*Research is partially supported by the National Institutes of Health (NIH) grant AI110288, and Merck. The content is solely the responsibility of the authors and does not necessarily represent the official view of the funders. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. UN AIDS FACTSHEET. 2016. <http://www.unaids.org/en/resources/fact-sheet>
2. Josefsson L, Dahl V, Palmer S. Can HIV infection be eradicated through use of potent antiviral agents? Current opinion in infectious diseases. 2010; 23:628–632. [PubMed: 20847693]
3. Maldarelli F. Targeting viral reservoirs: ability of antiretroviral therapy to stop viral replication. Current Opinion HIV AIDS. 2011; 6:49–56.
4. Dornadula G, Zhang H, VanUitert B, et al. Residual HIV-1 RNA in blood plasma of patients taking suppressive highly active antiretroviral therapy. JAMA. 1999; 282:1627–32. [PubMed: 10553788]
5. Furtaldo MR, Callaway DS, Phair JP, et al. Persistence of HIV-1 transcription in peripheral blood mononuclear cells in patients receiving potent antiretroviral therapy. N Engl J Med. 1999; 340:1614–22. [PubMed: 10341273]
6. Sharkey ME, Teo I, Greenough T, et al. Persistence of episomal HIV-1 infection intermediates in patients on highly active anti-retroviral therapy. Nat Med. 2000; 6:76–81. [PubMed: 10613828]
7. Zack JA, Arrigo SJ, Weitsman SR, Go AS, Haislip A, Chen ISY. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. Cell. 1990; 61:213–22. [PubMed: 2331748]
8. Zack JA, Haislip AM, Krogstand P, Chen ISY. Incompletely reverse-transcribed human immunodeficiency virus type 1 genomes function as intermediates in the retroviral life cycle. J Virol. 1992; 66:1717–25. [PubMed: 1371173]
9. Morrow CD, Park J, Wakefield JK. Viral gene products and replication of human immunodeficiency virus type 1. Am j Physiol. 1994; 266:c1135–c1156. [PubMed: 8203479]
10. Hu, Wei-Shau, Hughes, Stephen H. HIV-1 Reverse Transcription. Cold Spring Harbor Perspectives in Medicine 2.10. 2012 a006882. PMC. Web. 20 Sept. 2016.
11. Craigie, Robert, Bushman, Frederic D. HIV DNA Integration. Cold Spring Harbor Perspectives in Medicine 2.7. 2012 a006890. PMC. Web. 20 Sept. 2016.
12. Purcell DF, Martin MA. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol. 1993; 67:6365–6378. [PubMed: 8411338]
13. Schwartz S, Felber BK, Pavlakis GN. Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. Mol Cell Biol. 1992; 12:207–219. [PubMed: 1729599]
14. Murray JM. HIV dynamics and integrase inhibitors. Antivir Chem Chemother. 2009; 19:157–164. [PubMed: 19374143]
15. Pace MJ, Graf EH, O'Doherty U. HIV 2-long terminal repeat circular DNA is stable in primary CD4+T Cells. Virology. 2013 Jun 20; 441(1):18–21. [PubMed: 23537959]
16. Huang Y, Wu H, Acosta EP. Hierarchical Bayesian inference for HIV dynamic differential equation models incorporating multiple treatment factors. Biomet J. 2010; 52:470–486.
17. Putter H, Heisterkamp SH, Lange JMA, de Wolf F. A Bayesian approach to parameter estimation in HIV dynamical models. Stat Med. 2002; 21:2199–2214. [PubMed: 12210633]

18. Luo, Rutao, Fabian Cardozo, E., Piovoso, Michael J., Wu, Hulin, Buzon, Maria J., Martinez-Picado, Javier, Zurakowski, Ryan. Modelling HIV-1 2-LTR dynamics following raltegravir intensification. *J Royal Society Interface*. 2013; 10(84)
19. Luo, Rutao, Piovoso, Michael J., Zurakowski, Ryan. Modeling Uncertainty in Single-Copy Assays for HIV. *Journal of Clinical Microbiology*. 2012; 50(10):3382–3383.
20. Burnham, Kenneth P., Anderson, David R. Model selection and multimodel inference: a practical information-theoretic approach. Springer; 2002.
21. Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, Bullinger E, Allgöwer F, Gilles ED. A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. *Genome Res*. 2004 Sep; 14(9):1773–85. [PubMed: 15342560]
22. Anderson, James, Papachristodoulou, Antonis. On validation and invalidation of biological models. *BMC Bioinformatics*. 2009; 10(1):132. [PubMed: 19422679]
23. Melykuti, Bence, August, Elias, Papachristodoulou, Antonis. Discriminating between rival biochemical network models: three approaches to optimal experiment design. *BMC Systems Biology*. 2010; 4:38. [PubMed: 20356406]

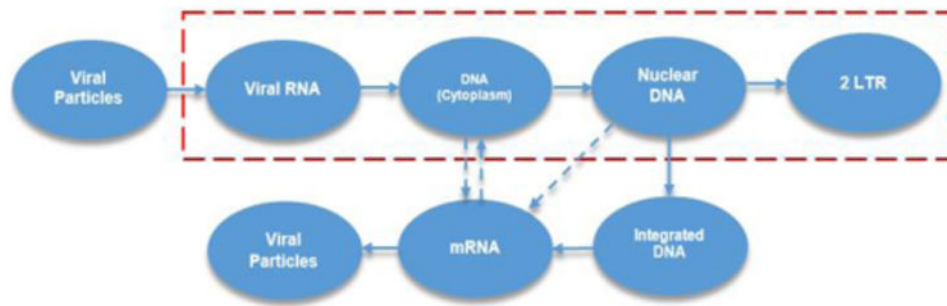


Figure 1. Compartmental Model: Life Cycle of HIV (Highlighted area in the red box indicate the different stages that occur in the presence of Integrase Inhibitor)

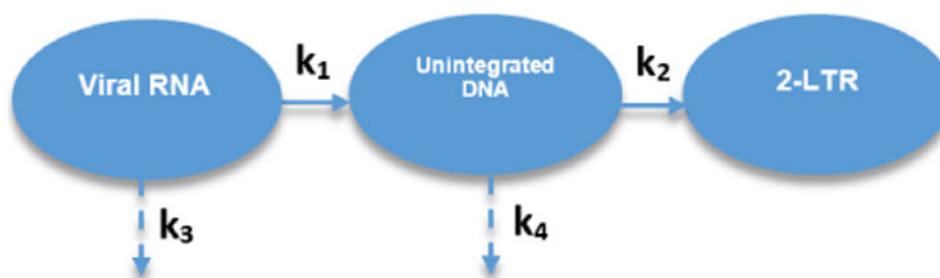


Figure 2. Compartmental Model of Linear Dynamics

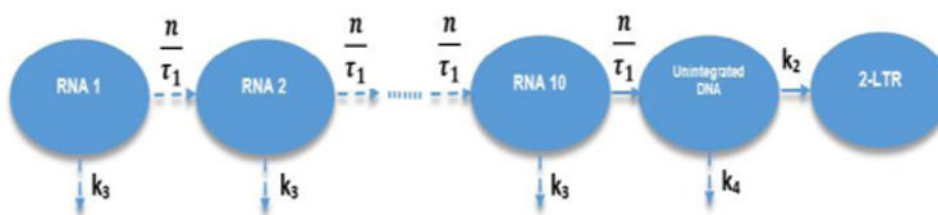


Figure 3. Compartmental Model of Single Delay Dynamics

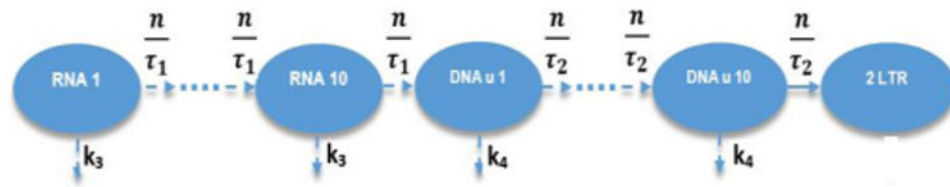


Figure 4. Compartmental Model of Double Delay Dynamics

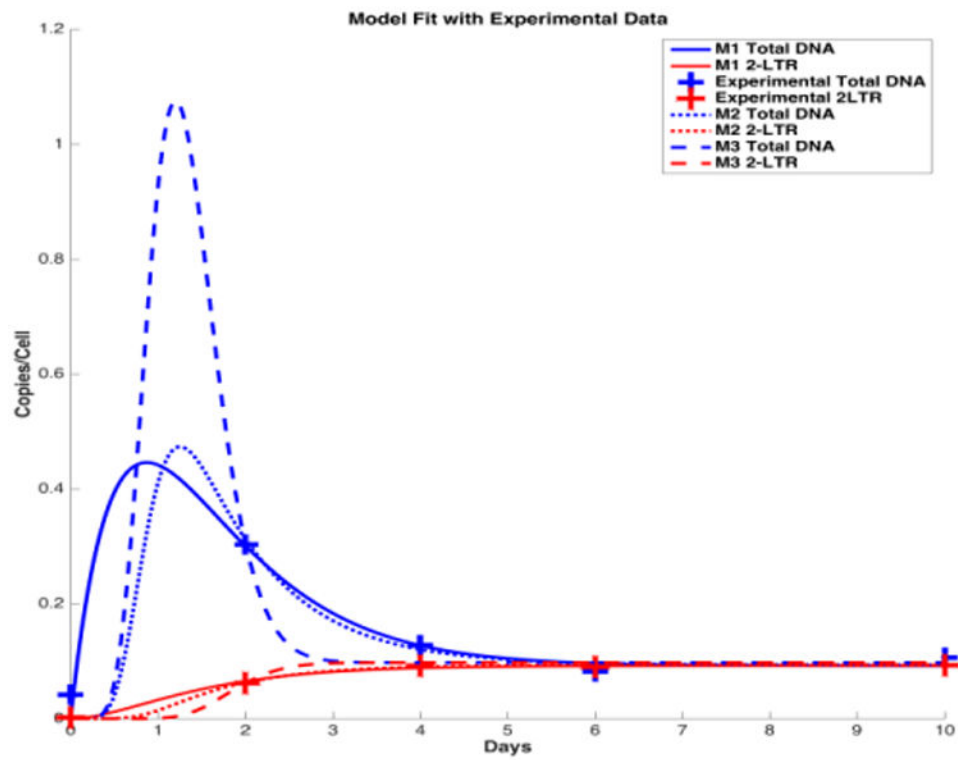


Figure 5. Model Fit with Experimental Data obtained from Pace et. al.[15]

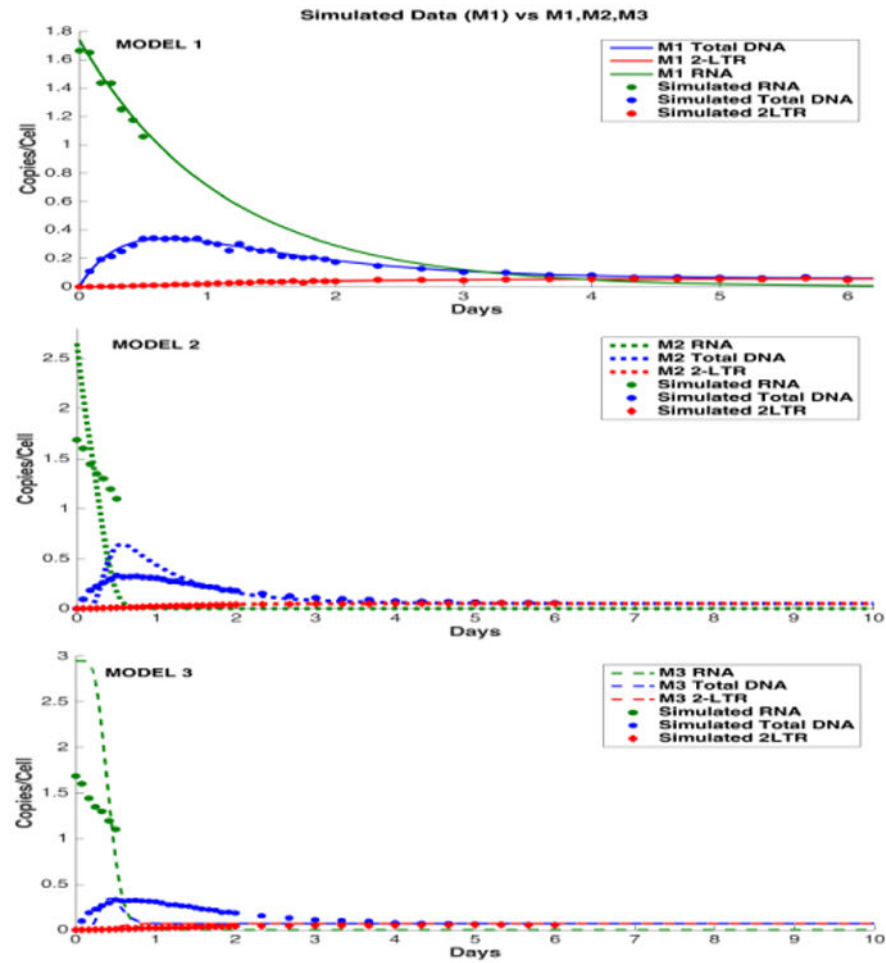


Figure 6.

Model Fit across all the models with simulated data generated from Model 1. Model distinguishability can be observed as Model 1 fits the simulated data from Model 1 while Model 2 and 3 do not fit properly.

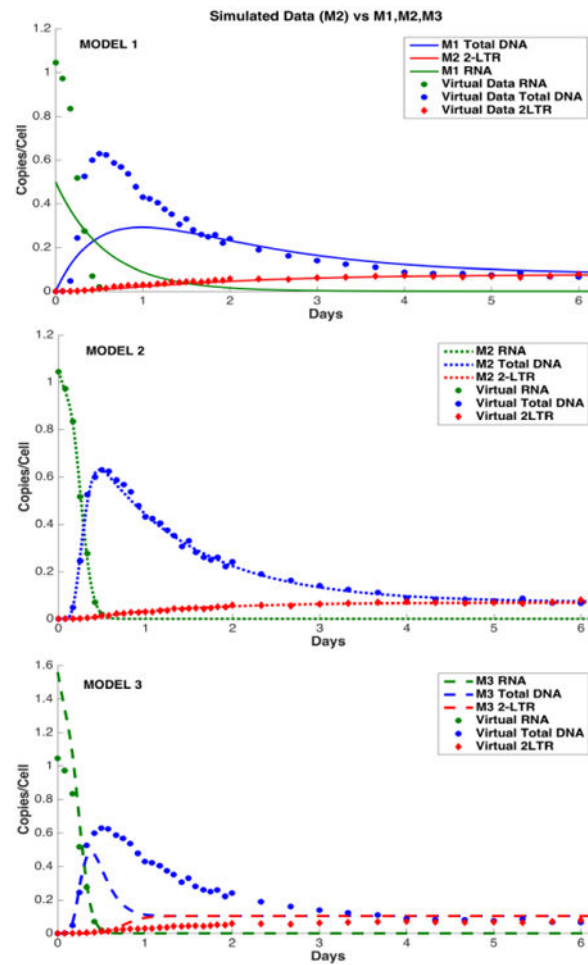


Figure 7.

Model Fit across all the models with simulated data generated from Model 2. Model distinguishability can be observed as Model 2 fits the simulated data from Model 2 while Model 1 and 3 do not fit properly.

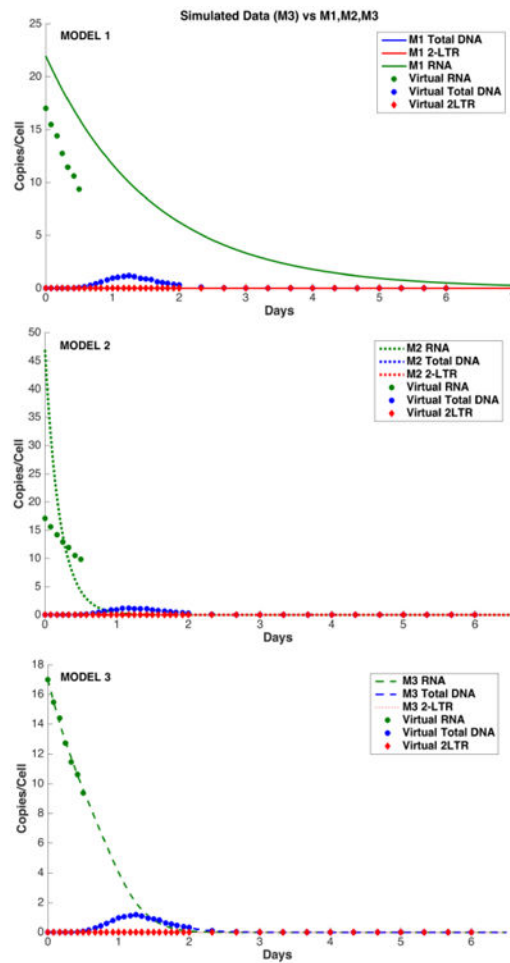


Figure 8.

Model Fit across all the models with simulated data generated from Model 3. Model distinguishability can be observed as Model 3 fits the simulated data while Model 1 and 2 do not fit properly.

Table I
Maximum likelihood values for the three models with experimental data (Log-Likelihood)

MODEL	M1	M2	M3
Max Likelihood Value	-95.9446	-95.9209	-96.9999

Table II
Maximum Likelihood Values Across Different models vs Simulated Data Across different Models

Maximum Likelihood		Simulated Data		
		M1	M2	M3
Model	M1	-926.1	-1735.1	-3390.1
	M2	-1458.9	-921.31	-3964.8
	M3	-2167.7	-1729.1	-649.0

Table III
AIC Values Across Different Models vs Simulated Data Across Different Models

AIC		Simulated Data		
		M1	M2	M3
Models	M1	1862.2	3480.2	6790.2
	M2	2927.8	1852.6	7939.6
	M3	4345.4	3468.2	1308.0

Table IV
AIC Values Across Different Models vs Simulated Data Across Different Models

AIC		Simulated Data		
		M1	M2	M3
Model	M1	0	-1627.5764	-5482.19
	M2	-1065.572	0	-6631.59
	M3	-2483.172	-1615.5764	0

Table V

Range For The Values of Parameters Obtained From Simulated Data In Comparison To Data Obtained From Pace et al.

Range		MODEL		
		M1	M2	M3
PARAMETERS	K_1 (days ⁻¹)	0.2048(1)	-	-
		2.5140 (P)	-	-
	K_2 (days ⁻¹)	0.0855 (1)	0.0142 (2)	-
		0.0961 (P)	0.1288(P)	-
	K_3 (days ⁻¹)	0.2393 (1)	1.5002(2)	0.2351(3)
		0.6847 (P)	5.5817(P)	1.1270(P)
	K_4 (days ⁻¹)	0.8424 (1)	0.1565(2)	0.3071(3)
		1.8134 (P)	1.3016(P)	2.1868(P)
	τ_1 (days)	-	0.0167(2)	0.0330(3)
		-	1.2381(P)	0.7175(P)
	τ_2 (days)	-	-	0.1965(3)
		-	-	0.7588(P)