Stochastic Modeling of Vesicular Stomatitis Virus (VSV) Growth in Cells
An Application of Order Statistics to Predict Replication Delays

RISHI SRIVASTAVA, SEBASTIAN HENSEL, JOHN YIN and JAMES B. RAWLINGS[a]

Department of Chemical and Biological Engineering, University of Wisconsin-Madison

[a]rawlings@engr.wisc.edu

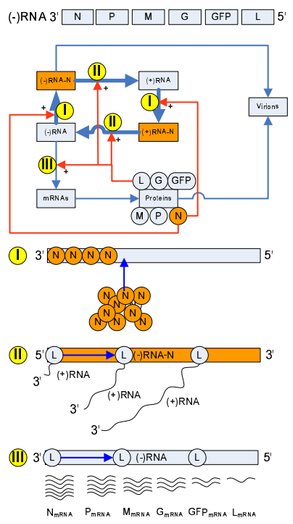
ABSTRACT

Although viruses are the smallest organisms with the shortest genomes, they have major impacts on human health, causing deadly diseases (e.g., influenza, AIDS, cancer) on a global scale. To reproduce, a virus particle must infect a living cell and divert biosynthetic resources toward production of virus components. A better mechanistic understanding of the infection cycle could lead to insights for more effective antiviral strategies. However, simulating the virus infection cycle is a computationally hard problem because some species fluctuate rapidly while others change gradually in number. We study vesicular stomatitis virus, an experimentally accessible virus for which we have developed a deterministic kinetic model of growth (Lim et al, PLoS Comp Bio, 2006). Stochastic simulation of VSV genome encapsidation is a computationally-intensive chain reaction that produces rapid fluctuations in the nucleocapsid(N) protein that associates with the genome. Analytical results from order statistics enable us to avoid explicit tracking of all intermediate species, with significant reduction in computational burden. This approach can be generalized to a broad class of stochastic polymerization reactions where multi-step chain reactions are modeled as a single reaction with time-delay.

Characteristics of VSV

- Single negative strand RNA virus
- Rapid growth but non pathogenic to humans
- Well characterized molecular processes
- Each virus particle carries a single 11-kb (-)RNA genome that encodes five genes, and a recombinant virus (VSV-GFP) encodes an additional gene for green fluorescent protein, enabling tracking of protein expression within infected cells
- Every gene encodes exactly one protein

Overview of the Infection Cycle

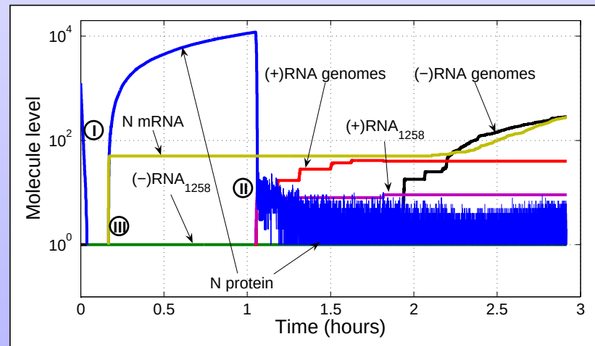


- **I** is Encapsidation of (-)RNA or (+)RNA
- **II** is Replication of encapsidated (-)RNA or (+)RNA
- **III** is formation of mRNAs from naked or partially encapsidated (-)RNA genome

Parts of Infection Cycle and Modeling Approaches

- **Transcription**: Transcribe all genes of the genome into their mRNA's-Modeled as **6** delayed reactions
- **Translation**: Produce viral proteins by translating the produced mRNA's using host cell ribosome-Modeled as **6** parallel non-delayed reactions
- **Replication**: Synthesize a copy of the genome-Modeled as **2516** chain reactions and **2** delayed reactions

A Stochastic Simulation of Infection Cycle: Issues and Challenges



- **Issues**
  - Presence of fast fluctuating N protein
  - Quasi steady state approximation (QSSA)of N protein questionable due to presence of Delayed reactions
- **Challenges**
  - Approximating N protein level without losing the stochasticity
  - Later in the infection cycle L protein also starts switching rapidly due to delayed reactions
  - Approximating L protein level with suitable assumptions

Simulation Methodology

- Different parts of the infection cycle need different simulation techniques

Type of reactions	Algorithm used
With **Delay**	**Delayed** Stochastic Simulation Algorithm (SSA)
Involving Species with **high level**	Langevin Approach
With **rapidly switching** species	SSA with QSSA for that species

- Motivation to develop analytical methods to predict the delays of Transcription and Replication reactions

Genome Encapsidation

Encapsidation process and its representation as a chain reaction

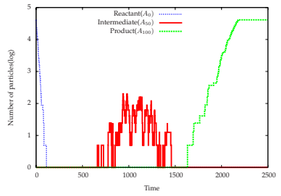
$$(-)RNA + N \xrightarrow{k} (-)RNA_1$$
$$(-)RNA_1 + N \xrightarrow{k} (-)RNA_2$$
$$\cdot \xrightarrow{k} \cdots$$
$$\cdot \xrightarrow{k} \cdots$$
$$(-)RNA_{1257} + N \xrightarrow{k} (-)RNA_{1258}$$

- Two classes of genome fully encapsidated or partially encapsidated
- The fully encapsidated VSV genome serves as a template for replication of the VSV anti-genome.
- Large number of simulations are to be run to get a good stochastic picture of the fully encapsidated genome

The Generalized Chain Reaction

$$A_0 + B \xrightarrow{k} A_1$$
$$A_1 + B \xrightarrow{k} A_2$$
$$\cdots \xrightarrow{k} \cdots$$
$$\cdots \xrightarrow{k} \cdots$$
$$A_{n-2} + B \xrightarrow{k} A_{n-1}$$
$$A_{n-1} + B \xrightarrow{k} A_n$$

A_0:	Reactant
A_i:	Intermediate (for $i \in [1, n-1]$)
A_n:	Product
$\mathbf{N_0}$:	Starting number of particles of A_0
n:	Chain length
Delay Time:	Time at which first particle of product A_n is produced



Semi-log plot of number of particles with time

Space Efficiency and Time efficiency

- Space Efficiency is fraction of stored species that we actually care about
- Time Efficiency is fraction of Stochastic simulation algorithm(SSA) time steps producing product A_n
- Space Efficiency of SSA is O($1/n$)
- Time Efficiency of SSA is O($1/n$)
- Time of Simulation is O(n^2)

Analytical Solution for Product Distribution

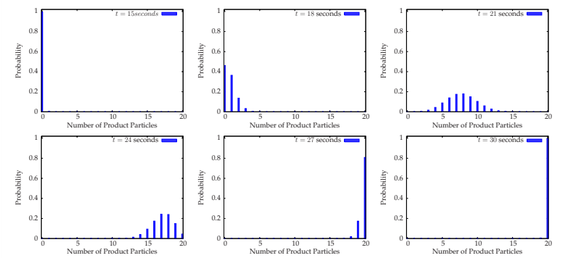
Use of Binomial and Gamma Distribution

Main result

$$P(N_p = j) = \binom{N_0}{j} \mathbf{p}(\mathbf{t})^j \left(1 - \mathbf{p}(\mathbf{t})\right)^{N_0 - j}$$
$$j \in [0, N_0]$$

- N_p: Number of product particles
- $P(N_p = j)$: Probability of having exactly j product particles
- $\mathbf{p}(\mathbf{t})$: Probability that a particular reactant molecule has converted into product by time t (**CDF** of a Gamma distribution)
- An approximation of this binomial distribution to Poisson or Normal is possible if we have large number of reactant particles

Analytical Prediction of Product Particle Distribution



Comparison of computation time[a] to get product distribution

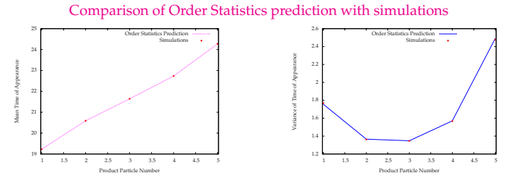
Chain Length (n)	Using SSA (10,000 runs)	Time per run (of SSA)	Analytically (from Main result)
10	9m	0.054s	0.12s
50	2h 29m	0.89s	0.12s
100	8h 30m	3.06s	0.12s
150	21h 18m	7.70s	0.13s
200	34h 4m	12.30s	0.14s
1258	302h 20m	109s	0.14s

[a]All computations were done on AMD Athlon Dual core processor with clock speed 2.2 GHz, L1 cache memory 256KB and L2 cache memory 2MB using Octave

Use of Order Statistics to Predict Time Delays

- Delay time for appearance of first particle of our desired product can be obtained using order statistics analysis
- In general case a j^{th} order statistic can be used where we are interested in the distribution of time of appearance of an specified number of product particles

Comparison of Order Statistics prediction with simulations



Looking Ahead

- Incorporation of this analytical approach in complete infection cycle simulation
- Extending this approach to the case where the second reactive species (N Protein in our case)is not deterministic
- Use of Order Statistics to tune characteristic parameters of chain reactions
- Verification of this approach with experiments

Acknowledgments

This work was supported by an NIH Phased Innovation Award (R21 AI071197), an NSF Focused Research Group Award (DMS-0553687) and the German Academic Exchange Service (DAAD).