Agent based modeling of viral Infections: an investigation across several spatio-temporal scales

Mancini, E.

Citation for published version (APA):
3. A Simulation Framework to Investigate \textit{in vitro} Viral Infection Dynamics

3.1 Introduction

In the past ten years there has been a growing interest in modeling viral infections to study the characteristics of their dynamics. Early mathematical models were typically based on ordinary differential equations (ODEs) and focused on extracting key parameters of the infections dynamics [1, 2]. Later on, interest moved to investigating how spatial relations affected the system dynamics, thus moving to cellular automata (CA) models [3,4,5]. More recently the attention has shifted focus to viral respiratory diseases due to the increasing danger of pandemics such as the “Spanish flu” that caused over 50 million deaths worldwide in 1918 or the more recent H1N1 swine flu. The risk of new pandemics has a high impact on society and new research projects are being funded to investigate the different scales of threat, from the epidemiological spread of a virus in a population to its early phases of the infection in a cell culture. Among the papers published on this topic, in particular on the influenza virus [20,21,22], there are Bocharov’s ODE model [6] and Beauchemin’s CA model [3,4,7]. Both models described the dynamics of the influenza infection in the upper epithelial areas of the lungs and were validated by clinical data. Although they take into account the effect of the immune system on the dynamics of the disease, describing the infection until its final outcome (more than a week after the first infection), none of them analyzed which mechanisms where critical in the dynamics of the first and second round of infection. To understand the key mechanisms of viral respiratory diseases, we focus instead on early phases of viral infection by using \textit{in vitro} experiments observing lung cells up to 24 hours post infection (PI). The \textit{in vitro} experiments provide measurements of virus titer, spatial characteristics of cell growth and, through green fluorescent protein (GFP) imaging, of the infection spread.

Our computational model focuses on simulating the early stages of a
viral infection in a population of cells plated on a culture well. The choice of a CA model was natural since the in vitro infections being studied use host lung cancer cell lines that form a fixed mono-layer in which spatially dependent aspects of infection may be present [8,9]. We developed this computational model using the Multi-Agent System Visualization (MASyV) platform [3]. In contrast to previous models, we explicitly focus on the dynamics of virus spread on a population of cells, supported by experimental data from an in vitro model system. We also explicitly model the infectious viral particles as discrete entities, whereas in previous models the infection of cells followed simple CA rules depending on the states of neighboring cells. These viral particles are released by infected cells according to a specific function based on time post infection, and move over the well with a random walk algorithm. This representation allows us to model the mechanisms of virus spread in an environment where the virus is not confined and can also infect cells not adjacent to the infected ones.

In Section 2 we describe the model design and its main features. We also describe the SARS infection experimental data used to parameterize the model. In Section 3 we present a sensitivity analysis that identifies the critical mechanisms characterizing the early phases of the infection. We also show that the model can explain the experimentally observed virus titer data and allows a deeper understanding of the infection dynamics in the in vitro experiments.

3.2 Materials and Methods

3.2.1 Simulation Environment

The computational model is built using Beauchemin’s MASyV platform. The software consists of a server providing I/O and supervisory services to the various client modules where the simulation is actually coded. Our choice to use MASyV was partially driven by flexible and powerful graphical visualization routines that facilitate comparison to images provided by the experimental collaborators. MASyV has a C-based API and is open source allowing
finalized custom models to be easily shared. We discuss novel contributions and differences from the previous modules. The original module details are presented in Beauchemin *et al.* [3].

Our model reproduces a viral infection on a population of cells plated on a culture well. In our client we consider, as the target of the viral infection, Calu-3 cells that are a human airway epithelial cell line derived from human lung cancer. We model these host cells using a 130x130-site CA model where each site represents either one calu-3 cell or an empty space. At the beginning of the simulation each lattice site is initialized and labelled with “uninfected” or “empty” states as described below in Section 2.2. Uninfected cells are initially stochastically infected with virus through a first round of infection at the beginning of the simulation, described in Section 2.3, and once infected progress through the following states:

- **Containing:** initial infection state representing viral entry and hijacking of host cell mechanisms necessary for viral replication.
- **Expressing:** cell is actively producing and assembling virus capsids and genomes internally, but has not begun releasing virion.
- **Infectious:** Assembled virion is being released from the host cell according to the release function (Section 2.4)
Figure 1. SARS-CoV Infection Virus Titer Measurement: Virus titer resulting from Calu-3 cells infected with SARS virus using an estimated Multiplicity of Infection (MOI) of 2. Measured time points are indicated in hours post infection.

By examining the experimental viral titer data shown in Figure 1 we derived temporal delay of the state transition between Containing and Infectious. The ~ \log_{10} 3 viral titer measurements at time points 0, 4 and 7 hours post infection are residual virus left over from the initial infection after washing. We set the delay for release of new infectious viral particles to 7 hours to represent the jump in observed virus titer between 7 and 12 hours. Cells release virus particles according to a viral release function, described below. Virus particles are entities that diffuse from one lattice point to another according to a random walk algorithm (Section 2.4). These free-floating virus particles stochastically infect uninfected cells in a second round of infection. The CA lattice is therefore like a tissue of immobile cells with infective virions moving over it. We represent only infectious viral particles to
relate their count in the model with the viral titer measured in terms of Plaque Forming Unit per mL (PFU/mL). A PFU is a measure of the number of viral particles capable of forming plaques per unit of volume, thus only infectious particles are counted in the viral titer. The CA lattice is updated synchronously and the boundary conditions for both the epithelial and viral particles are periodic, i.e. epithelial cells and viral particles can grow or move outside the boundary of the virtual well. We use a Honeycomb neighborhood: each lattice site is considered adjacent to six sites and cells can replicate only if a neighboring site is empty or if it contains a dead cell. Only uninfected cells are allowed to replicate. When cells on the lattice are initialized all uninfected cells are stochastically assigned a TIME_TO_DEATH value between 1 and an arbitrary CELL_LIFESPAN variable. CELL_LIFESPAN was fitted in order to have natural cell death for 5% of the cells within the simulated 24 hours. CELL_LIFESPAN is not affected by infection as in vitro experimental observations showed no increase in cell death due to infection within 24 hours.

3.2.2 Cell placement

In the laboratory before an in vitro infection is performed, $10^6$ cells are first plated in each of the 6 culture wells present in a dish, and given time to settle down on the well surface to create a cell mono-layer, before the virus is applied to the cell culture. In the plating process, a suspension of cells is plated onto the surface of the well that adhere to the plastic in small clumps or as single cells forming "islands" that are homogeneously distributed throughout the well. Cells located on the inside of these islands (surrounded by other cells) do not replicate due to contact inhibition, whereas cells on the outer edges of each island replicate until they are completely surrounded by other cells. As a result each island continues to grow until there is no adjacent empty space on the well surface.
Figure 2. Microscopy Images of *in vitro* Calu-3 Cells: Images used to determine simulated island size. Cells grow in groups or “islands” with an initial average size of 116 cells.

To represent this behavior we used a set of images (Figure 2) that shows how cells are distributed on the culture well just before the infection is performed. The number of cells forming each isolated island was counted for each image obtaining a distribution of island sizes with an average of 116 cells per island and a standard deviation of 74 cells per island. This data was used to generate randomly positioned clusters of cells on the grid, each with a size extracted from the measured distribution. Islands are placed iteratively until the simulated confluency matches the experimentally measured confluency.

Figure 3 shows lattice initialization using different CONFLUENCY parameter values. A value of .45 was derived based on calculations from well surface area, the total number of cells and the size of a single cell (~20 μm² see ATCC catalogue). This value was used for simulations performed in the results section.
A cell line doubling time is defined as the time required for the population of cells to double in number. We used a doubling time of 48 hours for Calu-3 cells (Ardizzoni Lab, personal communication). Cell density directly affects local cell growth since cells completely surrounded by other cells are not able to replicate. For this reason in our model the initial spatial distribution of uninfected cells affects the overall cell growth on the virtual culture well. In order to obtain a correct cell growth in our model we used a parameter called DIVISION_TIME that measures the time necessary to a uninfected cell to duplicate. We adjusted DIVISION_TIME to match the simulated doubling time with the experimentally-derived one. This was accomplished by analyzing cell growth of uninfected cells over 48 hours as a function of time for different values of the parameter. This led us to fix the value of DIVISION_TIME to 10 hours.

3.2.3 First round of infection

Cells in the culture well are initially infected using a Multiplicity of Infection (MOI) of 2 (see section 2.7.4). The MOI is calculated as the ratio of infectious viral particles to the number of Calu-3 target cells.
By definition the proportion of infected cells is given by the Poisson distribution that describes the infection process [12]. Using the experimental MOI would theoretically give us the exact percentage of infected cells at the beginning of the experiment but even though the number of particles can be measured with good accuracy not all the viral particles used in this first infection process are infective and the estimate is derived using VeroE6 cells. This leads to the definition of an effective MOI, that is the MOI given by the number of particles able to truly infect the Calu-3 host cells. Initially viral particles infect a proportion of the plated Calu-3 cells before being washed away. We use the resulting proportion of infected cells estimated through the standard MOI definition as a starting point of our simulation.

After cells are initialized on the lattice, they are assigned an infected state according to Equation 2, which describes the probability of at least one viral particle entering a cell. In the equations below, \( n \) represents the number of virus particles and MOI is the multiplicity of infection used in the experiment. Equation 2, the probability of a cell being infected by 1 or more virus particles, is derived from Equation 1, the probability of being infected by a single virus particle, commonly used to describe viral infection as a Poisson process [13]. Although the expected MOI is experimentally known, as mentioned in the “Experimental Data” section it is often over-estimated. We add the free parameter INFECTIOUSNESS for two purposes: 1) to scale MOI over-estimation and 2) predict the number of initially infected cells when this data is not available. In the sensitivity analysis we discuss the importance of this parameter.

\[
P(n) = \frac{MOI^n \times e^{-MOI}}{n!} \quad (1)
\]

\[
P(n > 0) = 1 - P(n = 0) = 1 - e^{-INFECTIOUSNESS \times MOI} \quad (2)
\]
3.2.4 Viral release and diffusion

After cells transition to an infectious state virus particles are released with a probability described by a sigmoidal function shown in Figure 4. The probability that one or more infectious viral particles are released in each time step is given by $V(t)$ (Equation 3). In this equation $t$ represents time since the cell was infected and parameters $A$, $B$, and $C$ are derived by fitting an experimentally-derived viral release curve produced by Hui et al. [11]. Since we had no direct measure of the amount of virions per cell released by SARS infected cells we rescaled the sigmoidal function with the free parameter VIR_RELEASE. When $V(t) > 1$ an infectious cell will release one or more virions with a probability of 1. Decimal digits less than one are treated as a probability for an additional virus particle to be released.

Figure 4. Viral Release Function: After cells transition to an infectious state virus particles are released with a probability described by this sigmoidal function. Its equation is shown in Eq. 5 and rescaled by the VIRAL_RELEASE parameter.
Infectious virions diffuse in the virtual well according to a simple random walk: for each virion in the virtual well at each time step of the simulation we perform a number of Random Walk steps given by the free parameter NUM_DIFF_STEPS. Each viral particle has an equal chance to move to one of the six neighboring lattice sites and at each time step performs NUM_DIFF_STEPS movements on the lattice. The need of using a free parameter for the diffusion of virions was given by the lack of data regarding the virion diffusion coefficient on the well with specific conditions of the media used to cover the cell culture. Viral titer measures only infectious particles present in the media using Plaque Forming Units (PFU) per ml. For this reason we model only infectious particles [13]. In addition it is more computationally efficient to release and diffuse only the infectious viral particles.

3.2.5 Second round of infection

Once released, virus particles may infect cells with an uninfected state that are located at the same lattice site. We assume each viral particle independently infects a nearby cell with a probability following the binomial distribution \( P(n) \) (Equation 4) For each lattice site we compute the probability \( P(n) \) that an uninfected cell gets infected by \( n \) viral particles given \( N \) total local virus particles as follows:

\[
P(n) = \binom{N}{n} p_{BP}^n (1 - p_{BP})^{N-n}
\]

where \( p_{BP} \) represents the probability of a virus-receptor binding event leading to a cell’s infection by a single viral particle during a time step. We refer to the free parameter \( p_{BP} \) as BINDING_PROB below. Once a cell is successfully infected, \( n \) viral particles are removed from the virtual media located at the lattice site of the infected cell.
3.2.6 Experimental data

3.2.6.1 Viral Stocks

For these experiments we used a virus derived from our severe acute respiratory syndrome coronavirus (SARS-CoV) wild type infectious clone in which we engineered the green fluorescent protein (GFP) in place of open reading frame 7a/b. SARS-CoV GFP stock titers were calculated using standard plaque tittering methods. Briefly, confluent monolayers of VeroE6 cells in 6 well plates were infected with serial 1:10 dilutions (usually 10e-1 to 10e-6) of stock virus for 1 hour at 37°C. The monolayers were covered with a solution of 0.8% low melting point agarose (Seachem), 1X minimal essential media high glucose (MEM, Invitrogen), 10% fetal clone II (Hyclone), and 1X antibiotic/antimycotic (Invitrogen), which solidifies trapping the virus to allow cell to cell viral spread but not release of virions into the media. Plates were incubated at 37°C for 36 hours, stained with neutral red for 2 to 5 hours, the stain removed and the plaques (holes in the monolayer generated when viruses kill the host cells) are visualized and counted on a light box. Stock titers were calculated as plaque forming units (pfu) per mL.

3.2.6.2 Viral infection

Calu-3 cells were plated at a density of 1x10^6 per well in 6 well plates in MEM containing 10% fetal clone II and 1X antibiotic/antimycotic. Cells were incubated at 37°C at 5% CO2 levels for 2 days prior to infection with a media change 24 hours post plating. SARS-CoV GFP stocks (7.5x10^7) were used to infect each well at a multiplicity of infection of 2 or the addition of 2 infectious virions per cell in each well. At each time point, 100uL of media was harvested from each well to titer via plaque assay. (Number of cells per well assumed to be 2x10^6 at 2 days post plating.)

3.2.6.3 Microscopy

Images shown in Figure 2 were taken using phase contrast microscopy at standard exposure times and at 10X magnification. At the indicated
times post infection, images of the infected and mock-infected Calu-3 cells were taken in living cells real time using a fluorescent light to excite GFP [23]. At each time point, three to five fields of SARS-GFP infected cells were assessed using ImageJ (http://rsbweb.nih.gov/ij/). In ImageJ, we gated the light signal to generate spots in each cell and used the spot counting algorithm to determine the total number of cells per field. GFP positive cells were then counted and averaged.

### 3.2.7 Quantifying model fit to experimental data

For each candidate parameter set \( \theta \) a simulation fitness (Eq. 5) was calculated based on a comparison of two experimental measurements: virus titer and proportion infected cells. Both components of simulation fitness, \( F(\theta) \), are normalized by a maximum error (ME) term to balance their contribution.

\[
F(\theta) = \frac{F_{\text{titer}}}{ME_{\text{titer}}} + \frac{F_{\text{GFP}}}{ME_{\text{GFP}}}
\]

\[
F_{\text{titer}} = \sum \left( \log_{10}(V_{\text{exp}}(t)) - \log_{10}(\alpha \times V_{\text{sim}}(t)) \right)^2
\]

\[
F_{\text{GFP}} = \left| I_{\text{exp}} - I_{\text{sim}} \right|
\]

Randomized weight bootstrapping (1000 iterations) was used to determine MEtiter and the derivation of MEGFP is described below. Equation 6 shows \( F_{\text{titer}} \) is the sum of squares difference between experimentally derived virus titer, \( V_{\text{exp}} \), (Figure 1) and scaled virus titer produced by the simulation, \( V_{\text{sim}} \). Only the last three time points (7, 12, 24 hours) are compared because earlier time point titer values were skewed by residual virus. Simulation output, \( V_{\text{sim}} \), is scaled by \( \alpha \) to account for differences in the number of simulated cells versus the number of in vitro cells. Using a lattice size of 130x130 an average of 7,636 total starting simulated cells were produced. The virus titer was scaled by a factor of \( 10^6/7,636 \) because the experimental virus titer was produced from a population of 106 in vitro cells.
As described in the Experimental Data Section 2.7, in vitro GFP measurements were used to quantify the proportion of infected cells at 12 hours post infection. This additional information was used to constrain our model’s parameter space such as the number of infected cells. The FGFP portion of simulation fitness (Equation 7) is the absolute value of the difference between simulated proportion of infected cells $I_{sim}$ to experimentally measured proportion infected, $I_{exp}$. Since $I_{exp}$ was measured to be .11 (on average 11 of 100 cells are infected) at 12 hours, a maximum error, $ME_{GFP}$, of .89 was used to normalize FGFP between zero and one.

3.3 Results

3.3.1 Sensitivity analysis of free parameters

To assess the importance of model free parameters to simulation outcome, a Latin Hypercube sampling (LHS) sensitivity analysis was performed [14,15]. This two-tiered sampling method was implemented by dividing each free parameter’s range into 1000 intervals. Intervals for each parameter were shuffled without replacement and then randomly sampled using a uniform distribution. All possible probability parameter values were considered for INFECTIOUSNESS AND BINDING_PROB between 0 and 1. Maximum values for NUM_DIFF_STEPS and VIR_RELEASE were arbitrarily chosen (12 and 30 respectively). A Spearman rank correlation was then used to measure statistical dependence, Rho, between free parameter and single run simulation fitness.

The following are free parameters (described above) contained in our model and assessed using LHS:

- **INFECTIOUSNESS** : scales number of infectious viral particles for the initial infection
- **BINDING_PROB** : Probability with which a single virion infects an uninfected cell in a single time step
- **NUM_DIFF_STEPS** : Number of steps for the diffusion of virions at each iteration (random walk)
• VIR_RELEASE: scales the number of virus particles released from each infected cell

Figure 5. Sensitivity Analysis Results: Using Latin Hypercube sampling, we examine the relationship between free parameters and simulation fitness. A significant correlation was identified for INFECTION (A) and VIR_RELEASE (D). Spearman correlations and p-values derived using the AS 89 method are shown. Blue lines indicate Loess regression fitted curve. Armand will update these graphs to specify Rho instead of cor, infectiousness instead of viral_fitness and remove p-value estimates.
Figure 5 shows that INFECTIOUSNESS and VIR_RELEASE were significantly correlated with the simulation fitness while BINDING_PROB and NUM_DIFF_STEPS are not. We see that lower values of INFECTIOUSNESS and VIR_RELEASE result in improved fitnesses whereas larger values result in dramatically poor fitness.

Figure 6. Time-Dependent Parameter Contributions: Dynamics of free parameter correlations with virus titer and proportion of infected cells over time are shown above. (A) INFECTIOUSNESS and VIR_RELEASE parameters are shown to be significantly correlated with virus titer. (B) indicates that INFECTIOUSNESS has a strong correlation with the number of infected cells, but eventually becomes insignificant. Spearman correlations are shown and a grey box indicates a lower quality correlation threshold of |.2|.

We also examined how free parameters affect simulation output with time. Using LHS sampling, we tested relations between free parameters and two simulation outputs: virus titer and % of infected cells. Figure 6A shows that INFECTIOUSNESS and VIR_RELEASE are significant over 24 hours of simulated infection and that their correlations start to diverge at ~12 hours. Again, BINDING_PROB and NUM_DIFF_STEPS do not show a significant correlation with virus titer. However, parameters necessary for the second round of infection do impact the proportion of infected cells shown in Figure 6B
where we see an increase in BINDING_PROB and VIR_RELEASE correlation between 7 and 12 hours, corresponding to a decrease in the INFECTIOUSNESS correlation. This decrease can be attributed to the second round of infection in which a greater population of cells has been infected. The simultaneous increase in correlation of BINDING_PROB and VIR_RELEASE indicates that additional rounds of infection are playing a significant role in viral spread.

3.4 Discussion

Although simplistic compared to in vivo model systems, the interpretation of in vitro experiments is still confounded by biological complexity and disparate data types. Explanatory models are critical for understanding and hypothesis generation. This agent-based modeling framework may be used to investigate first and second round of infection mechanisms using free parameters able to be tuned to allow the model to incorporate disparate types of experimental data. We also take into account spatial aspects of infection, including biases in culture well cell growth and diffusion of infectious virions. Virus titer data and GFP infectivity data from a SARS infection of Calu-3 cells is used as an example to illustrate the model’s capacity to interpret experimentally derived data. LHS sensitivity analysis indicates that a small population of cells is initially infected and that additional rounds of infection are responsible for virus titer measurements.

A significant relationship between INFECTIOUSNESS and both the simulation fitness and simulation outputs (virus titer and proportion infected cells) indicates the importance of this parameter on the resulting infection dynamics. This result also demonstrates the importance of the accurate cell and infectious virus particle counts for the MOI calculation. Finally our model highlights the importance of intracellular processes leading to virus release. One possible future step is to include additional detail regarding intracellular processes of virus replication and move to a multi-scale spatio-temporal model.
Future work is planned to incorporate microarray data and make predictions regarding host response and examine connections between infection status and host response. Simulated annealing has been used to identify free parameters that fit the described model to virus titer data and may be used to predict the number of infected cells or other un-measured data types to support experimental modeling efforts. These off-line predictions could then be used to interpolate a single-cell function representing host response post-infection. We also plan to train the model with data from multiple virus strains to investigate how virus population and host response dynamics differ. Finally, we also plan to investigate the effects of initial spatial distribution of infected cells on viral pathogenesis for multiple virus strains.

References


