Kinetics of Vesicular Stomatitis Virus mRNA and Genomes During Infection

By

Collin M. Timm

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemical and Biological Engineering)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 6/26/2013

The dissertation is approved by the following members of the Final Oral Committee:

John Yin, Chemical and Biological Engineering
James Rawlings, Chemical and Biological Engineering
Sean Palecek, Chemical and Biological Engineering
Paul Ahlquist, Oncology and Molecular Virology
Mark Craven, Biostatistics and Medical Informatics

Acknowledgements

This work would not be the same without the contributions and support of many people. I am grateful for the support of my family and friends and the guidance of my peers and advisors.

Throughout my life I have been supported and challenged by my parents and sister. Without their guidance, excitement, and work ethic I would not be where I am today. My parents bought me my first microscope, my first chemistry set, and taught me about negative numbers in the first grade. I am deeply grateful for the support they gave me and their high expectations throughout all of my schooling. My father's attention to detail and "do it right the first time" attitude coupled with my mother's "work hard, play hard" style is the basis of my success in the academic world. My sister was always a source of competition: her unique approaches and skills always kept me asking how problems could be approached or solved differently. And the rest of my family (Grandman, Carole and Tim Conama, Felicia and Jim Paulzine, Chris and David Timm, all the other extend family) for always being proud of me and not being afraid to poke fun or challenge me, which has helped keep me modest and working hard.

A more recent but extremely impactful addition to my life is my beautiful and insightful wife: Andrea. Andrea has provided emotional and academic support throughout graduate school. Her encouragement and grounded advice has helped me focus and develop ideas into real world applications. Without Andrea, this thesis would certainly be very different. You really have changed me and my view on the world for the better.

I also acknowledge the Department of Chemical and Biological Engineering at UW Madison. In this department I have made friendships guaranteed to last long into the future. My friends in the department have been a familiar group who intimately understand the difficulties that come with graduate school. My interactions with both faculty and graduate students have helped shape my academic career and goals for the future. Andrea, Claudio, Josh, Derek, Jeff, Joe and Becky you have been great sources of inspiration and relief.

I am grateful for the input and questions from the Yin Group: John Yin, Musarat Ishaq, Kathy Vielhuber, Stephen Lindsay, Emily Voigt, Andrea Timm, Bahar Inankur, Greg Potts and Ashley Baltes have always kept meetings interesting by asking unique and challenging questions and helping to bring the research to the next level. I learned many important lessons, both professional and lifelong, that have helped me develop the experiments and conclusions present in this thesis. I must especially acknowledge Adam Swick, Jay Warrick, Fulya Akpinar, and Ankur Gupta for challenging me and working with me on multiple projects.

I am deeply indebted to John for setting an example as a successful researcher. John always asked the next question and made sure we always had something new to consider. One of his many lessons that has stuck with me: “When you recognize good writing, it doesn’t matter if it is about sports or politics, study it and learn what makes it is that makes it so good”. I apply this lesson not only to writing, but other skills I wish to develop in my life. His teaching and impact as a role model will always be with me as a work to develop my career.

I also thank my committee for their unique viewpoints and advice on this document and my work presented here. While our interactions were infrequent, your input really has helped me motivate and advance my project.

This work was funded in part by an NLM Training Grant (NLM 5T15LM007359) through the Computation and Informatics in Biology and Medicine (CIBM) program at UW Madison. This program was a great way to interact with computer scientists and statisticians and gave me a different perspective from my home department of Chemical and Biological Engineering.

I am grateful to all you, and hope I can return as much to you as you have given me over the years!

Abstract

An approach to combating viral infections is to rationally design treatments to target sensitive mechanisms in the viral infection. Quantitative, predictive models of the viral infection are excellent tools for evaluating the effect of a small perturbation to a mechanism on the overall infection cycle. However, quantitative models must be validated with quantitative data from an infection.

Vesicular stomatitis virus has broad cell tropism in the laboratory. In order for the virus to be successful across the wide range of cell types it can infect, there must be robust mechanisms that are conserved between infections. In this work I investigated the kinetics of VSV mRNA and genomes during infections under multiple conditions. Using qRT-PCR to measure viral RNA and ODE modeling to analyze the data I showed that viral mRNA production is linearly related to genomes during an infection, and that the rate constant of production is similar across many experimental conditions. These results suggest that the viral mRNA production depends only on genomes; mRNA production rate is independent of translation, genome replication and host cell environment. This conserved relationship is likely one strategy that allows VSV to infect a wide range of cell types. If other viruses show similar conserved, quantitative relationships then those relationships are potential targets for anti-viral drugs or treatments.

To investigate the quantitative features of VSV infection I also developed a mechanistic and modifiable mathematical model of VSV infection. Using known VSV mechanisms, new data, and literature values for other parameters the model is able to match observed experimental trends. By parameterizing the model to wild-type infection data, the model was able to qualitatively match experimentally observed trends of two fluorescent strains of VSV. The

model also predicts that VSV N protein is very efficiently used: all of the N protein in the simulation is consumed for genome replication.

Natural viral infections occur over multiple generations. Under high multiplicity of infection conditions VSV spontaneously forms defective interfering particles (DIPs). These particles have the replication promoters but are lacking a major portion of the VSV genome which renders them non-infective. However, when co-infected with normal VSV the DIPs replicate and sequester viral proteins, an interfering process that can completely inhibit VSV production in cells. This interference is a potential antiviral strategy that we can engineer. Because the DIPs use the viral replication promoters, it is less likely that the virus will be able to evolve away from such an antiviral treatment. We used RNA sequencing to examine the population dynamics of VSV and its defective interfering particles. This technology was shown to be unbiased in its detection, identification and quantification of VSV population makeup. Using complementary measures of qRT-PCR, TEM, and biological activity we examined the population of VSV infectious and defective particle growth. The study suggests that nearly all defective particles have interfering activity. The RNA sequencing technology will be an excellent addition to virus-host interaction kinetics studies as the cost of the technology decreases.

The quantitative approaches presented here taught us new biology about VSV. I identified that mRNA production is linearly related to the number of genomes in the cells, independent of multiple environmental conditions. A mechanistic model of VSV replication was developed to provide a modifiable simulation to complement existing experimental techniques. Finally, I examined the ability of RNA sequencing as an unbiased and quantitative tool to characterize dynamics of VSV and DIP populations. The quantitative methods presented here can also be

applied in other viral systems to help determine conserved relationships between viral intermediates.

Table of Contents

Acknowledgements ii

Abstract v

List of Figures xv

Chapter 1: Motivation and Background 1

1.1 Motivation 1

1.1.1 Virus infection modeling 2

1.1.2 Modeling VSV 3

1.1.3 VSV growth over multiple generations 4

1.2 Thesis overview 5

1.3 Background 7

1.3.1 One-step infection 7

1.3.2 Reverse transcription (RT) 8

1.3.3 qPCR 9

1.3.4 Chemical reaction kinetics and ODE modeling 10

1.4 References 12

Chapter 2: Kinetics of transcription and replication for an RNA virus: from simple to complex 15

2.1 Abstract 15

2.2 Introduction 16

2.3 Results 18

2.3.1 Development of a qRT-PCR assay to measure the kinetics of viral N mRNA and genomes during an infection 18

2.3.2 Rate of mRNA production depends linearly on genome levels in the absence of translation 19

2.3.3 Rate of mRNA production depends linearly on genome levels in the presence of translation and replication 20

2.3.4 Primary is slower than secondary mRNA production rate in PC3 cells 21

2.3.5 Stimulation of the innate response by IFN does not affect the dependence of mRNA production on genome levels 22

2.4 Discussion 24

2.5 Supplemental Figures 28

2.6 Acknowledgements 33

2.7 Supplemental Information: Model Development 34

2.8 Methods 42

2.8.1 Cell Culture 42

2.8.2 Virus Culture 42

2.8.3 Virus Infections 42

2.8.4 Cycloheximide treatments 43

2.8.5 IFN treatments 43

2.8.6 RNA Extractions 43

2.8.7 qPCR 44

2.8.8 qRT-PCR Assay Optimization 45

2.8.9 Conversion to number per cell 45

2.8.10 Model Solution and Data Fitting 46

2.8.11 Significance testing for model parameters 47

2.9 References 48

Chapter 3: A mechanistic model of vesicular stomatitis infection 50

3.1 Abstract 50

3.2 Introduction 51

3.3 Results 52

3.3.1 Model formulation 53

3.3.2 Primary Transcription 55

3.3.3 Translation 57

3.3.4 Replication 58

3.3.5 Viral detection and host shut down 60

3.3.6 Late viral processes and packaging 61

3.3.7 Multiplicity of infection 63

3.3.8 Predicting gene insertion at different locations 64

3.4 Discussion 65

3.5 Supplemental Figures 68

3.6 Acknowledgements 72

3.7 Methods 72

3.7.1 Reaction formulation 72

3.7.2 Maximal loading rate determined by spacing on template 72

3.7.3 Delayed ODE solution 73

3.7.4 Model structure 73

3.7.5 Reaction List 74

3.7.6 Algorithm pseudocode 75

3.7.7 Cell Culture 76

3.7.8 Virus Culture 76

3.7.9 Virus Infections 76

3.7.10 RNA Extractions 77

3.7.11 qPCR 77

3.7.12 qRT-PCR Assay Optimization 78

3.7.13 Conversion to number per cell 78

3.8 References 80

Chapter 4: Quantitative characterization of defective virus emergence by deep sequencing .. 83

4.1 Abstract 83

4.2 Introduction 84

4.3 Results 85

4.3.1 Deep sequencing can detect differential coverage profiles in related populations ... 85

4.3.2 Normalizing coverage to the initial population coverage reduced the noise in coverage profiles ... 87

4.3.3 Coverage analysis reveals two distinct genome deletion mutants in the population 90

4.3.4 *De novo* read assembly suggests a genome deletion mutant ... 90

4.3.5 Relative coverage levels were used to quantify population fraction of full length and genome deletion mutants ... 91

4.3.6 Complementary measures agree with quantification by relative coverage ... 91

4.3.7 Quantification by relative coverage is limited by statistical and empirical measurement noise ... 93

4.3.8 Absolute quantification of population suggests a strong correlation between genome structure and viral activity ... 93

4.4 Discussion ... 95

4.5 Supplemental Figures and Tables ... 98

4.6 Methods ... 106

4.6.1 Cell culture ... 106

4.6.2 Preparation of viral passages at fixed MOI ... 106

4.6.3 Plaque Assay ... 107

4.6.4 Interference Assay ... 107

4.6.5 RNA Extraction ... 108

4.6.6 RNA Sequencing ... 108

4.6.7 Sequence alignment and P0 consensus determination 108

4.6.8 Quantification and analysis of coverage data 109

4.6.9 qRT-PCR 109

4.6.10 Particle Quantification by Transmission Electron Microscopy 110

4.7 Acknowledgements 110

4.8 References 112

Chapter 5: Summary and Conclusions 117

5.1 Quantification of VSV processes reveals new biology 117

5.1.1 VSV transcription is captured with a single reaction 117

5.1.2 The rate constant for VSV transcription is similar between many conditions 118

5.1.3 A simple model captured VSV kinetic data 119

5.2 Simulation of VSV using quantitative data generates new hypotheses 119

5.2.1 Model predictions 120

5.2.2 Model flexibility 120

5.3 Usable and modifiable virus models 121

5.4 Quantification of VSV population changes 121

5.5 Quantitative biology as a complement to molecular mechanisms 122

Chapter 6: Future work 123

6.1 Introduction 123

6.2 Experimental work 123

6.2.1 Longevity of linear transcription suggests active recycling of polymerases 123

6.2.2 VSV transcription is linear throughout replication, suggesting that the majority of genomes are used as templates for transcription rather than genome replication..... 126

6.2.3 What limits VSV transcription? 127

6.2.4 Quantification of VSV proteins by coupling qRT-PCR and mass spectrometry.... 129

6.3 Virus infection modeling work 133

6.3.1 Evolution of virus reaction network and application to N1-N4 mutants 133

6.3.2 Localized models 135

6.3.3 Innate response models 135

6.4 References 137

List of Figures

Figure 1.3-1 Schematic of a one-step infection. The left shows a normal culture of cells which has an average virus per cell (MOI) added to infect all cells. The right shows data of the production of infectious virions (PFU=plaque forming units) from a population of cells over time. 8

Figure 1.3-2 qPCR workflow schematic. DNA or cDNA samples are loaded into wells, amplified and monitored for fluorescence, and quantified by comparing to the known levels in the standard curve (10-fold dilutions) 10

Figure 1.3-3Genome and mRNA data for modeling example. The data is from Chapter 2, Figure 1, medium MOI infection. Genomes are shown as open symbols, and mRNA as closed symbols. Axes are time, in hours post-infection and number per cell, an average value measured from ~10^5 cells. Replication is not observed because translation is inhibited through treatment with cycloheximide. 11

Figure 2.3-1 **VSV N mRNA and genome kinetics in cycloheximide treated BHK cells**. BHK cells were treated with 50 ug/mL cycloheximide. Medium and low MOI are 10 and 100 fold dilutions of high MOI samples, respectively. Samples were collected in duplicate as described in Methods. A) mRNA and genome data from cycloheximide treated cells. B) model reactions and differential equation solutions. Parameter Tables: k_t is the transcription constant (mRNA/genome/hr), k_d is the genome degradation constant (hr^{-1}), and IC is the initial condition (genomes). 19

Figure 2.3-2 **VSV N mRNA and genome kinetics in BHK cells**. Medium and low MOI are 10 and 100 fold dilutions of high MOI samples, respectively. Samples were collected in duplicate as described in Methods. A) mRNA and genome data from cycloheximide treated cells. B) model

reactions and differential equation solutions. Parameter Tables: k_t is the transcription constant (mRNA/genome/hr), τ is the time delay before replication, k_r is the genome replication constant (hr^{-1}), k_d is the genome degradation constant ($genomes^{-1}hr^{-1}$), and IC is the initial condition (genomes)..... 20

Figure 2.3-3 **VSV N mRNA and genome kinetics in PC3 cells.** Medium and low MOI are 10 and 100 fold dilutions of high MOI samples, respectively. Samples were collected in duplicate as described in Methods. A) mRNA and genome data from cycloheximide treated cells. B) mRNA and genome data from untreated cells. Parameter Tables: k_t is the transcription constant (mRNA/genome/hr), τ is the time delay before replication, k_r is the genome replication constant (hr^{-1}), k_d is the genome degradation constant ($genomes^{-1}hr^{-1}$), and IC is the initial condition (genomes)..... 21

Figure 2.3-4 **VSV N mRNA and genome kinetics in IFN stimulated or untreated PC3 cells.** PC3 cells were treated with 1000 Units/mL human IFN-beta 1 hour before infection, and co-treated; or not treated. Medium and low MOI are 10 and 100 fold dilutions of high MOI samples, respectively. Samples were collected in duplicated as described in Methods. A) N mRNA and genome measures and model for untreated PC3 cells B) N mRNA and genome measures for IFN treated PC3 cells Parameter Tables: k_t is the transcription constant (mRNA/genome/hr), τ is the time delay before replication, k_r is the genome replication constant (hr^{-1}), k_d is the genome degradation constant ($genomes^{-1}hr^{-1}$), and IC is the initial condition (genomes)..... 23

Figure 2.5-1 **Reverse transcription validation.** Infected cell RNA (harvested at 6 hpi) was diluted into mock-infected cell RNA. Parallel RT reactions were performed to show the linearity and lower limit of detection of the RT step. Fit lines are included for the linear detection region. Mock lines (dashed) show higher qPCR signals than some of the dilutions, but samples below

LOD were distinguished by melt curves. Mock samples were included in all experiments and data below mock values was omitted from analysis. 28

Figure 2.5-2 **Reverse sampling is equivalent to traditional sampling.** BHK cells were infected at MOI ~5 in multiple 6 well plates (traditional) or using reverse method in 96 well plates (reverse). Cells were sampled serially and stored in RNAProtect (traditional) or simultaneously (reverse) and N mRNA was counted by qRT-PCR method. As shown, the results give similar trajectories. The reverse sampling is much more high throughput allowing for 96 samples while traditional sampling is much more labor intensive. 29

Figure 2.5-3 **Linear scale plots of N mRNA per cell for cycloheximide treatments**. **A.** The three MOI treatments on BHK cells are plotted together to show the effect of incoming MOI on primary transcription. (high, medium, low = 160, 16, 1.6 PFU/cell). The lines show the model fit as shown in Figure 2. The line is not linear because during primary transcription genomes are degrading. **B.** Individual plots for each MOI show data and model fit. Note: the y-axis scale is different between plots. **C-D.** Same as A-B, but on PC3 cells 30

Figure 2.5-4 **VSV N mRNA and genome kinetics in cycloheximide treated PC3cells**. PC3 cells were treated with 50 ug/mL cycloheximide. Medium and low MOI are 10 and 100 fold dilutions of high MOI samples, respectively. Samples were collected in duplicate as described in Methods. A) mRNA and genome data from cycloheximide treated cells. B) model reactions and differential equation solutions. Parameter Tables: k_t is the transcription constant (mRNA/genome/hr), k_d is the genome degradation constant (hr^{-1}), and IC is the initial condition (genomes). 31

Figure 2.5-5 **IFN treated measures relative to un-treated infection.** Data from Figure 5 was normalized to the no IFN case and averaged over the duplicate measures. **A.** relative genome

levels. **B.** relative mRNA levels. Notice that the medium and low MOI are inhibited to the same level, but the high MOI ratio stays closer to 1, indicating that the high MOI does a better job at overcoming the immune response as compared to medium or low MOI. 32

Figure 3.3-1 **The VSV genome contains all information for transcription, translation, and packaging of the virus.** The attenuation between genes determines the relative mRNA production and the availability of polymerases. The number of nucleotides and polymerase elongation rate are used to calculate production delays for mRNA using t_{delay}=(gene length, nt) / (pol. rate, nt/s). Similarly, delays are calculated for protein translation. The genome acts as a reactant for transcript production, mRNA acts a reactants for translation, and all information is referenced back to a single viral genome. 54

Figure 3.3-2 Model prediction of primary transcription for VSV in a BHK cell environment. ... 57

Figure 3.3-3 Viral translation leads to a switch in polymerase activity to replication. A. Genome and mRNA data are shown as points, and simulation as lines. Data are collected from multiple infections as described in Methods. The lines are higher than the data, as we have not yet included mechanisms which inhibit replication. B. The fraction of total N protein that is consumed during infection. 59

Figure 3.3-4 Detection of viral genomes and inactivation of ribosomes in the cell. The left plot shows the number of detection molecules growing over time. The detection molecules interact with ribosomes to inactivate them (right plot). 61

Figure 3.3-5 Full simulation of VSV infection produces virions. 62

Figure 3.3-6 Different MOI approach the same genome levels 63

Figure 3.3-7 Model predicts RFP1 virus will be brighter than RFP5, but produce fewer infectious virions. 65

Figure 3.5-1Generic pool structure used to account for infection intermediates. Items that are visible to the simulation are included in “Available Items”. Newly produced items may be delayed and are stored in “Delayed Items”. Finally, the “Rates” structure is used to account for rates of change of intermediates. 68

Figure 3.5-2 We simulate a cell containing distinct pools with specific contents. At every time step, the defined reactions are used to calculate rates of changes based on concentrations defined by the rate calculation. The arrows show the interactions that occur in a simulation of VSV infection. For example, templates are used to make transcripts or can be combined with proteins to produce virions. Virions are created in the simulation but not used to modify other pools. 69

Figure 3.5-3Production of VSV at the end of infection. In BHK cells the production is independent of MOI, while in PC3 higher MOI is able to produce more particles at 24 hours. .. 70

Figure 3.5-4 Simulated data used to calculate predicted percent of viral protein remaining in cell. We calculated how much protein was left after all particels were produced. Note that N protein is essentially 0.0. After that, P proteins (466 molecules per virion) are the most used proteins for producing viriosn, relative to the production level. 71

Figure 4.3-1 **Serial passage culture of VSV.** Virus was cultured on BHK-21 cells at MOI 10 for 3 passages, established conditions that promote the emergence of DIPs. RNA extracted from each passage was submitted for RNA sequencing. 86

Figure 4.3-2 **Raw coverage data.** Coverage is plotted on y axis vs. position on the viral genome on the x-axis (genome cartoon below x-axis). Genome is from 3’ to 5’ with base 1 representing the first base on the 3’ end. The P3 plot has two sequencing replicates, one in red and one in black. Notice the relatively flat coverage in P0 and the distinct plateaus that are seen in later passages..... 87

Figure 4.3-3 **Normalization by P0 coverage reduces variability** P1 coverage values are divided to P0 coverages at the corresponding positions shifted by an offset. **A)** The plots show the normalized P0 (blue) and P1(green) coverages with increasing offset from top to bottom. The x-axis represents the nucleotide positions on VSV genome. **B)** As the reference P0 coverages shifted by a higher offset, the noise in data increases, which can be seen in the increase of coefficient of variance (COV, standard deviation of normalized coverages divided by the coverage mean across genome) of P1. The COV approaches its maximum after about 30 bases. Similar behavior was observed when this analysis was performed on P2 and P3..... 88

Figure 4.3-4 **Comparison of all quantification techniques. A)** Coverage values from deep sequencing were normalized to the corresponding base in P0, ignoring any zero values. The technical replicate of P3 sample is shown as the red curve. **B-E)** Population fractions are calculated for each method. The black segment corresponds to full size species, the gray in the large pie represent all defective particles. In **B** and **C** a second pie chart shows the fractions of medium and short genomes in the defective population of P2 and P3. Values reported are percent of population ± standard deviation. **B)** Fraction calculated from sequencing coverage. The standard deviation was calculated by coverage noise and propagation of uncertainty. **C)** Fractions by qRT-PCR. Standard deviation calculated by 2 replicates of extraction – qPCR workflow. **D)** Fraction calculated from TEM images. Standard deviation calculated from counts of two sets of ~15 images from each passage. **E)** Fractions calculated from plaque forming units and interfering units measured by plaque and interference assays. Standard deviation calculated from duplicate measures. 89

Figure 4.3-5 **Absolute virus and DIP concentrations using different methods** Genomes were measured by qRT-PCR, particle by TEM and activity by plaque assay or interference assay. A)

Full length genomes, particles and infectious units decline at a similar rate. B) Defective genomes, particles and interfering units follow a similar trend. C) Total genomes, particles and activity follow similar trends. Note that in P2 and P3 the counts are dominated by defective particles. 94

Figure 4.5-1 **Plateau boundary determination** 98

Figure 4.5-2 **Contig detected by *de novo* alignment has complementary ends and similarity to full length virus.** The *de novo* alignment tool in CLCBio was used with default parameters on P3 reads to generate this contig. The first base of the contig aligns with the 5,116 base of the parental genome (P0). There is a short segment where there is no alignment, then a long stretch of perfect match. Finally, the DIP sequence aligns to the end but is truncated before the final base of the parent sequence. 99

Figure 4.5-3 **Calculation of population fractions from coverage data.** Data from P3 is shown. The coverage at the right contains coverage from short, medium and full length genomes. We used the equations above to determine the fraction (f) of the population corresponding to full, medium, or short length genomes. These equations were also used with propagation of uncertainty rules (assuming no covariance) to determine standard deviations on the population fraction estimates. 100

Figure 4.5-4 **TEM Image of P3 Virus and DIPs.** In this image we see cellular debris, full-length VSV particles and short defective particles. The scale bar is 200nm in the lower right corner of the image. 101

Figure 4.5-5 **Limit of detection calculated two ways.** 102

Figure 4.5-6 **Mapped reads do not correlate with other absolute measures**. The total number of mapped reads (x-axis) were plotted against biological activity per mL, genomes per mL (by

qRT-PCR), or particles per mL. We do not see a strong correlation between mapped reads and any of the absolute measurements. 103

Figure 6.2-1 A polyribosome. (A) Schematic drawing showing how a series of ribosomes can simultaneously translate the same eucaryotic mRNA molecule. (B) Electron micrograph of a polyribosome from a eucaryotic cell. (B, courtesy of John Heuser.) 124

Figure 6.2-2 Potential structure of VSV genome during transcription. Top: unfolded genome. Bottom: This structure would allow for close proximity of polymerase 'exit sites' and the 3' transcriptional promoter, allowing for active recycling of polymerases. 125

Figure 6.2-3 VSV particles were serially diluted 1:5 and adsorbed to PC3 cells. **A.** Genome measures. The points are the data, and the line is the theoretical 1:5 line from the most concentrated sample. **B.** VSV N protein counts plotted against VSV genomes at that dilution. **C.** VSV P protein vs. genome counts. **D.** VSV M **E.** VSV G **F.** VSV L 131

Chapter 1: Motivation and Background

1.1 Motivation

Examples of major diseases linked to viral infections include AIDS, influenza, the common cold, and a number of cancers. For many of the diseases caused by viral infections, there is little we can do to stop the virus once the infection is established. In the case of acute infections like the cold or flu, the solution is usually to wait until the immune system is able to clear the virus. In the case of chronic infections like AIDS or cancers, the infection is slowed by drugs or treatments but usually the result is death.

Viral infections are transmitted between individuals by virion particles that carry the genome of the virus. The particle itself serves to protect the genetic code of the virus, as well as trigger delivery of the genome to the cell the virus will infect. After genome delivery the virus hijacks cell resources to produce viral proteins and replicate genomes. The viral proteins are used both to further reprogram the cell's normal function and to produce new particles that can transport the genomes to new, uninfected cells. It is important to note that the virus cannot replicate without infecting a healthy cell. The virus is tightly coupled to the function of the cell and a study of a virus infection cannot neglect the presence of the host cell.

Viruses are often classified by the type of genome they carry: either DNA or RNA, which may be either double-stranded or single-stranded. RNA viruses are further classified as either positive-sense (mRNA sense, can be translated directly into proteins) or negative-sense (acts as template for producing mRNA). Another distinction is whether the virus particle is built with a lipid envelope inherited from the infected cell, or if the structure is purely made up of structural proteins.

In this work I focus on vesicular stomatitis virus (VSV). VSV is a member of the *Mononegavirales* order which includes the Ebola virus, measles virus and rabies virus. All members of *Mononegavirales* have a single-stranded, negative-sense, RNA genome and form enveloped particles. While VSV does not cause human disease, it does cause important diseases in livestock. VSV has also been widely studied as a potential anti-cancer therapy because it is able to infect almost every cell type (1, 2).

In order to help develop treatments and drugs for viral infection researchers attempt to discover vulnerable mechanisms of the infection. The goal is to determine specific functions that the virus uses, independent of the host cell, and then target those mechanisms with drugs or other treatments. In theory, because the mechanisms are independent of the cell, the treatment will not have undesired side effects on the host cell. The mechanism discovery approach has been applied widely and we know a great deal about the mechanisms many viruses use to infect and reprogram cells.

1.1.1 Virus infection modeling

The first substantial *in vivo* computational models of virus infection were used to help understand the replication dynamics of human immunodeficiency virus (HIV) in infected individuals (3). The so-called SIR models (Susceptible-Infected-Resistant, referring to measured states in the model) were previously used in epidemiological studies (4) but showed to be informative for predicting viral clearance rates and other lumped parameters in HIV (3, 5, 6), herpes virus (7), and hepatitis C virus (HCV) systems (8–11). These lumped models have shown to be very useful in the clinical setting and are still widely applied today.

Another approach to modeling virus infections is to create a reaction network that incorporates all known mechanisms of the infection cycle. These models serve multiple purposes:

1) The models validate the infection network by showing that all the included mechanisms can work together
2) The models force us to ask even more questions about the details of the infection
3) The models serve as predictive tools to help estimate sensitive aspects of the infection and the overall effect of perturbing specific mechanisms

Mechanistic intracellular models were developed first for bacteriophage infections and were used to investigate system dynamics (12, 13) or energetic requirements of the phage infection (14). These works confirmed that phage biology can be simulated *in silico*, as well as provided interesting and important implication on the metabolic needs of the phage during replication. Models of intracellular phage replication are continually explored with the goal of engineering phage as anti-bacterial treatments (15).

Intracellular mechanistic models have also been extended to virus infections of mammalian hosts. Influenza modeling has shown that the infection cycle is limited by production rate of matrix protein and that the virus has little effect on the bulk availability of nucleotides or amino acids (16). A model of poliovirus replication suggests that the poliovirus replication strategy, specifically the ratio of genomes to anti-genomes, is near the evolutionary optimum (17). These types of models have also been used to determine nucleotide substitution rates in RNA viruses and the effect the substitution rate has on genome evolution (18, 19). The main goal of using mechanistic intracellular models is to help identify sensitive aspects of the infection and determine how they affect the integrated infection reaction network. This is important for the rational design of anti-viral treatments or for engineering virus growth for vaccine production. In the next section I discuss models of VSV infection and their implications.

1.1.2 Modeling VSV

Vesicular stomatitis virus has been especially well studied because of its ability to infect nearly all cells in culture and to grow to high titers in the lab. These traits made it an ideal virus to study as molecular biology and cell culture methods were developing. Because VSV was so historically well studied, we understand many of the mechanisms the virus uses during infection.

An early strategy for modeling VSV incorporated all known VSV mechanisms into an "average cell" model (20). Most of the available data is measured as a population average. Differential equation models can be thought of as "average" because they allow for fractional genomes and other intermediates in the cell. For the differential equation model the known VSV interactions were modeled as chemical reactions, and mass action kinetics were used to develop differential equations to describe how the concentration of intermediates changes during an infection. Another major strategy used was to account for delays in infection by modeling delays as chained reactions. For example, the VSV genome is 11,161 nucleotides and takes roughly one hour to synthesize. In this model the genome was modeled as a series of 40 reactions where a new VSV genome was the end product. This model showed that the translational resources of the host are likely what limit the production of virus particles.

A second strategy to model viral infections is to use stochastic modeling (21). A stochastic model more closely represents what happens in a single cell as it accounts for integer numbers of molecules and random fluctuations in the environment. It is very common that a virus infection be initiated with small particle numbers that would be sensitive to random fluctuations. In this work the model predicted that early genome and protein expression would be tightly clustered, such that subsets of cells would behave in similar groups, as opposed to a continuous, uni-modal distribution of genomes and proteins.

1.1.3 VSV growth over multiple generations

Until now we have discussed the growth of virus at the single cycle level, or how the virus infects and grows through a single generation. In natural infections of tissues the virus is spreading between cells and sustains growth over multiple generations. Therefore, it is important to consider how the virus population may change over the course of an infection.

VSV has been shown to have dramatic population changes under specific growth conditions. When VSV is passaged at high multiplicity of infection (>10 particles per cell), defective interfering (DIPs) particles can form (22). DIPs are genome deletion mutants: they have replication promoters but are missing a significant portion of the coding sequence in the genome. The deletion renders the particles non-infective. However, the presence of the replication promoters allows the truncated genome to replicate when sufficient viral proteins are present. When DIPs co-infect with a full length VSV genome they sequester proteins created by the full length virus and thus interfere with the normal replication cycle. DIPs have been observed in natural viral infections (23, 24) and are hypothesized to act as a regulatory response to a virus that is growing rapidly in an organism.

The interference by DIPs can be so extreme that it completely shuts down normal virus progression (25). Because of this these particles have been proposed as potential live vaccines: the carry the viral sequence (in the promoters) and thus may be more resistant to viral escape by mutation.

1.2 Thesis overview

Both VSV models (20, 21) depended on predicting absolute mRNA and protein levels within a cell to describe production of particles from cells. Little data exists to validate the model predictions of intracellular intermediates. The goal of this work was to obtain quantitative estimates of the number of VSV genomes and mRNA during an infection, then apply those in a

mechanistic modeling framework. In Chapter 2 I describe how I developed a qRT-PCR assay and observed the kinetics of viral gene expression and genome replication under multiple experimental conditions. In addition to being the first study to determine the absolute relationship between mRNA and genomes, the data collected here can be used to validate complex models of VSV infection.

The quantitative measurements of viral mRNA and genomes were used to help develop a delayed differential equation model similar to the one described above. However, there are some differences and the model is described in detail in Chapter 3. The most notable difference was the simulation strategy. My co-author Dr. Jay Warrick and I treated the viral mechanisms as chemical reactions, but we used complex reactions that may contain a variety of information to help simplify the simulation. For example, a viral genome is the primary reactant for transcription, but in our case the genome also is linked to the information about the number of genes, the protein products, and the length to help calculate delays and further products in the infection. This simulation strategy is similar to the natural system, in that the viral genome has all the necessary "information" to infect a cell and produce thousands of progeny virions.

A single cycle of virus infection may not be enough to fully understand the impacts of virus-caused disease. In Chapter 4 we use the powerful technique of high throughput RNA sequencing applied to virus populations as they change over serial passages. We use RNA sequencing coverage data to detect, identify, and quantify genome deletion mutants as they grow in the population. While this technology is powerful and quantitative, it is currently too expensive for detailed kinetic experiments.

This work serves as an advancement to the biology of VSV and to the field of quantitative biology. In Chapter 5 I summarize the results of this work, and in Chapter 6 I discuss future directions for the described projects.

1.3 Background

This section will give a brief background on the experimental and modeling techniques used in the thesis.

1.3.1 One-step infection

A one-step infection is an infection in cell culture where the goal is to infect all cells in a sample simultaneously and with the same amount of virus. The term “one-step” refers to the fact that all cells are infected simultaneously, meaning we only will observe one cycle of the infection process. The opposite of a one-step infection is a low multiplicity infection, where only a small fraction of cells in a population are infected at time zero. In a low multiplicity infection several rounds of replication may be required to infect every cell in the plate.

One-step infections are traditionally used to get an average measure from a large population of cells. Many assays require a large amount of starting material to detect viral activity. Figure 1.3-1 shows the schematic for a one step infection where output virus (PFU) is measured over time.

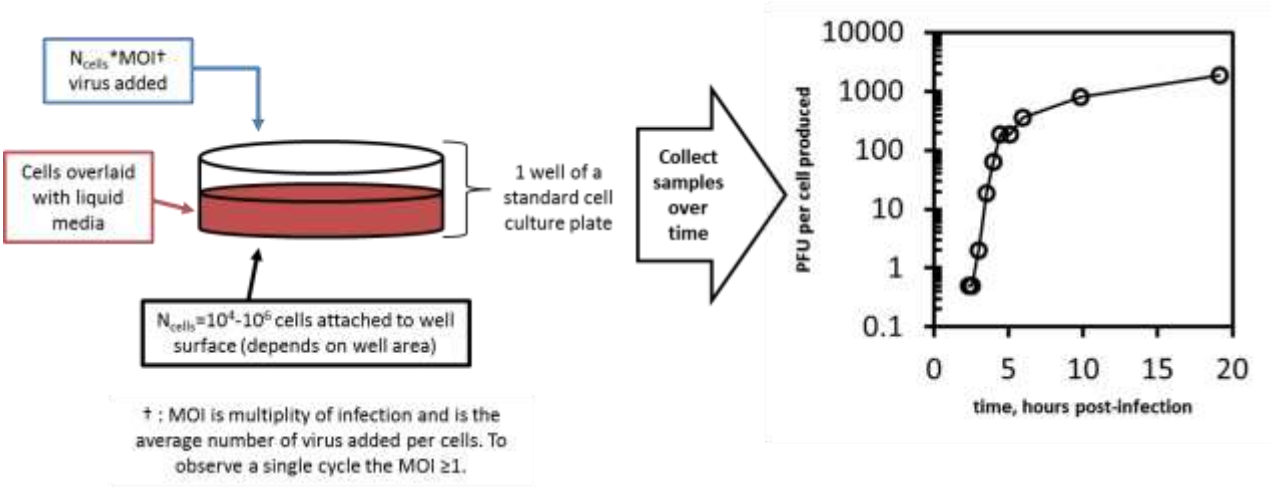


Figure 1.3-1 Schematic of a one-step infection. The left shows a normal culture of cells which has an average virus per cell (MOI) added to infect all cells. The right shows data of the production of infectious virions (PFU=plaque forming units) from a population of cells over time.

A similar strategy can be used to collect samples for measuring viral RNA or viral protein. It is important to note that while a one-step infection is usually used to consider an average cell, the actual progression of infection in any individual cell may be very different from the average (26). However, in this work I only consider population averages.

1.3.2 Reverse transcription (RT)

Reverse transcription (often abbreviated RT) is a molecular biology method to convert RNA to DNA. The conversion is important because methods exist to quantify DNA in a sample, but not RNA. Reverse transcription utilizes enzymes originally isolated from viral sources, which are optimized and sold by a variety of vendors. The specific RT used here is from Promega initially isolated from MMLV (Moloney Murine Leukemia Virus).

The process of reverse transcription is dependent upon an initial DNA primer binding to a specific sequence of RNA in the sample. By taking advantage of this requirement we design DNA primers that are able to convert specific molecules in a sample into DNA. The DNA produced by the reverse transcription enzyme is called cDNA, for **c**omplementary DNA. The produced DNA has the complimentary sequence to the target RNA.

1.3.3 qPCR

qPCR stands for **q**uantitative **P**olymerase **C**hain **R**eaction. The exponential amplification of PCR is made quantitative by adding a fluorescent dye that binds to double-stranded DNA during the amplification. The dye intensity is measured and increases exponentially with the DNA. A threshold is set to capture exponential growth, and the threshold cycle is determined for each reaction. Because the DNA doubles in quantity with every cycle, a doubling of initial DNA results in a threshold cycle offset of one unit. By interpolating threshold cycles a qPCR assay can be used to quantify relative amounts of DNA in parallel samples. It is important to note that the readout we get is exponential or fold-change differences between sample which results in error that is normally distributed about the exponential change and not the absolute level of molecules. Figure 1.3-2 shows the process of qPCR and how it can be used to quantitatively measure DNA counts in a sample.

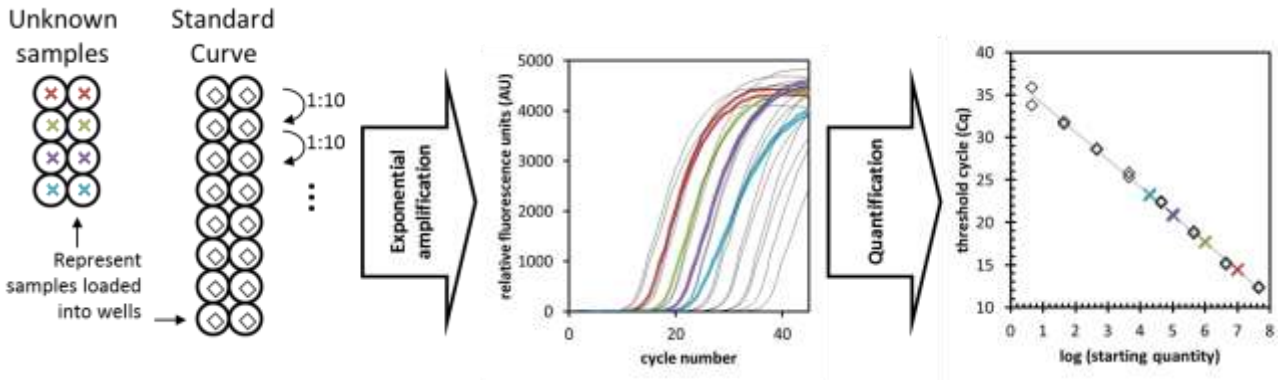


Figure 1.3-2 qPCR workflow schematic. DNA or cDNA samples are loaded into wells, amplified and monitored for fluorescence, and quantified by comparing to the known levels in the standard curve (10-fold dilutions)

1.3.4 Chemical reaction kinetics and ODE modeling

In this work I use mass-action chemical reaction kinetics to describe the interactions of viral intermediates in an infection. Mass-action kinetics treats the chemical reactions as elementary such that rates are calculated based on the number of molecules involved in a reaction. Traditionally this style of modeling is applied to large scale reactions where the number of molecules is very high. In my modeling I use "average-cell" data to model process in an "average cell". This is an important distinction because my measurements are able to detect (on average) ~1 molecule per cell for viral genomes. By treating this value as a continuous concentration as opposed to a discrete integer value we are still able to fit the resulting models to data collected.

I will use an example from Chapter 2 to explain the modeling process. Figure 1.3-3 shows the data that will be modeled for this example: VSV mRNA production and genome degradation in the presence of a translation inhibiting drug.

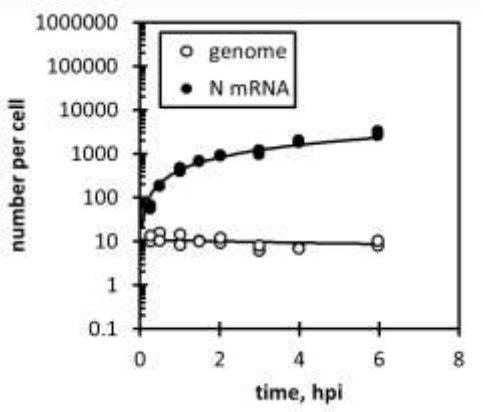


Figure 1.3-3Genome and mRNA data for modeling example. The data is from Chapter 2, Figure 1, medium MOI infection. Genomes are shown as open symbols, and mRNA as closed symbols. Axes are time, in hours post-infection and number per cell, an average value measured from ~10^5 cells. Replication is not observed because translation is inhibited through treatment with cycloheximide.

As shown, the genomes start at about 10 per cell, and degrade slowly over time. We see that mRNA grows linearly (note the log axis) over time. Because mRNA is produced from genomes, we chose the follow two reactions to represent this system:

$$2g \overset{k_d}{\rightarrow} 0$$

$$g \overset{k_t}{\rightarrow} m + g$$

The first reaction implies that two genomes (g) interact and are destroyed with rate constant kd (this choice is explained in detail in Chapter 2). The second reaction states that a genome (g) produces mRNA (m) at a rate k_t, and the genome is not consumed in the process.

These reactions lead to the following differential equations and initial conditions to describe the kinetics of the system:

$$\frac{dg}{dt} = -2k_d g^2$$
$$\frac{dm}{dt} = k_t g$$

$$g(0) = g_0$$
$$m(0) = 0$$

For this system there are three adjustable parameters: k_d, k_t, and g_0. In this case m(0) is set to 0, but it could also be treated as adjustable. The parameters are adjusted to minimize

$$obj = \sum_{i=1}^{n} [ln(y_i) - ln(x_i)]^2$$

Where y_i is the model solution at time index i, and x_i are the experimental observations (replicates) at that time. While an analytical solution is available for this system, the above equations are used with numerical integration and the optimization problem was solved in Matlab using parest. The lines in Figure 1.3-3 (above) show the fits to the data.

1.4 References

1. **Lichty BD, Power AT, Stojdl DF, Bell JC**. 2004. Vesicular stomatitis virus: re-inventing the bullet. Trends in Molecular Medicine **10**:210–6.

2. **Lyles DS, Rupprecht CE**. 2007. Rhabdoviridae, p. 1363–1408. *In* Knipe, DM, Howley, PM (eds.), Fields Virology, 5th ed. Lippincott Williams & Wilkins, Philadelphia, PA.

3. **Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD**. 1996. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. Science **271**:1582–1586.

4. **Anderson RM, May RM**. 1979. Population biology of infectious diseases: Part I. Nature **280**.

5. **Herz A V, Bonhoeffer S, Anderson RM, May RM, Nowak M a**. 1996. Viral dynamics in vivo: limitations on estimates of intracellular delay and virus decay. Proceedings of the National Academy of Sciences **93** :7247–7251.

6. **Bonhoeffer S, Coffin JM, Nowak MA**. 1997. Human immunodeficiency virus drug therapy and virus load. Journal of Virology **71** :3275–3278.

7. **Schiffer JT, Swan D, Al Sallaq R, Magaret A, Johnston C, Mark KE, Selke S, Ocbamichael N, Kuntz S, Zhu J, Robinson B, Huang M-L, Jerome KR, Wald A, Corey L**. 2013. Rapid localized spread and immunologic containment define Herpes simplex virus-2 reactivation in the human genital tract. eLife **2**:e00288.

8. **Adiwijaya BS, Kieffer TL, Henshaw J, Eisenhauer K, Kimko H, Alam JJ, Kauffman RS, Garg V**. 2012. A Viral Dynamic Model for Treatment Regimens with Direct-acting Antivirals for Chronic Hepatitis C Infection. PLoS Computational Biology **8**:e1002339.

9. **Dahari H, Layden–Almer JE, Kallwitz E, Ribeiro RM, Cotler SJ, Layden TJ, Perelson AS**. 2009. A Mathematical Model of Hepatitis C Virus Dynamics in Patients With High Baseline Viral Loads or Advanced Liver Disease. Gastroenterology **136**:1402–1409.

10. **Dixit NM, Layden-Almer JE, Layden TJ, Perelson AS**. 2004. Modelling how ribavirin improves interferon response rates in hepatitis C virus infection. Nature **432**:922–4.

11. **Zeremski M, Talal AH**. 2007. Hepatitis C viral kinetics during treatment of hepatitis C virus/HIV coinfected patients. Current opinion in HIV and AIDS **2**:489–95.

12. **Heidtke KR, Schulzećkremer S**. 1998. Design and implementatio of a qualitative simulation model of λ phage infection. Bioinformatics **14**:81–91.

13. **You L, Yin J**. 2001. Simulating the growth of viruses. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing 532–43.

14. **Kim H, Yin J**. 2004. Energy-efficient growth of phage Qβ in Escherichia coli. Biotechnology and Bioengineering **88**:148–156.

15. **Cairns BJ, Timms AR, Jansen VAA, Connerton IF, Payne RJH**. 2009. Quantitative Models of <italic>In Vitro</italic> Bacteriophage–Host Dynamics and Their Application to Phage Therapy. PLoS Pathog **5**:e1000253.

16. **Sidorenko Y, Reichl U**. 2004. Structured model of influenza virus replication in MDCK cells. Biotechnology and Bioengineering **88**:1–14.

17. **Regoes RR, Crotty S, Antia R, Tanaka MM**. 2005. Optimal replication of poliovirus within cells. The American Naturalist **165**:364–73.

18. **Sardanyés J, Solé R V, Elena SF**. 2009. Replication mode and landscape topology differentially affect RNA virus mutational load and robustness. Journal of Virology **83**:12579–89.

19. **Drake JW, Holland JJ**. 1999. Mutation rates among RNA viruses. Proceedings of the National Academy of Sciences of the United States of America **96**:13910–3.

20. **Lim K, Lang T, Lam V, Yin J**. 2006. Model-Based Design of Growth-Attenuated Viruses. PLoS Computational Biology **2**.

21. **Hensel SC, Rawlings JB, Yin J**. 2009. Stochastic kinetic modeling of vesicular stomatitis virus intracellular growth. Bulletin of Mathematical Biology **71**:1671–92.

22. **Lazzarini RA, Keene JD, Schubert M**. 1981. The origins of defective interfering particles of the negative-strand RNA viruses. Cell **26**:145–154.

23. **Li D, Lott WB, Lowry K, Jones A, Thu HM, Aaskov J**. 2011. Defective interfering viral particles in acute dengue infections. PloS One **6**:e19447.

24. **Pesko KN, Fitzpatrick KA, Ryan EM, Shi P-Y, Zhang B, Lennon NJ, Newman RM, Henn MR, Ebel GD**. 2012. Internally deleted WNV genomes isolated from exotic birds in New Mexico: function in cells, mosquitoes, and mice. Virology **427**:10–7.

25. **Roux L, Simon AE, Holland JJ**. 1991. Effects of defective interfering viruses on virus replication and pathogenesis in vitro and in vivo. Advances in Virus Research **40**:181–211.

26. **Timm A, Yin J**. 2012. Kinetics of virus production from single cells. Virology **424**:11–17.

Chapter 2: Kinetics of transcription and replication for an RNA virus: from simple to complex

2.1 Abstract

In order to persist viruses have evolved strategies to grow in diverse host environments. To probe such strategies we used qRT-PCR to measure viral RNA during infections by the non-segmented negative-sense RNA virus, vesicular stomatitis virus (VSV). Absolute levels of the VSV major transcript and genome were measured for different MOI (1, 10, 100) and diverse host cell environments spanning from simple (permissive BHK cells treated with cycloheximide to inhibit protein synthesis) to complex (resistant PC-3 cells pre-stimulated with type I interferon to activate an innate immune response). While viral genome replication was delayed in more resistant host cell environments, kinetic modeling showed a simple linear dependence of transcription rate on genome levels that spanned all host conditions. These results indicate that while transcription and replication both depend on the availability of the viral RNA-dependent RNA polymerase and host cellular resources, transcription proceeds without apparent limits on these resources. This relative insensitivity of viral transcription to essential viral and host cellular resources may be a characteristic of other RNA viruses.

2.2 Introduction

Vesicular stomatitis virus (VSV) has a broad cell tropism that makes it an attractive candidate for vaccine vectors and engineered, attenuated virus (1). In order to be successful at infecting a wide range of cells types virus growth must be robust to different host cell conditions. Variability in the host environment can have dramatic effects on the production of infectious virus particles (2). Viruses may overcome such limitations by using robust strategies that are insensitive to host variability.

Non-segmented negative-sense (NNS) RNA viruses such as VSV have a unique transcriptional process. The viral particles contain the genome, polymerases, and necessary helper proteins to initiate primary transcription. Transcription initiates at the 3' transcriptional promoter so the gene most proximal to the 3' end is the first transcribed (3). At each intergenic junction the polymerase either initiates transcription of the next gene or does not initiate transcription, in which case the polymerase may dissociate from the template (4, 5). By this way a gradient of gene expression is produced depending on the position of the gene relative to the 3' end (5).

Primary transcripts are translated and the feedback of new viral proteins leads to genome replication (5, 6). Genome replication is performed using the same catalytic polymerase units (L and P proteins) as used for transcription. The presence of sufficient levels of viral N protein is required to initiate genome synthesis (6). The N protein coats the nascent RNA produced by the polymerase such that full length RNA templates in the cell are all encapsidated by the N protein (5). Genome synthesis occurs through a full length positive-stranded template intermediate (5).

Previous studies have reported the transcriptional attenuation and polymerase elongation rates for vesicular stomatitis virus *in vitro* (7). *In silico* models of VSV have used this and other transcriptional data to predict mRNA production rates as a part of a simulation of the larger viral

infection (8, 9). The current model of NNS RNA virus transcription suggests that genomes should act independently during primary transcription, so the initial multiplicity of infection (MOI) should not affect the production rate of mRNA when appropriately scaled by genome numbers (5).

However, open questions on resource allocations remain. While we may be able to predict primary mRNA production, it is unknown if the feedback of new viral polymerases associated with secondary transcription will increase the mRNA production rate. It is also possible that the switch of polymerase activity from transcription to replication could bring about an observable decrease in mRNA production. Further, some transcriptional studies were performed *in vitro* (7); due to the interactions among resources required for transcription, such as availability of templates, viral RNA-dependent RNA polymerase (RdRp), and pools of nucleotide triphosphates (NTPs), mRNA production rates *in vitro* may be different from rates *in vivo,* in either permissive or resistant host cellular environments.

In this work we initially examined how the VSV mRNA production rate changes in the presence or absence of translation. New polymerase molecules produced during translation have the potential to bring about different rates of primary and secondary mRNA production. Further, translation also leads to genome replication which potentially competes with the transcription process by using the same active unit for polymerization. We also investigate how the host environment affects transcription by measuring mRNA production rates on permissive baby hamster kidney (BHK) cells or resistant human prostate cancer (PC3) cells. VSV grows efficiently on BHK cells, but has been shown to be inhibited in infections of PC3 cells (10). PC3 resistance to VSV has been attributed to an active immune response (11), so we also investigate mRNA production and genome replication in un-stimulated and interferon (IFN) stimulated PC3

cells. To determine how the mRNA production rate is related to genomes we employed a qRT-PCR assay to quantitatively measure viral nucleocapsid (N) mRNA and genomes. To facilitate analysis we further developed a simple three reaction model to describe genome replication and test the relationship between viral mRNA synthesis and genome levels.

2.3 Results

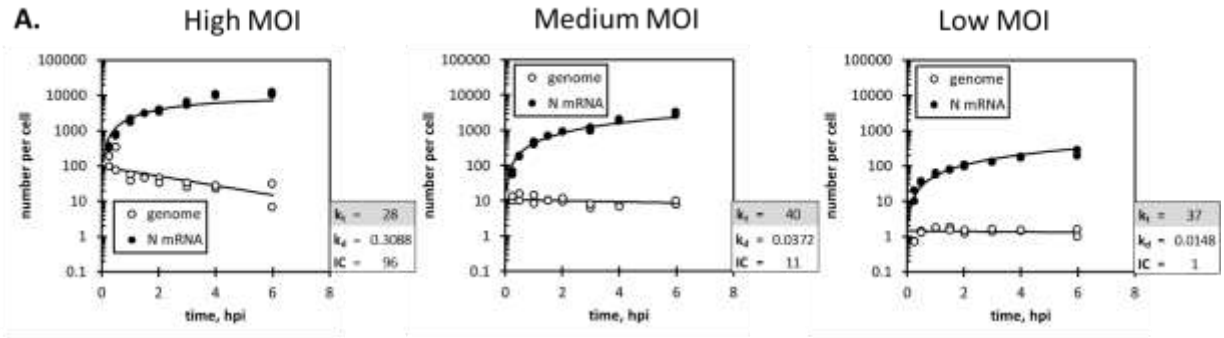
2.3.1 Development of a qRT-PCR assay to measure the kinetics of viral N mRNA and genomes during an infection

qRT-PCR primers were designed to target viral N mRNA and genomes as described in the Methods. We chose the N mRNA to act as an indicator of viral transcription, as it is the first transcribed gene and attenuation has been investigated elsewhere (7). We took extra precaution to ensure that the primers could linearly detect viral mRNA and genomes in cellular samples. To do this, we diluted infected cell RNA (12 hpi, BHK cells) into mock-infected RNA to determine a lower limit of detection and establish that the primers detect RNA linearly over a biologically relevant range of conditions, as shown in Figure 2.5-1. These measurements represent averages from ~10^5 cells.

Measuring rates of mRNA production and replication require multiple time points with replicates over an appropriate time range for the experiment. Briefly, cells were infected at multiple times before harvest allowing a full 96 well plate to be collected and analyzed simultaneously. Using this technique we were able to generate data sets with eight time points in duplicate for six conditions in a 96-well format. This approach gave consistent results with the traditional time course infection as shown in Figure 2.5-2.

2.3.2 Rate of mRNA production depends linearly on genome levels in the absence of translation

BHK cells were infected in the presence of cycloheximide to inhibit translation of new viral proteins at three initial MOI. As shown in Figure 2.3-1, viral genomes degraded with first-order kinetics over the six hour infection period. During this time, mRNA were produced linearly from genomes. Using a model based only on genome levels (and not protein levels) we are able to fit mRNA production data. Further, the rate constants between the three MOI conditions do not follow an observable trend and were of similar magnitude, suggesting that the primary mRNA production rate from genomes is independent of MOI.



$$g \xrightarrow{k_t} m + g$$

$$g \xrightarrow{k_{d,1}} 0$$

$$\frac{dm}{dt} = k_t g$$

$$\frac{dg}{dt} = -k_{d,1} g$$

$$g(0) = g_0$$

$$m(0) = 0$$

Figure 2.3-1 **VSV N mRNA and genome kinetics in cycloheximide treated BHK cells**. BHK cells were treated with 50 ug/mL cycloheximide. Medium and low MOI are 10 and 100 fold dilutions of high MOI samples, respectively. Samples were collected in duplicate as described in Methods. A) mRNA and genome data from cycloheximide treated cells. B) model reactions and differential equation solutions. Parameter Tables: k_t is the transcription constant (mRNA/genome/hr), k_d is the genome degradation constant (hr^{-1}), and IC is the initial condition (genomes).

2.3.3 Rate of mRNA production depends linearly on genome levels in the presence of translation and replication

BHK cells were infected using ten-fold dilutions of initial MOI. As shown in Figure 2.3-2, genome replication began after a delay that increases with decreasing MOI, and genome replication was essentially complete after six hours of infection. Message mRNA production is observed immediately and continued throughout the six hour infection. The mRNA levels over time were described using a model that is only based on genome levels, while the genome levels could be described empirically using an exponential replication reaction, a second order decay reaction, and a time delay. As emphasized in Figure 2.5-3 we observe that mRNA production is linear over the course of this infection.

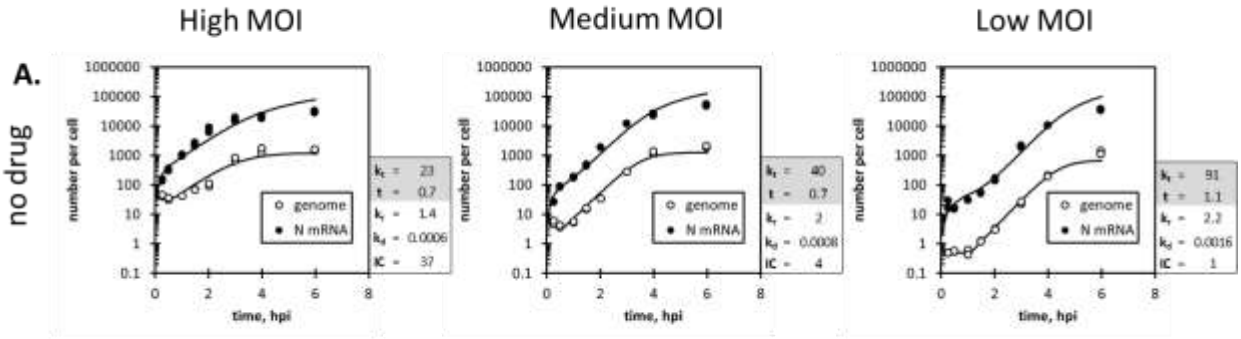


Figure 2.3-2 **VSV N mRNA and genome kinetics in BHK cells**. Medium and low MOI are 10 and 100 fold dilutions of high MOI samples, respectively. Samples were collected in duplicate as described in Methods. A) mRNA and genome data from cycloheximide treated cells. B) model reactions and differential equation solutions. Parameter Tables: k_t is the transcription constant (mRNA/genome/hr), τ is the time delay before replication, k_r is the genome replication constant (hr^{-1}), k_d is the genome degradation constant ($genomes^{-1}hr^{-1}$), and IC is the initial condition (genomes).

2.3.4 Primary is slower than secondary mRNA production rate in PC3 cells

We challenged the virus by infecting PC3 cells using ten-fold dilutions of MOI. PC3 cells have shown to be resistant to VSV in culture due to inhibition at multiple steps (11). As shown in Figure 2.3-3, we observed a long delay (2.3-3.6 hours) before genome replication; the delay was longer at lower MOI. During the delay, mRNA production was slower than following the delay, behavior that was also observed in infections treated with cycloheximide, as shown in Figure 2.5-4. To describe this behavior, we set the primary mRNA production rate constant to be a fraction (α) of the secondary rate constant as shown in the equations in Figure 2.3-3B. We estimated α to be 0.06, consistent with previous findings (11).

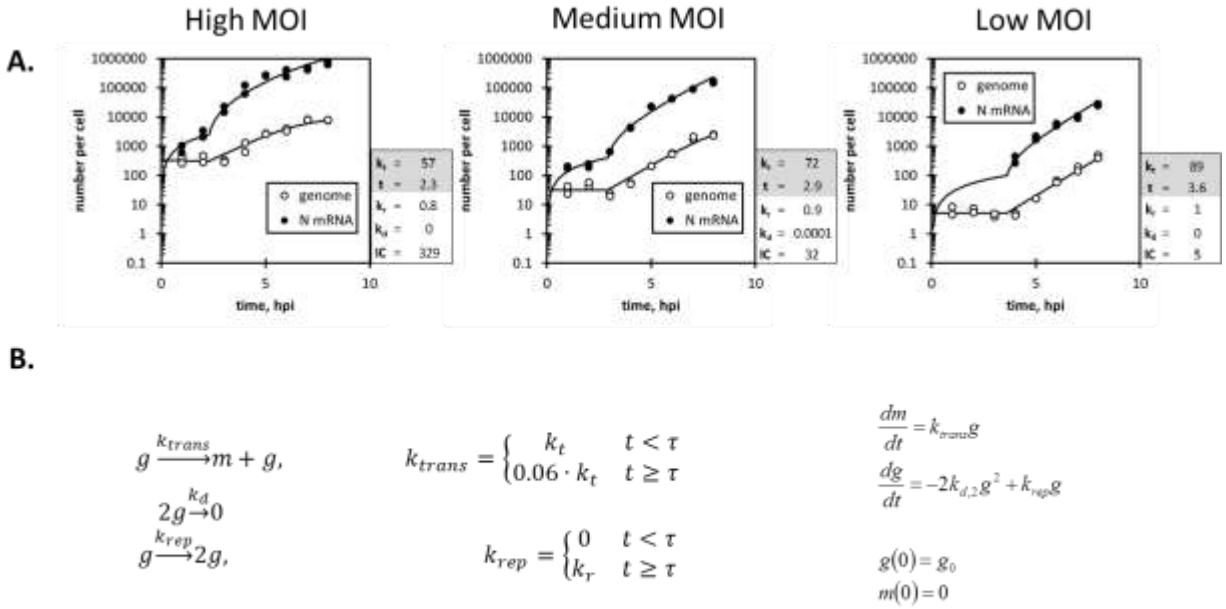


Figure 2.3-3 **VSV N mRNA and genome kinetics in PC3 cells.** Medium and low MOI are 10 and 100 fold dilutions of high MOI samples, respectively. Samples were collected in duplicate as described in Methods. A) mRNA and genome data from cycloheximide treated cells. B) mRNA and genome data from untreated cells. Parameter Tables: k_t is the transcription constant (mRNA/genome/hr), τ is the time delay before replication, k_r is the genome replication constant (hr^{-1}), k_d is the genome degradation constant ($genomes^{-1}hr^{-1}$), and IC is the initial condition (genomes).

2.3.5 Stimulation of the innate response by IFN does not affect the dependence of mRNA production on genome levels

Mechanisms for PC3 resistance to VSV infection arise from a functional innate immune response (10, 11) that is activated upon detection of viral intermediates through a variety of receptors. Activation of receptors stimulates innate response genes and the production of the anti-viral cytokines. The cytokines, which include type I interferons, act in positive feedback to produce a strong and fast innate response to viral infection (12). We chose to investigate how the innate response affects viral mRNA production and replication further by stimulating PC3 cells with 1000 U/mL of interferon-β (IFN-β). IFN pretreatment has been shown to inhibit VSV at the level of entry (13) and primary transcription (14). Figure 2.3-4 shows the data and model fits for mock-treated and IFN-treated cells. IFN stimulation did not affect genome delivery to cells as evidenced by similar fitted initial conditions. The mRNA production rate and genome replication rate were decreased by IFN treatment while the delay and degradation constants were increased by IFN treatment. Interestingly, the high MOI infection was better able to overcome the inhibitory effects of IFN as compared to medium or low MOI infections. As shown in Figure 2.5-5, the relative levels of mRNA and genomes decline more rapidly for medium and low MOI infections after six hours whereas the relative levels for high MOI stay higher.

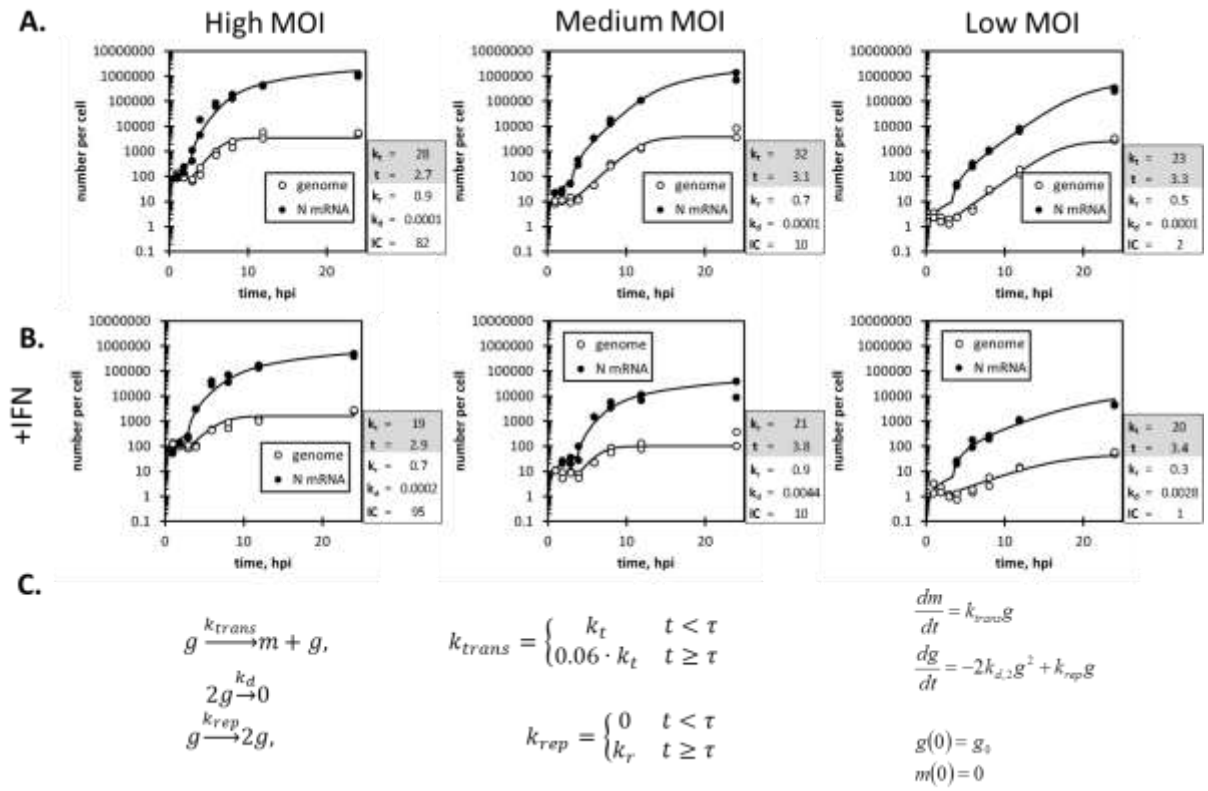


Figure 2.3-4 **VSV N mRNA and genome kinetics in IFN stimulated or untreated PC3 cells.** PC3 cells were treated with 1000 Units/mL human IFN-beta 1 hour before infection, and co-treated; or not treated. Medium and low MOI are 10 and 100 fold dilutions of high MOI samples, respectively. Samples were collected in duplicated as described in Methods. A) N mRNA and genome measures and model for untreated PC3 cells B) N mRNA and genome measures for IFN treated PC3 cells Parameter Tables: k_t is the transcription constant (mRNA/genome/hr), τ is the time delay before replication, k_r is the genome replication constant (hr^{-1}), k_d is the genome degradation constant ($genomes^{-1}hr^{-1}$), and IC is the initial condition (genomes).

2.4 Discussion

VSV is known for its broad cell tropism (1). In order to grow in multiple cell types under multiple conditions we hypothesized that features of the virus infection strategy would be relatively insensitive to host cellular environment. To test this hypothesis we measured mRNA and genome kinetics under a variety of infection conditions. We used a lumped modeling approach to help analyze and draw conclusions from the detailed kinetics of the infection. Our model consisted of three reactions: exponential genome amplification (delayed), genome degradation, and a rate of mRNA production that was linearly dependent on genome levels. The ability of this simple model to fit experimental data shows a predictable relationship between genomes and mRNA and supports the hypothesis that the mRNA production rate is produced linearly from genomes early in infection. This suggests that the virus is operating in an environment with excess polymerase, as the polymerase levels do not need to be included in the equations to fit the mRNA production data.

We did not observe a decrease in the rate of mRNA production upon initiation of genome replication. Because the same polymerase complexes are used to generate both mRNA and genomic templates, we might expect a shift to purely replication. The continued high mRNA production rate suggests that even in the presence of sufficient N protein, transcription is still the dominating role of the polymerases. This could potentially be explained by a strong transcription promoter relative to replication.

Another interesting result is that primary mRNA production occurred throughout the long cycloheximide treatment in both BHK and PC3 cells (Figure 2.3-1A and Figure 2.5-3). Given a VSV processivity rate of 3.7 nt/s and a genome length of 11161 nt, the residence time of a given polymerase on a genomic template is ~54 minutes. With a polymerase loading of 50 polymerases

per genome (15) we expect that the last polymerase would dissociate from the genome after two hours into the infection. As shown in Figure 2.3-1A and Figure 2.5-3, we see mRNA production that continues at a constant rate for up to eight hours. Continued mRNA production suggests that polymerases are recycled after intergenic attenuation or reaching the end of the genome, as opposed to diffusing freely away from genomes. This hypothesis is supported by a loop-like structure observed in influenza RNA that would facilitate polymerase recycling (16, 17).

Previous work has shown that polymerases can be inhibited or degraded by human MxA, known to be expressed in PC3 cells (14). However, we do not see a continuing decrease in PC3 primary mRNA production which would be evidence of continued degradation of polymerases. It is possible that they are degrading but without a detectable large effect on the transcript production rate.

Interestingly, the high MOI case was able to grow similarly in the IFN treated and untreated infections (Figure 2.3-4 and Figure 2.5-5). The high MOI rate constants for all reactions were similar in IFN treated and untreated infections, while in medium and low MOI the degradation and replication constants were altered by the treatment. The results presented here show that it is feasible for the virus to overcome an innate response by producing large amounts of virus.

It is important to note that the model did not fully capture all trends in the data, most notably the saturation of the mRNA data after ~6 hours in BHK infections and ~20 hours in PC3 infections. Data after these times are lower than the model, highlighting a limitation of the model at later time in the infection. The saturation of mRNA data observed in longer infections may be explained by matrix protein inhibition of polymerase activity seen in other related viruses (18–20). We chose not to include this mechanism because most of the experimental data was before observable mRNA saturation. This means that the model presented here is more applicable for

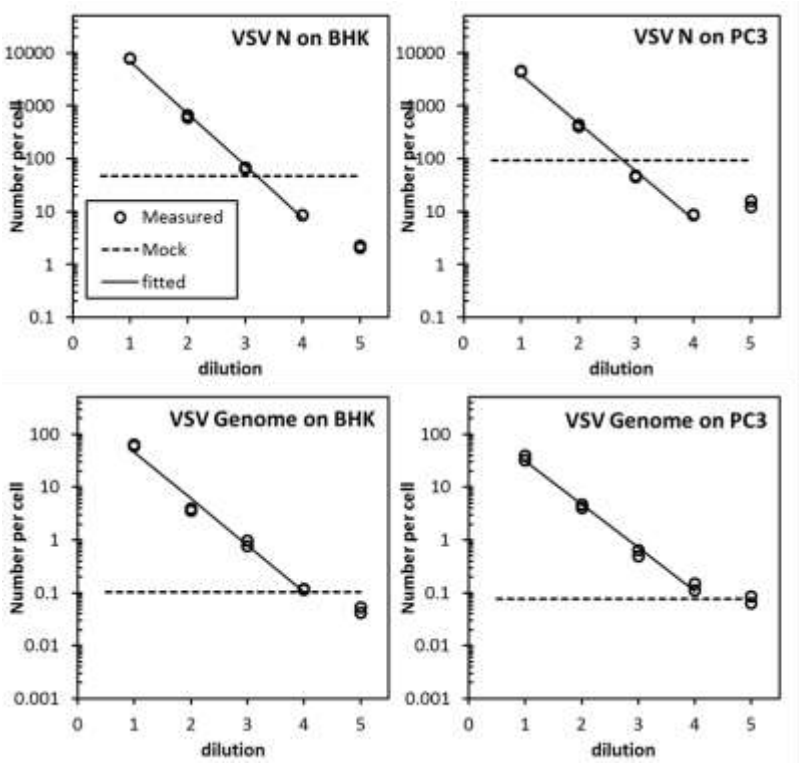
early infection kinetics and could be advanced by including more details of the later infection process. Future models coupled with quantitative protein kinetics will help extend predictions to all aspects of the virus infection.

MOI does not appear to have a large effect on mRNA production rates as evidenced by the model fit to all MOI conditions. The lack of any trend in mRNA production rate constants with MOI suggests that genomes act independently during transcription, especially during primary mRNA production. However, we did observe some trends with MOI for other model parameters. In general, the delay before replication increased with a decrease in MOI, especially in PC3 cells. An explanation for shorter delay at higher MOI is that more genomes can make more mRNA and thereby establish a replication environment more quickly. We also observed a general decrease in the replication rate constant with decreasing MOI in PC3 cells. This may be explained by early innate response events which lead to a less-permissive replication environment. The IFN treatment helps support this by showing that the high MOI infections are able to overcome the innate stimulation better than the medium or low MOI infections. While the reactions used to model genome kinetics are not mechanistic, the rate constants do provide a summary of the data and allow us to measure a bulk rate of genome replication in the system.

We were able to capture the mRNA production and genome replication data using a relatively simple model of only 3 reactions. The model only uses genome levels and does not depend on protein concentration in the cells to fit mRNA production data. This suggests that the VSV replication strategy is constant even in the presence of an anti-viral response: the virus establishes an environment of excess protein such that the mRNA production rate is constant from genomes. Being able to approximate the transcription process with this simple model does not suggest that the details of the transcription mechanism are not important. It does, however,

provide a template for estimating the quantitative effects of mutations and drugs on the early infection process as evidenced by the IFN treatment experiment in Figure 2.3-4. Further experiments can be used to investigate other drug effects or the infection process in different cellular environments. Quantitative protein data and particle production data can also be incorporated to generate a full quantitative description of the viral infection.

2.5 Supplemental Figures



	BHK N mRNA	PC3 N mRNA	Ideal
slope	-0.99	-0.91	-1
intercept	4.83	4.50	NA

	BHK Genome	PC3 Genom	Ideal
slope	-0.88	-0.82	-1
intercept	2.54	2.32	NA

Figure 2.5-1 **Reverse transcription validation.** Infected cell RNA (harvested at 6 hpi) was diluted into mock-infected cell RNA. Parallel RT reactions were performed to show the linearity and lower limit of detection of the RT step. Fit lines are included for the linear detection region. Mock lines (dashed) show higher qPCR signals than some of the dilutions, but samples below LOD were distinguished by melt curves. Mock samples were included in all experiments and data below mock values was omitted from analysis.

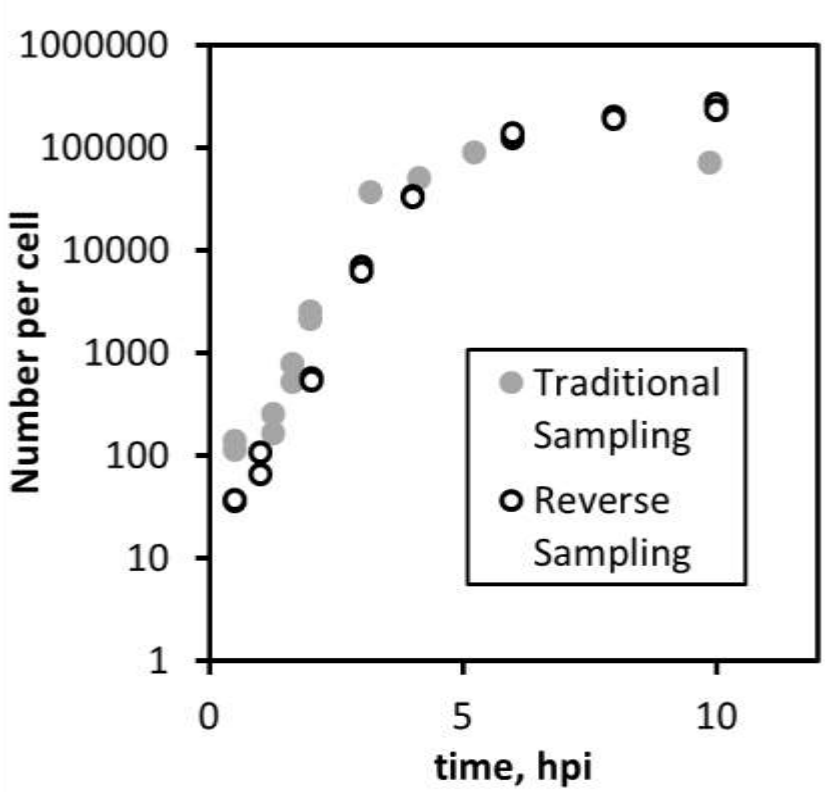


Figure 2.5-2 **Reverse sampling is equivalent to traditional sampling.** BHK cells were infected at MOI ~5 in multiple 6 well plates (traditional) or using reverse method in 96 well plates (reverse). Cells were sampled serially and stored in RNAProtect (traditional) or simultaneously (reverse) and N mRNA was counted by qRT-PCR method. As shown, the results give similar trajectories. The reverse sampling is much more high throughput allowing for 96 samples while traditional sampling is much more labor intensive.

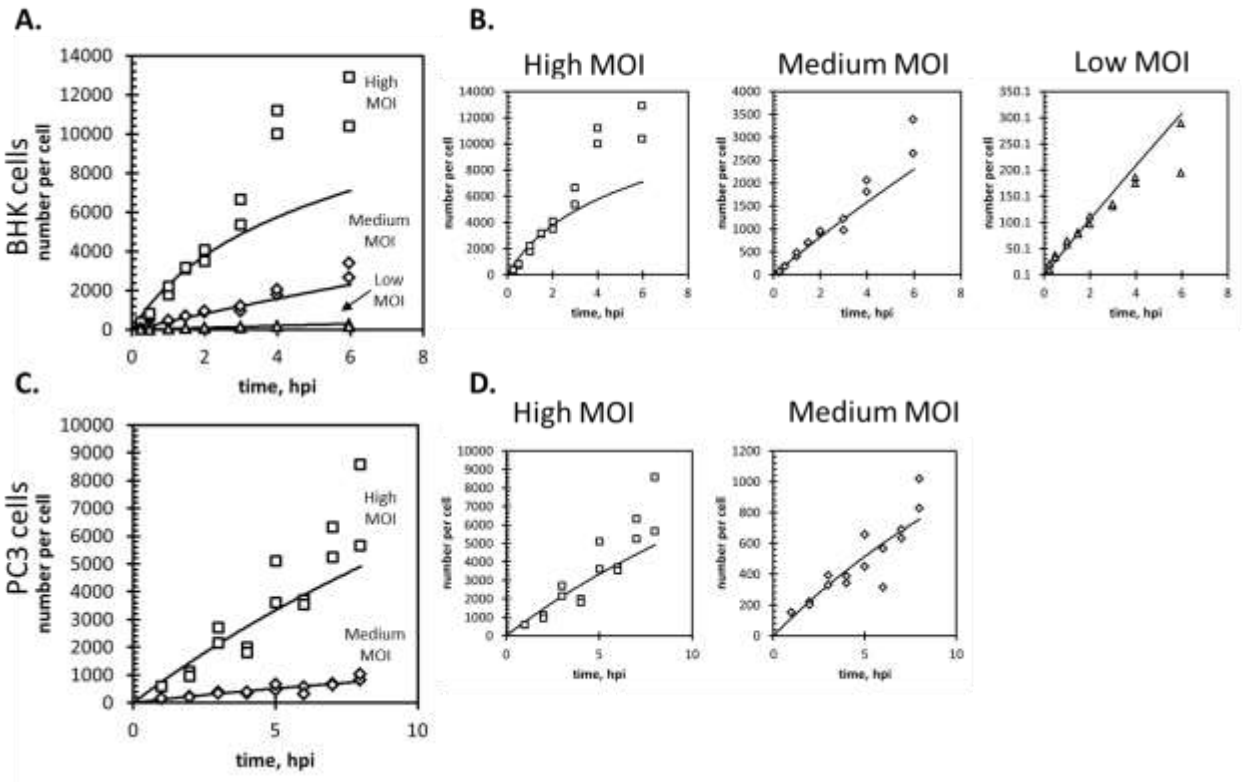
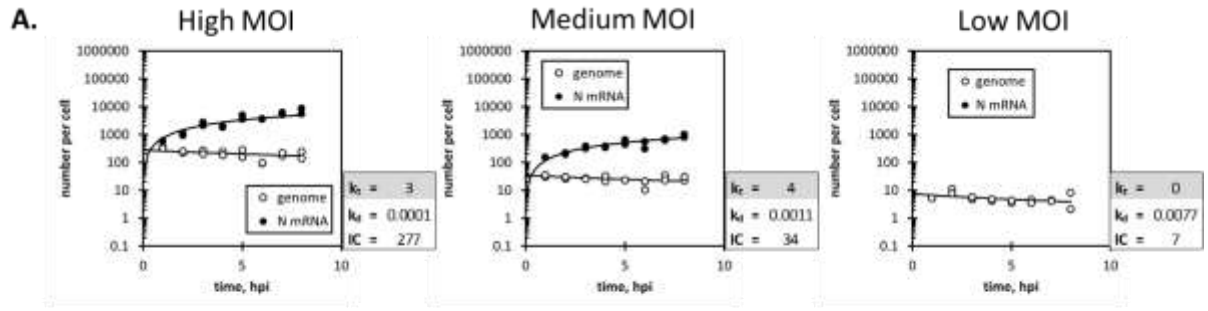


Figure 2.5-3 **Linear scale plots of N mRNA per cell for cycloheximide treatments**. **A.** The three MOI treatments on BHK cells are plotted together to show the effect of incoming MOI on primary transcription. (high, medium, low = 160, 16, 1.6 PFU/cell). The lines show the model fit as shown in Figure 2. The line is not linear because during primary transcription genomes are degrading. **B.** Individual plots for each MOI show data and model fit. Note: the y-axis scale is different between plots. **C-D.** Same as A-B, but on PC3 cells



B.

$$g \xrightarrow{k_t} m + g$$

$$g \xrightarrow{k_{d,1}} 0$$

$$\frac{dm}{dt} = k_t g$$

$$\frac{dg}{dt} = -k_{d,1} g$$

$$g(0) = g_0$$

$$m(0) = 0$$

Figure 2.5-4 **VSV N mRNA and genome kinetics in cycloheximide treated PC3cells**. PC3 cells were treated with 50 ug/mL cycloheximide. Medium and low MOI are 10 and 100 fold dilutions of high MOI samples, respectively. Samples were collected in duplicate as described in Methods. A) mRNA and genome data from cycloheximide treated cells. B) model reactions and differential equation solutions. Parameter Tables: k_t is the transcription constant (mRNA/genome/hr), k_d is the genome degradation constant (hr^{-1}), and IC is the initial condition (genomes).

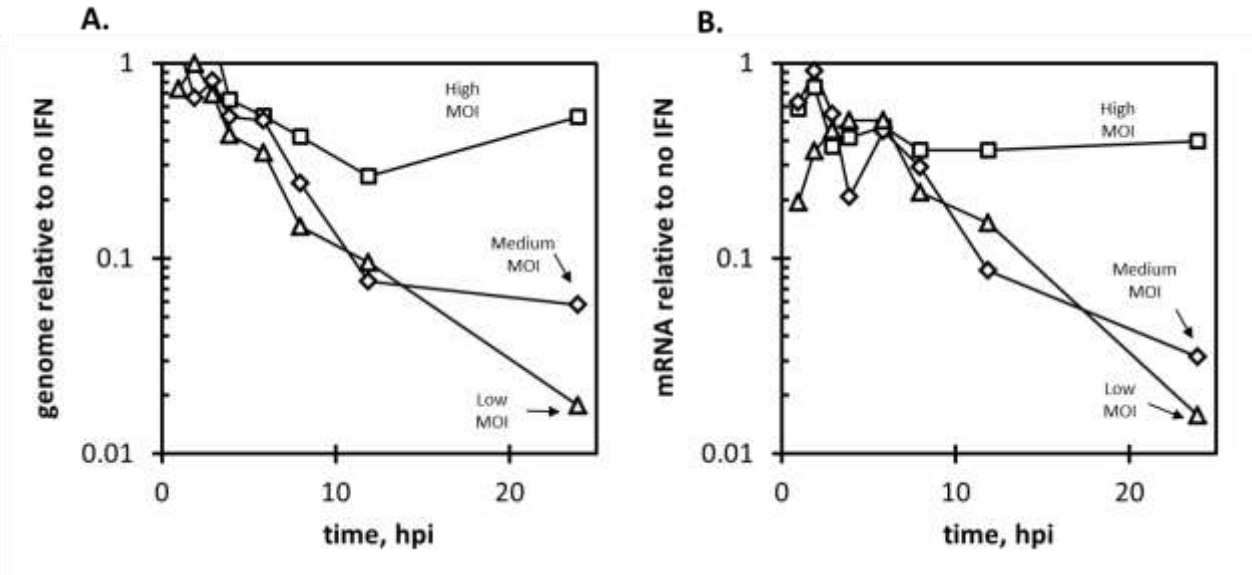


Figure 2.5-5 **IFN treated measures relative to un-treated infection.** Data from Figure 5 was normalized to the no IFN case and averaged over the duplicate measures. **A.** relative genome levels. **B.** relative mRNA levels. Notice that the medium and low MOI are inhibited to the same level, but the high MOI ratio stays closer to 1, indicating that the high MOI does a better job at overcoming the immune response as compared to medium or low MOI.

2.6 Acknowledgements

Collin Timm acknowledges Ankur Gupta as a 2nd author for his contributions to this work. Ankur helped with the implementation of parest for parameter estimation and early model development efforts.

2.7 Supplemental Information: Model Development

The final model consists of three reactions,

$$g \xrightarrow{k} m + g, \qquad k = \begin{cases} BHK & PC3 & condition \\ k_t & k_t & t < \tau \\ k_t & 0.06 \cdot k_t & t \geq \tau \end{cases}$$

$$2g \xrightarrow{k_d} 0 \qquad (1)$$

$$g \xrightarrow{k_{rep}} 2g, \qquad k_{rep} = \begin{cases} 0 & t < \tau \\ k_r & t \geq \tau \end{cases}$$

where m is N mRNA, and g is VSV genomes. Using reaction kinetic principles we get the following differential equations

$$\frac{dm}{dt} = k_t g$$
$$\frac{dg}{dt} = -2k_d g^2 + k_r g \qquad (2)$$
$$g(0) = g_0$$
$$m(0) = 0$$

where transcription constant k_t, genome degradation constant k_d, replication constant $_{kt}$, replication delay τ, and initial genomes g_0 are adjustable parameters fit to the data. The following section describes in detail how we arrive at these equations.

Introduction

Our qRT-PCR assay provides quantitative data on average number of viral N mRNA and genomic templates during an infection. We sought to build a model based on both: 1) known

VSV mechanisms and 2) features of the data we collected. We chose to develop a model with minimal parameters to identify which known mechanisms that are most important for describing the experimental data during the early infection process

It is important to note that the model choice was not immediately obvious to us. There were multiple iterations to determine how best to capture experimental data. In the following text we will explain the iterative process for developing our model, as well as describe some approaches that did not work.

Linear transcription

In Iverson 1981 [ref] the authors showed that viral mRNA was produced linearly in *in vitro* reactions. We first chose to measure a similar system: cells treated with cycloheximide to inhibit translation of viral mRNA and thus remove any autocatalytic effects of new viral proteins.

We observed linear growth of VSV N mRNA with a linear rate that decreased with MOI (**Error! Reference source not found.**A and Figure 2.5-3A). Our first approach was to calculate a linear slope to follow

$$NmRNA(t) = k_t \cdot t \qquad 3$$

We immediately asked whether the slope depended on initial genome counts, and then how we can use the genome data to help predict mRNA. A very simple approach would be to determine if the following equation could describe the data with a single slope

$$NmRNA(t) = k_t \cdot g_0 \cdot t \qquad 4$$

Where k_t is the transcription rate, g_0 is the initial number of genomes, and t is time. However, we noticed that the genome counts declined during the infection; they are not constant at g_0. To

capture this decline we needed a model to describe how genome numbers were changing. The model we first chose was a simple, linear degradation describing the rate of change of genomes as

$$\frac{dg}{dt} = -k_d \cdot g \qquad 5$$

(Note that this is NOT the final degradation equation we use for genomes in the paper.) The differential equation can be solved to determine how genomes should change in time

$$g(t) = g_0 e^{-k_d \cdot t} \qquad 6$$

Because genomes numbers are decreasing while producing mRNA, we write the mRNA equation as a differential equation

$$\frac{dNmRNA}{dt} = k_t \cdot g(t) \qquad 7$$

Notice that this equation is linear: the rate of change of mRNA depends only on the number of genomes in the cell, and a transcription rate constant. The N mRNA rate equation above can be solved analytically to give

$$NmRNA(t) = k_t \frac{g_0}{k_d} \cdot \left(1 - e^{-k_d \cdot t}\right) \qquad 8$$

where k_t is the transcription rate constant with units of mRNA/genome/hour. We now have two equations, one to describe how genomes change (Equation 6), and one to describe how N mRNA change (Equation 8). This model captures the data from cycloheximide treated cells (no translation= no replication) well, for both BHK and PC3 cells. While the model can capture the cycloheximide data, this is not the final model we applied to the cycloheximide treated cases.

The data and the model development described in this section were the initial motivation for testing the linear relationship of mRNA production rate to genomes.

Exponential Genome Growth and Saturation

In the parallel infections (not treated with cycloheximide) we observed genome replication. In BHK cells, the genome values for all MOI approached a single saturation value. Exponential growth and a constant saturation can be captured using a two-reaction carrying capacity model for genomes:

$$\begin{aligned} g &\xrightarrow{k_r} 2g \\ 2 \cdot g &\xrightarrow{k_d} 0 \end{aligned} \qquad 9$$

This model allows for exponential replication capture by the first reaction, and a concentration dependent saturation captured by the second equation. These reactions give the following differential equation

$$\begin{aligned} \frac{dg}{dt} &= k_r g - 2k_d g^2 \\ g(0) &= g_0 \end{aligned} \qquad 10$$

where the initial condition for genomes is g_0. Integrating Equation 10 and solving for g(t) gives

$$g(t) = \frac{\frac{k_r}{2k_d} g_0 e^{k_r t}}{\frac{k_r}{2k_d} + g_0 (e^{k_r t} - 1)} \qquad 11$$

We can apply the same linear relationship for mRNA growth

$$\frac{dNmRNA}{dt} = k_t \cdot g(t) \qquad 12$$

And integrate 11 to get a complex function for how mRNA changes over time.

$$NmRNA(t) = \frac{k_t}{2k_d} ln\left(1 + \frac{2k_d g0}{k_r}(e^{k_r t} - 1)\right) \qquad 13$$

Now Equation 11 and 13 can be fit to experimental data with genome replication. This works well for the BHK replication data, but in PC3 cells there is a long, clear delay before VSV

genomes begin replicating. The model cannot capture this delay so the next section will describe how we addressed the delay.

Incorporating a delay

Strict time delays introduce a discontinuity in the model, which makes robust parameter estimation difficult. First we briefly will describe a method that did not work well for this data, then we will describe how we incorporated the delay for the final publication.

A strategy to approximate a delay is to add an upstream reaction and make genomes an intermediate. We chose to do this by adding a reaction to simulate activation of entering genomes

$$\begin{aligned} g^* &\xrightarrow{k_a} g \\ g &\xrightarrow{k_t} m + g \\ 2g &\xrightarrow{k_d} 0 \\ g &\xrightarrow{k_r} 2g \end{aligned} \qquad 14$$

where g* must be activated with rate constant ka before the genome can be degraded, replicated, or transcribed. This is to simulate the slow release of genomes from endosomes as observed in Carey et al (ref). The reactions in Equation 14 give the following differential equations

$$\begin{aligned} \frac{dg^*}{dt} &= -k_a g^* \\ \frac{dm}{dt} &= k_t g \\ \frac{dg}{dt} &= k_a g^* - 2k_d g^2 + k_r g \end{aligned} \qquad 15$$

$$\begin{aligned} g^*(0) &= g_0, \textit{ fitted} \\ g(0) &= 0 \\ m(0) &= 0 \end{aligned}$$

As shown, the fitted initial condition is now applied to g* only. Using this model we expect to see significantly higher activation constants in BHK cells (fast replication) than in PC3 cells (long delay).

We did not use this model for two reason: 1) it does not fit the data as well as later models, and 2) the measurements cannot distinguish between g and g*, which confounds the estimation of k_a, k_r, and k_d.

After the activation model (Equations 14 and 15) did not fit the data well, we chose to introduce a strict time delay before replication. The model can be described by the following reactions

$$\begin{gathered} g \xrightarrow{k_t} m + g \\ 2g \xrightarrow{k_d} 0 \\ g \xrightarrow{k_r} 2g, if(t > \tau) \end{gathered} \qquad 16$$

The "if" statement in the 3rd reaction means that replication can only occur if the simulation time is greater than the time delay, τ. Before the time delay, genomes can be transcribed or degraded; after the time delay, genomes can be transcribed, replicated, or degraded. The following differential equations describe this model

$$\begin{aligned} \frac{dm}{dt} &= k_t g \\ \frac{dg}{dt} &= -2k_d g^2 + k_r g * if\left(t > \tau\right) \\ g(0) &= g_0 \\ m(0) &= 0 \end{aligned} \qquad 17$$

where the "if" statement is only applied to the replication term in the model.

The parameter estimation software was unable to optimize over delays implemented in this fashion. For simplicity, we fit the candidate time delays over a grid of 0, 0.25, 0.5… 6 hours and chose the optimal delay which gives the lowest objective function value.

Different primary transcription and secondary transcription rates in PC3 cells

The delay model (Equations 16 and 17 above) was close to fitting all the data. However, for the PC3 infections the model tended to overestimate early mRNA data, and underestimate late data. This indicates that the primary transcription rate could be slower than the secondary rate. We also had data from the PC3 cycloheximide treatment infections which suggested that primary transcription was about an order of magnitude lower than secondary transcription. This led us to the final model

$$\begin{aligned} & g \xrightarrow{k_t} m + g, \left(kt, primary = \begin{cases} k_{t,secondary}, t < \tau, BHK\ cells \\ 0.1 \cdot k_{t,secondary}, t < \tau, PC3\ cells \end{cases} \right) \\ & 2g \xrightarrow{k_d} 0 \\ & g \xrightarrow{k_r} 2g \left(k_r = \begin{cases} 0, t < \tau \\ k_r, t > \tau \end{cases} \right) \end{aligned} \qquad 18$$

which give the following discontinuous differential equations

$$\begin{aligned} \frac{dm}{dt} &= k_t g \\ \frac{dg}{dt} &= -2k_d g^2 + k_r g \\ g(0) &= g_0 \\ m(0) &= 0 \end{aligned} \qquad 19$$

In this model we picked a multiple of 0.1 on the transcription rate for PC3 cells. Note that the model is different from what is shown in Equation 17. The equations are discontinuous across $t=\tau$.

It is important to note that we did not allow the transcription factor of 0.1 to be adjustable. We chose not to let it be adjustable for two reasons: 1) we did not have sufficient data before the delay to estimate this parameter and 2) the PC3 cycloheximide data showed that the primary rate was approximately 10% of the secondary transcription rate. Further, by not adjusting we got a very good fit to mRNA data as shown in all model fits.

Conclusion

The modeling process was not linear. We tested many model variations to fit our data. All model variations had one thing in common: mRNA data was closely approximated by a linear integration of the genome model. Our main conclusion from this work is that the transcription rate is linearly related to the genomes in the cell, or

$$\frac{dNmRNA}{dt} = k_t \cdot g(t) \qquad 20$$

2.8 Methods

2.8.1 Cell Culture

Baby hamster kidney (BHK-21) cells were maintained in 10% fetal bovine serum (FBS, Atlanta Biologicals), 1% Glutamax I (Gibco) in in Eagle's minimum essential medium (MEM; CellGro). Cells were passaged every 2 or 3 days when they reached confluency. Prostate cancer cells (PC3) were maintained in 10% fetal bovine serum (FBS, Atlanta Biologicals) in RPMI 1640 medium (Gibco). Cells were passaged every 2 or 3 days when they reached confluency.

2.8.2 Virus Culture

Vesicular stomatitis virus (21) was passaged on BHK cells at MOI =0.001 to prevent formation of DIPs. Stocks were centrifuged and filtered to remove cell debris. Stock concentration was ~10^9 PFU/mL.

2.8.3 Virus Infections

Infections were performed in reverse order such that a single harvest of sample would yield the entire time course of the infection. For example, to sample an 8 and 12 hour time point the virus was added 12 hours before harvest-time to one set of culture wells, then again 8 hours before harvest-time to a separate set of culture wells, then cells were harvested simultaneously.

Cells were grown in a 96 well plate overnight to prepare for infections. At each time point the cell supernatant was removed and viral infections were performed by adding 20 uL stock virus (see above) or ten-fold dilutions in culture media supplemented with 2% FBS (Atlanta Biologicals). Virus was adsorbed for 5 minutes at 37°C, then the cells were rinsed with ~50 uL 37°C DPBS. 100 uL of media was added and the infection was allowed to proceed at 37°C. Cells

were infected in duplicate wells. In all cases multiple 96 well plates of cells were used to minimize the effect of removing the cells from the incubator. Cells plated in 3 parallel wells were used for counting to determine an average number of cells per well.

2.8.4 Cycloheximide treatments

Where indicated, cells were treated with 50 ug/mL cycloheximide in 2% media. Cycloheximide (Fisher) was originally dissolved in DMSO at a concentration of 100 mg/mL and the solution was diluted appropriately into 2% cell media. Media containing cycloheximide (or vehicle only control) was added to cells 30 minutes prior to infection. Media was removed and virus was added as above (virus inoculum did not contain cycloheximide). After the PBS wash described above fresh media containing cycloheximide or vehicle only was added to cells.

2.8.5 IFN treatments

Recombinant human IFN-β (PBL Interferon Source, Piscataway, NJ, USA) was added to cell culture media at 1000 U/mL. One hour prior to infection cells were overlayed with media containing IFN or control media. Media was removed and virus was added as above (virus inoculum did not contain IFN). After the PBS wash described above fresh media containing 1000 U/mL IFN-β or control media was added to cells.

2.8.6 RNA Extractions

Prior to RNA extraction infections were halted by placing the culture plates on ice. RNA extractions were performed using the RNEasy 96 kit (QIAGEN) with the vacuum protocol according to manufacturer's instructions. The 96-well extraction format allows for rapid parallel processing of the samples. Output RNA concentrations ranged from 2-20 ng/uL.

Reverse Transcription

Reverse transcription was performed using the GoScript Reverse Transcriptase system (Promega, Madison, WI, USA). Gene specific primers for either N mRNA or the VSV genome were used in parallel reactions from RNA samples. 3.15 uL of undiluted RNA (see above) was mixed with 2 uL of 10 µM gene specific primer and incubated at 70°C for 5 min, then immediately transferred to an EtOH/ice bath for 5 min. Then 1.85 µL of 25 mM MgCl2, 0.5 µL nucleotides (10 µM each), and 0.5 µL GoScript Reverse Transcriptase was added to a final volume of 10 µL. Reactions were incubated according to manufacturer's instructions.

2.8.7 qPCR

qPCR was performed with a forward primer and reverse primer (see below) from RT reaction using the SsoFast Supermix (BioRad, Hercules, CA, USA) on a C1000 thermal cycler (BIORAD). 2 µL of 1:5 diluted cDNA or plasmid standard was added to an 8 µL reaction mix containing 5 µL SsoFast Supermix, 1.2 µL of 5 µM each primer mix, and 1.8 µL RNAse free water. qPCR reactions were performed in duplicate. The standard was a purified plasmid containing the VSV genome was used as a known standard to measure copies per mL. Primers were optimized to detect viral RNA by dilutions of extracted viral genomes or infected cell RNA.

Target	**Primer**	**Sequence**	**Amplicon Length**
N mRNA	Gene specific RT and qPCR Reverse	ttccggatttgaggccttggtaga	119
	qPCR Forward	atccagtggaatacccggcagatt	
Genome	Gene specific RT and qPCR	atcctgctcggcctgagatacaaa	193

	Reverse		
	qPCR Forward	gggtggtgcgatccctaatttctt	

Amplification protocol: 1) 95°C for 45s, 2) 95°C for 5s, 3) 62.5°C for 10s + plate read, 3) Goto 2 45x. Melt curve conditions were 65°C-95°C in 0.5 °C increments with 5 s step times and a plate read at each step.

2.8.8 qRT-PCR Assay Optimization

Primers were initially tested for single product amplicons by running amplified reactions on an agarose gel. Priemrs were temperature and concentration optimized using a dilution of a plasmid with a full genome VSV insert as a PCR standard. After primer and amplification conditions were optimized we tested the ability of gene specific primers to detect dilutions of viral RNA over a background of cellular RNA to determine limit of detection and linearity of the reverse transcription step.

2.8.9 Conversion to number per cell

The starting quantity of the plasmid standard was determined by measured DNA concentration on the NanoDrop and dividing by the molecular weight of the plasmid to convert concentration to copy number. Infection samples were measured using the qRT-PCR assay described above. Starting quantity per well was converted to copy number per cell using the following equation:

$$\frac{\text{copies}}{\text{cell}} = \frac{\text{Starting Quantity}}{\text{qPCR well}} \cdot \frac{\text{qPCR well}}{\text{V uL cDNA}} \cdot \frac{\text{W uLcDNA}}{\text{X uLRNA}} \cdot \text{Y uL RNA} \cdot \frac{\text{mock RNA conc}}{\text{this RNA conc}} \cdot \frac{\text{well}}{\text{Z cells}}$$

Variable	Value	Description
V	2	volume of cDNA added to each qPCR well
W	50	total volume of cDNA (after 1:5 dilution) used for qPCR

X	3.15	volume of undiluted RNA added to each RT reaction
Y	93	eluate volume from the RNEasy 96 RNA extraction protocol and includes all of the RNA extracted from cells
Z	Variable ~10^5	cell count from a given infection experiment (see infection protocol above)
this RNA conc	Variable 2-20 ng/uL	The RNA concentration of each sample was used to normalize counts relative to the mock RNA concentration

2.8.10 Model Solution and Data Fitting

The following reactions and resulting differential equations were used as the model of VSV infection. See Model Development (Supplemental Information) for more information.

$$g \xrightarrow{k} m + g, \qquad k = \begin{cases} BHK & PC3 & condition \\ k_t & k_t & t < \tau \\ k_t & 0.06 \cdot k_t & t \geq \tau \end{cases}$$

$$2g \xrightarrow{k_d} 0$$

$$g \xrightarrow{k_{rep}} 2g, \qquad k_{rep} = \begin{cases} 0 & t < \tau \\ k_r & t \geq \tau \end{cases}$$

The reactions are transformed into differential equations assuming mass-action kinetics, with discontinuous rate constants following the rules above.

$$\frac{dm}{dt} = k_t g$$
$$\frac{dg}{dt} = -2k_d g^2 + k_r g$$

$$g(0) = g_0$$
$$m(0) = 0$$

The stated initial condition g_0, transcription constant k_t, genome degradation constant k_d, genome replication constant k_r, and replication delay τ were fitted parameters.

The models used were implemented in MATLAB and solved using an open source software parest, [http://jbrwww.che.wisc.edu/software/parest/]. We used parest to minimize the likelihood function as defined by the difference between the log of the model solution and the log of the qRT-PCR measurement (in copies/cell), which is equivalent to maximizing the likelihood function. Parameter values which minimize the squared error objective funcation are reported as parameter estimates. Details of the procedure can be found in Rawlings and Ekerdt (22).

2.8.11 Significance testing for model parameters

In cases where statistical tests were performed on parameter values, we used a standard t-test treating parameter values as independent measurements.

Models were compared for significant differences using the F-test (23).

2.9 References

1. **Lichty BD, Power AT, Stojdl DF, Bell JC**. 2004. Vesicular stomatitis virus: re-inventing the bullet. Trends in Molecular Medicine **10**:210–6.

2. **Zhu Y, Yongky A, Yin J**. 2009. Growth of an RNA virus in single cells reveals a broad fitness distribution. Virology **385**:39–46.

3. **Rose JK**. 1980. Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. Cell. Cell Press.

4. **Barr JN, Whelan SP, Wertz GW**. 1997. Role of the intergenic dinucleotide in vesicular stomatitis virus RNA transcription. Journal of virology **71**:1794–801.

5. **Lyles DS, Rupprecht CE**. 2007. Rhabdoviridae, p. 1363–1408. *In* Knipe, DM, Howley, PM (eds.), Fields Virology, 5th ed. Lippincott Williams & Wilkins, Philadelphia, PA.

6. **Davis NL, Wertz GW**. 1982. Synthesis of Vesicular Stomatitis Virus Negative-Strand RNA In Vitro: Dependence on Viral Protein Synthesis. Journal of Virology **41** :821–832.

7. **Iverson LE, Rose JK**. 1981. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. Cell **23**:477–84.

8. **Lim K, Lang T, Lam V, Yin J**. 2006. Model-Based Design of Growth-Attenuated Viruses. PLoS Computational Biology **2**.

9. **Hensel SC, Rawlings JB, Yin J**. 2009. Stochastic kinetic modeling of vesicular stomatitis virus intracellular growth. Bulletin of Mathematical Biology **71**:1671–92.

10. **Ahmed M, Cramer SD, Lyles DS**. 2004. Sensitivity of prostate tumors to wild type and M protein mutant vesicular stomatitis viruses. Virology **330**:34–49.

11. **Carey BL, Ahmed M, Puckett S, Lyles DS**. 2008. Early steps of the virus replication cycle are inhibited in prostate cancer cells resistant to oncolytic vesicular stomatitis virus. Journal of virology **82**:12104–15.

12. **Faul EJ, Lyles DS, Schnell MJ**. 2009. Interferon response and viral evasion by members of the family rhabdoviridae. Viruses **1**:832–51.

13. **Whitaker-Dowling PA, Wilcox DK, Widnell CC, Youngner JS**. 1983. Interferon-mediated inhibition of virus penetration. Proceedings of the National Academy of Sciences **80** :1083–1086.

14. **Staeheli P, Pavlovic J**. 1991. Inhibition of vesicular stomatitis virus mRNA synthesis by human MxA protein. Journal of Virology **65** :4498–4501.

15. **Thomas D, Newcomb WW, Brown JC, Wall JS, Hainfeld JF, Trus BL, Steven AC**. 1985. Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. Journal of Virology **54** :598–607.

16. **Arranz R, Coloma R, Chichón FJ, Conesa JJ, Carrascosa JL, Valpuesta JM, Ortín J, Martín-Benito J**. 2012. The Structure of Native Influenza Virion Ribonucleoproteins. Science **338** :1634–1637.

17. **Coloma R, Valpuesta JM, Arranz R, Carrascosa JL, Ortín J, Martín-Benito J**. 2009. The Structure of a Biologically Active Influenza Virus Ribonucleoprotein Complex. PLoS Pathog **5**:e1000491.

18. **De BP, Thornton GB, Luk D, Banerjee AK**. 1982. Purified Matrix Protein of Vesicular Stomatitis Virus Blocks Viral Transcription in vitro. Proceedings of the National Academy of Sciences of the United States of America **79**:7137–7141.

19. **Hankins RW, Nagata K, Kato A, Ishihama A**. 1990. Mechanism of influenza virus transcription inhibition by matrix (M1) protein. Research in Virology **141**:305–314.

20. **Suryanarayana K, Baczko K, ter Meulen V, Wagner RR**. 1994. Transcription inhibition and other properties of matrix proteins expressed by M genes cloned from measles viruses and diseased human brain tissue. Journal of Virology **68** :1532–1543.

21. **Wertz GW, Perepelitsa VP, Ball L a**. 1998. Gene rearrangement attenuates expression and lethality of a nonsegmented negative strand RNA virus. Proceedings of the National Academy of Sciences of the United States of America **95**:3501–6.

22. **Rawlings JB, Ekerdt JG**. 2002. Chemical Reactor Analysis and Design Fundamentals.

23. **Wu CFJ, Hamada MS**. 2009. Experiments: Planning, Analysis, and Optimization.

Chapter 3: A mechanistic model of vesicular stomatitis infection

3.1 Abstract

Vesicular stomatitis virus (VSV) is the most widely studied of the Mononegavirales family which includes measles, rabies and Ebola viruses. Owing to its extensive study, as well as its potential biomedical applications, researchers have developed mechanistic models based on well-described functions of the virus during an infection. These models were designed to help engineer the growth of VSV. However, the models are complex and require expert knowledge of the simulation environment when modifications are necessary. Here we developed a genome-based, easy to modify, delayed ODE model to describe processes of VSV infection. By incorporating new data and mechanisms we are able to provide bounds for the simulation using physical limitations and data-based trajectories. Our model predicts that genome replication is limited by the production rate of proteins, rather than consumption of any specific reactant. Although our model was only fit to a single initial condition, the model qualitatively predicts trend observed for multiple initial conditions and gene-insertion mutants. The simple, self-regulating model structure is ideal for fast simulation and prediction of VSV kinetics. The described model can aid in the design of growth-attenuated viruses as live vaccines or vectors.

3.2 Introduction

Vesicular stomatitis virus (VSV) is well known for its ability to infect and grow in a variety of cell culture conditions. Due to the high infectivity and ease of culture of the virus, much of the primary literature focused on determining mechanisms of VSV growth using the newest techniques as they were developed. The high amount of historical research has provided an extensive understanding of the mechanisms VSV employs during its replication cycle. A well-described system is ideal for generating a quantitative model of biological processes as it does not require extensive assumptions about specific mechanisms. In this case we develop a quantitative model of VSV infection using delayed ordinary differential equations.

Quantitative models of virus infections can be used to help identify sensitive mechanisms, test hypothesis, and design multiple aspects of the viral infection. Virus infection models were initially developed to predict growth rates of chronic infections on an organism scale (1–3). Such organism scale models were extended to investigate mechanisms of drug inhibition in HCV infections (4) and to design anti-viral treatment regimens for chronically infected HCV patients (5). Detailed molecular mechanistic models have also been developed to study the viruses of importance to human health, including HIV-1 (6), influenza virus (7) and polio virus (8), exploiting the wealth of knowledge on the intracellular replication of these viruses.

However these intracellular models of virus replication generally do not account for potentially significant delays in biopolymerization process such as production of mRNA or proteins or production of the viral genome: processes which may take on the order of minutes to hours. Delay models for biopolymerization processes have been investigated for multiple systems (9–11). In this work we apply similar methods to model the VSV infection as a system of delayed ODEs.

VSV is an attractive option for mechanistic models as it is well–described, has been widely studied as a model for viral evolution (12–15), and has been proposed as a potential oncolytic agent (16, 17) or vaccine vector (18). A computational and mechanistic model of VSV infection complements such evolution and viral design studies by providing an integrated description of viral processes and potential explanations for experimental observations. An ODE model based on VSV mechanisms showed that viral protein production is limited by a translational window with boundaries defined by viral takeover of ribosomes and host cell death (19). This model was also used to study the predicted behavior of VSV gene-shuffled mutants (20). While this model was shown to be predictive and relatively powerful, it was in practice difficult to modify the viral genome and modify specific mechanisms of the infection. As experimental hypotheses are tested, new mechanisms will likely be discovered, providing motivation to extend the initial model. Having an easy to modify, genome-based model will facilitate the interaction between the model and the experiments. While VSV is relatively easy to modify through reverse genetics, we lack the complementary mathematical model to describe the impact on VSV mechanisms.

In this work we generated a delayed differential equation model to describe VSV infection kinetics. The model was parameterized based on known mechanisms, physical limitations, and recent quantitative data of VSV infection. The model is also simple to modify: genes can be rapidly added to the VSV genome and new mechanisms can be added using the framework we developed. Our model also incorporates a simple host response that is effective in preventing unbounded viral genome replication. Finally, we use our model to interrogate the effects of multiple initial conditions and demonstrate the ability to easily simulate the effects inserted genes in the virus genome.

3.3 Results

3.3.1 Model formulation

Our goal was to generate a model structure for VSV simulations that is easy to modify and manipulate. There are often new discoveries made about viruses and virus host-interactions, so our biological model should be readily able to adapt existing reactions or incorporate new ones. The reverse genetics system for VSV is also relatively easy to manipulate in the lab allowing for the creation of new mutant strains of the virus, and we wanted to ensure that inserting a gene into the genome in the simulation was straightforward and simple for future users.

Many of the intermediates of interest in a viral infection require an extensive time for polymerization to come into being. For example, when transcription of an mRNA molecule is started there is a significant time before the molecule is completed. Because many of the products in the simulation have such delays, we partitioned the cell environment into "pools" of similar system intermediates. Figure 3.5-1shows the strategy we used to account for delayed and available items during an infection. The simulated cell contains 4 generic pools to account for templates, transcripts, proteins and virions (Figure 3.5-2).

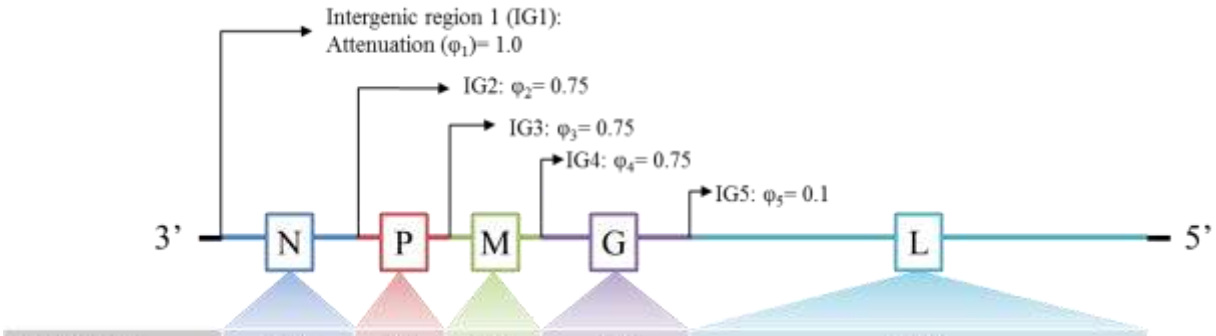
To ensure flexibility and ease of adding new mechanisms we developed a standardized reaction format to describe steps of the viral replication cycle:

$$w \cdot A, (\alpha, pool_A) + x \cdot B, (\beta, pool_B) \rightarrow y \cdot C, (t_C, pool_C) + z \cdot D, (t_D, pool_D) \quad , \quad k_{rxn}, \; r = f_x$$

In the standard reaction A, B, C, and D are potential reactants and products. The stoichiometry of the reaction is denoted by w, x, y, and z. The species in the reaction are followed by two quantities in parenthesis. For reactants, the quantities α and β represent the order of the reaction, to be used in calculating a rate based on mass action kinetics. The "pool" term denotes where

information about quantities of species are stored in the simulation. On the product side, t_C and t_D represent the delay time to produce the specified product. A rate constant k_{RXN} is specified. Finally, the rate can be either mass action (default) and calculated from the reaction orders, or it can be a user specified function (f_x) which may depend on catalyst concentration or other parameters.

To ensure that changing the VSV genome in a simulation was straightforward we developed our model to accept complex reactants, rather than just a single named item. In this case genome templates act as a reactant for transcription but also include information about the gene order, length, and other gene functions. In this way our modeling environment closely resembles the real system, as a single virus genome contains all the "information" to perform all identified virus functions. Figure 3.3-1 shows a schematic of the information that the genomes contains.



	N	P	M	G	L
Nucleotides [1]	1333	822	838	1672	6380
Amino Acids [1]	422	265	229	511	2109
Molecules per virion [2]	1258	466	1826	1205	50

Figure 3.3-1 **The VSV genome contains all information for transcription, translation, and packaging of the virus.** The attenuation between genes determines the relative mRNA production and the availability of polymerases. The number of nucleotides and polymerase elongation rate are used to calculate production delays for mRNA using t_{delay}=(gene length, nt) / (pol. rate, nt/s). Similarly, delays are calculated for protein translation. The genome acts as a reactant for transcript production, mRNA acts a reactants for translation, and all information is referenced back to a single viral genome.

1- NCBI Genome entry for Vesicular stomatitis Indiana virus, NC_001560.1

2- **Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis.**D Thomas, W W Newcomb, J C Brown, J S Wall, J F Hainfeld, B L Trus, and A C Steven, J. Virol. May 1985 54:2 598-607PMID: 2985822

The genome is referenced to determine information about gene lengths and potential delay times. We use the order of genes and names of specific gene products to determine the delay for production and maturation of gene products.

While there is not sufficient quantitative data to determine parameter values for all of the described virus mechanisms, there are however many physical and experimentally observed boundaries that allow us to estimate appropriate values for parameters. We developed the model in a stepwise fashion by first considering primary transcription, then translation, replication, late viral process and finally packaging of the virions. In the following sections we describe the physical, biological and data-driven mechanisms that helped us determine appropriate parameter values for a VSV infection.

3.3.2 Primary Transcription

We assume genome delivery is relatively fast compared to other processes. VSV entry was shown to occur on the order of minutes (21). Thus, our simulation begins after the genome has been delivered to the cytoplasm. Production of mRNA during primary transcription is dominated by the polymerase association with the genome and the polymerization rate of the polymerase. The spacing of polymerases on the genomic template is also a consideration: there is a physical size limitation that can only allow a single polymerase to operate on a given length of nucleotides. To capture both concentration dependence and physical limitations we used a rate calculation further described in the Methods. In Chapter 2 we observed that primary transcription (independent of translation) occurred for a much longer duration than expected for the residence time of a polymerase on a genome template. We hypothesized that the rate of re-association of the polymerase with the genome was very high, because we did not see a decrease in mRNA production rate. Based on this observation we parameterized the model such that

association of the polymerase with the template was almost instantaneous. This implies that the association constant for genomes and polymerase is very high, leaving two variables that allow us to modulate the production rate of mRNA: polymerase spacing and nucleotide processivity rate. The polymerase spacing was a fitted parameter in a previous VSV model, so we allowed it to vary. Similarly, the nucleotide processivity rate was initially estimated to be 3.7 nt/s (22). The initial estimate was from an *in vitro* system with uncertain relevance to the infection in cells, so we also allowed this rate constant to vary about 5-fold of the original 3.7 nt/s. Our goal was to match an mRNA production rate of ~100 mRNA per genome per hour (Chapter 2). To reach this production rate the polymerase spacing was set to 400 nucleotides per polymerase, and the processivity rate was set to 15 nucleotides per second. Figure 3.3-2 shows the production of mRNA from a single infection genome. We also notice a slight degradation of genomes over this time, as observed in Chapter 2.

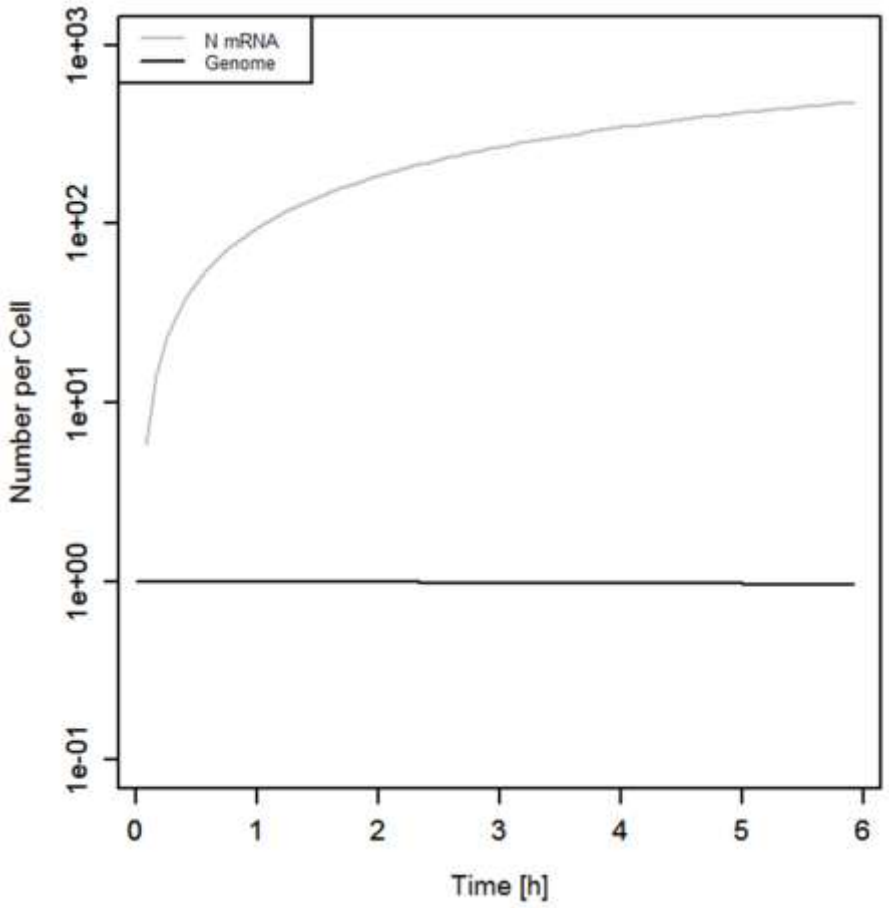


Figure 3.3-2 Model prediction of primary transcription for VSV in a BHK cell environment.

3.3.3 Translation

There is relatively little data available on the number of proteins produced from a single transcript during a VSV infection, but we do have estimates for loading of mRNA by ribosomes and the ribosome elongation rate. The maximal loading of mRNA with ribosomes is determined by ribosome spacing rate estimated to be as close as 80 nucleotides (23). The processivity rate of a ribosome has an average rate of about 20 amino acids per second (23). For our model we estimated 100-1000 proteins produced per mRNA per hour. Because these are both estimates and are not exactly known for our system, we allowed both these parameters to vary and set the spacing to be 180 nucleotides per ribosome with processivity rates of 12 amino acids per second.

3.3.4 Replication

VSV replication is controlled by the viral N protein encapsidating new RNA formed by the polymerase (24). The switch to polymerase replication activity should then depend on the level of N protein relative to genomes in the cell. Interestingly, a hard switch (every polymerase performing replication) was not observed at any point during replication. Rather, transcription and replication occur simultaneously until the end of the infection. However, there is a clear experimentally observed delay that precedes the start of genome replication. During this delay transcripts and protein are produced. After sufficient N protein has accumulated the strength of the replication promoter relative to the transcription promoter controls polymerase fate. During parameterization we observed that the transcription to replication promoter had a strong effect on the delay and subsequent production of genomes. If the replication promoter was too strong, we observed fast consumption of N protein with little production of mRNA. Therefore, we hypothesized that controlling the promoter strength ratio of transcription versus replication is important for capturing experimental observations. It is also important to note that we could not use strict mass action to describe the N protein concentration effect on replication: the exponential increase in N protein caused an exponential increase in replication rate relative to transcription rate. Instead, we used a logarithmic function to describe the N protein concentration's contribution to the replication reaction rate calculation (see Methods).

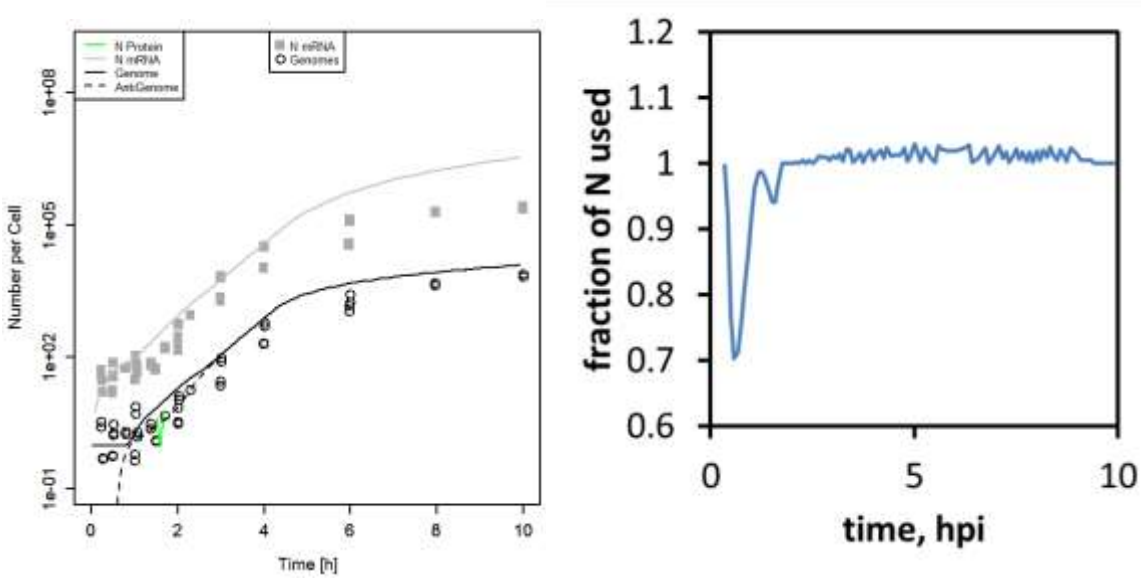


Figure 3.3-3 Viral translation leads to a switch in polymerase activity to replication. A. Genome and mRNA data are shown as points, and simulation as lines. Data are collected from multiple infections as described in Methods. The lines are higher than the data, as we have not yet included mechanisms which inhibit replication. B. The fraction of total N protein that is consumed during infection.

Using the described parameters we were able to simulate genome replication similar to observed values in experimental data. Using the described rates, viral N protein was the limiting factor for replication. The simulation essentially predicts that there is no free N protein in the system, which explains the absence of the N protein curve in Figure 3.3-3. While it is unlikely that all of the N protein is consumed during a real infection, we interpret this result to suggest that the vast majority of N protein is included in replication complexes.

Interestingly, the decrease in the genome production rate observed in Figure 3.3-3 is not explicitly coded into the model. The decrease in production rate is attributed to a switch to linear genome growth, driven by the limit of translational resources in the cell. The explanation for this is that genome replication depends on N protein availability and after genomes have reached a certain level the ribosomes cannot produce enough protein to continue exponential replication of genomes. The behavior makes intuitive sense, because the virus does not have a mechanism to

slow itself down, and the host translation resources are limited. The host also has active mechanisms to inhibit the virus which we explore those in the following section.

3.3.5 Viral detection and host shut down

The host is not idle throughout the infection process but responds actively to inhibit the replication cycle of the virus, a process that depends heavily on the cell type. The host cell senses the viral infection through either the Toll-like or RIG-I receptors, both which recognize dsRNA as a ligand (25). The activation of these receptors leads to production of an antiviral state that is unfavorable for viral replication. While the host inhibition is a complex function of multiple interactions, we chose to simplify the process into two reactions: detection and inhibition. The production of detection molecules depends on viral genome count which we argue is a proxy for detection of dsRNA. Detection molecules interact with ribosomes to inactivate them, which makes the simulated mechanism of inhibition translation-based. Protein production for the virus is thus internally limited by the ability of the host to detect viral genomes. Figure 3.3-4 shows the production of detection molecules and the inactivation of ribosomes over the course of the infection.

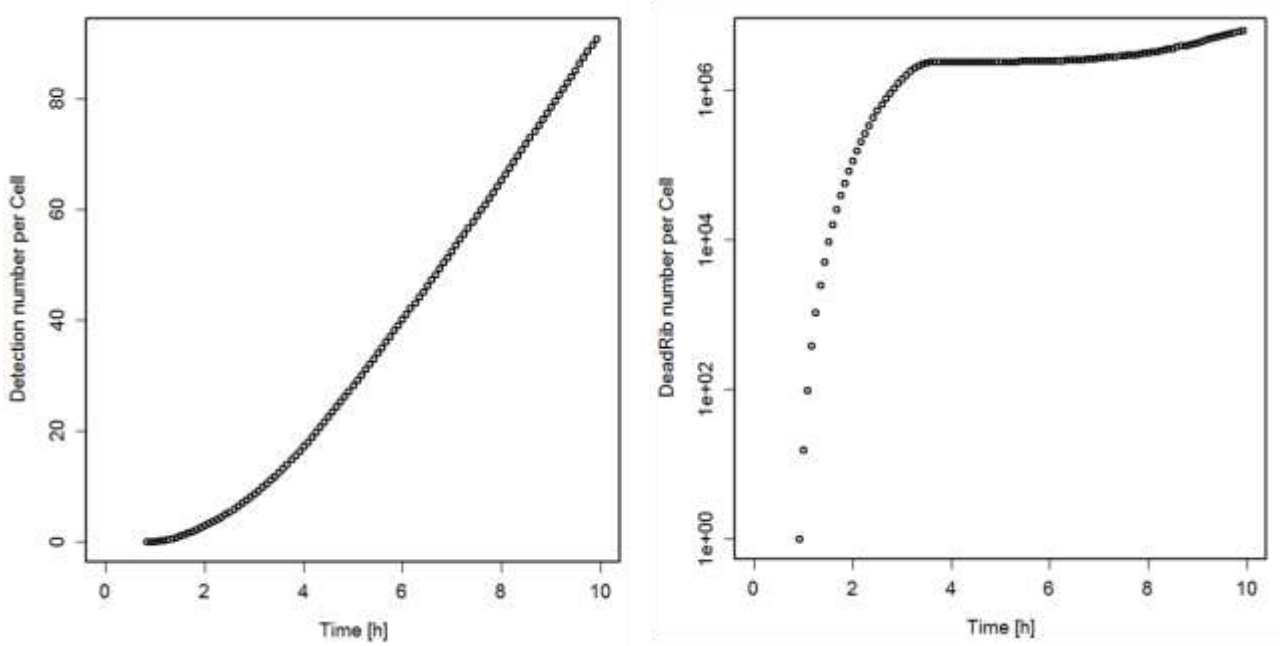


Figure 3.3-4 Detection of viral genomes and inactivation of ribosomes in the cell. The left plot shows the number of detection molecules growing over time. The detection molecules interact with ribosomes to inactivate them (right plot).

As shown, detection molecules grow exponentially but slow later in the infection. The detection molecules interact with ribosomes to inactivate them and remove them from the simulation. The result of this is a decrease in overall translation rate.

3.3.6 Late viral processes and packaging

The viral M protein assists in nucleocapdsid condensation (26) and has been shown to inhibit the activity of the VSV polymerase (27). We capture both observed trends by treating the condensed nucleocapsid as a replication and transcriptionally inactive genome species. The condensed nucleocapsid is the final precursor to the viral particle and interacts with viral G protein on the cell surface to produce new particles. Figure 3.3-5 shows the result of activating condensation and final packaging of the virion.

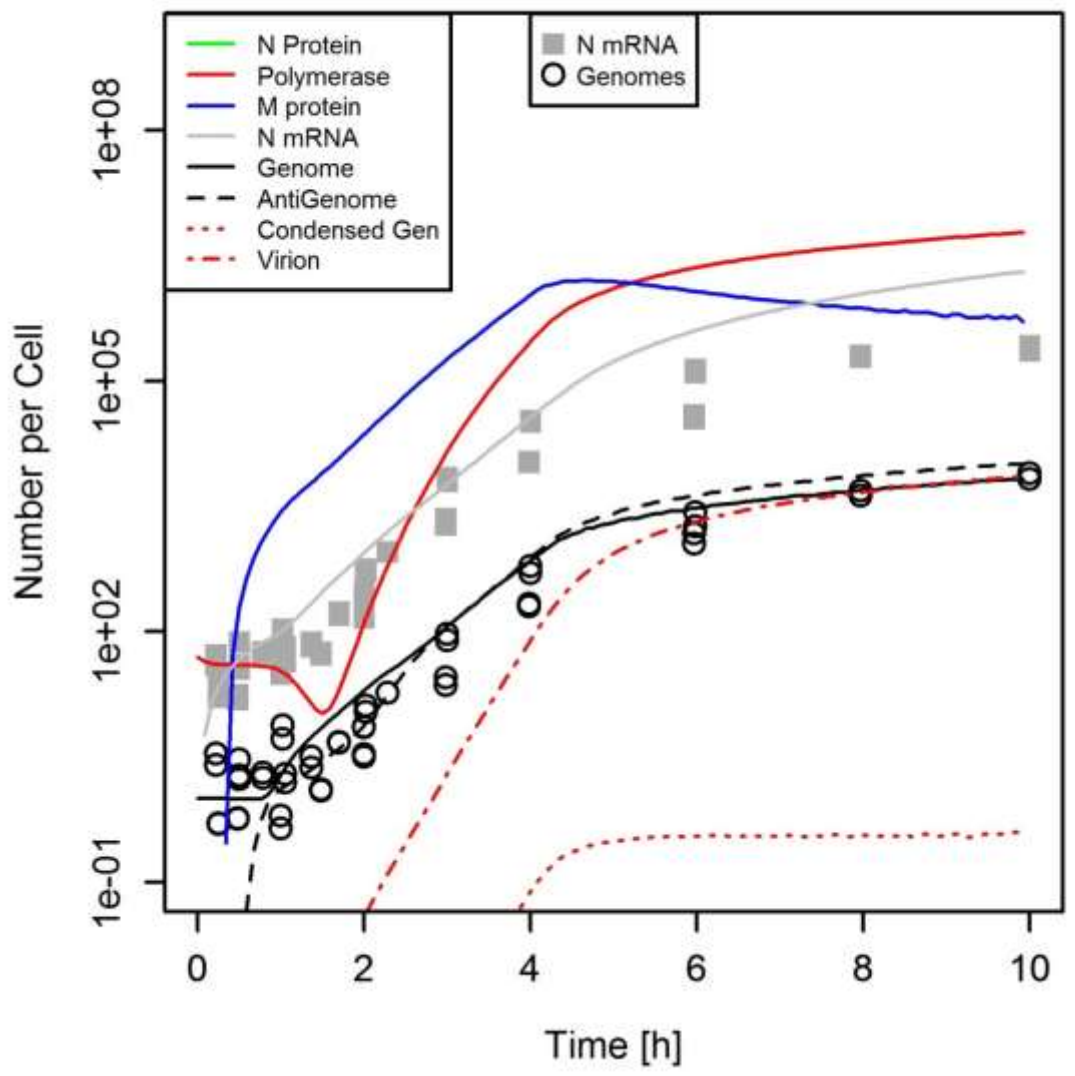


Figure 3.3-5 Full simulation of VSV infection produces virions.

As shown, the first infectious virions are produced at about 3 hours post-infection (corresponding to the time the virion curve crosses y=1). We do not have data for the condensation rate of genomes within the cell. To model condensation and particle formation the simulation uses a relatively slow condensation rate with a high rate of packaging of condensed genomes into particles. The result is a small concentration of condensed but unpackaged genomes within the cell. The packaging steps were parameterized to produce about 1000 particles by 10 hrs, as

demonstrated by data for BHK cell infections (Figure 3.5-3). Interestingly, while M protein limits the rate of condensation, we see a much larger fraction of P protein consumed in the production of particles. This suggests that after N protein, P is the most limiting of the proteins when normalized by expression value (Figure 3.5-4).

3.3.7 Multiplicity of infection

In VSV infections it is known that different initial conditions change the kinetics of virus production, but do not greatly affect the total particles produced (Figure S3). Infections initiated with different multiplicity of infection (MOI) have been shown to approach similar maximums in genomes (Chapter 2) and particles produced (Figure 3.5-3) in BHK cells. Figure 3.3-6 shows the results of multiplicity of infection (MOI) interrogation of the model.

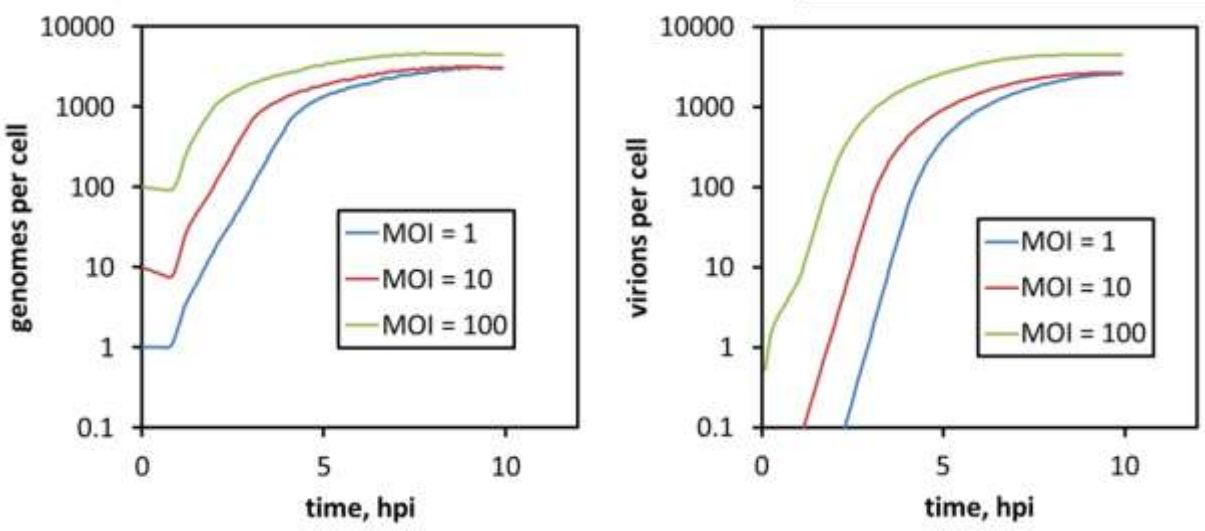


Figure 3.3-6 Different MOI approach the same genome levels

The model was parameterized to match experimental data from an average of ~1 particle per cell, but we also considered the behavior of different initial conditions throughout the parameterization process. The model converges to a similar answer for Medium and Low MOI, but High MOI produces higher genome levels and particle numbers. Interestingly, we saw this

same trend in Chapter 2, Figure 4 where resistant cells were pre-treated with human IFN and the High MOI virus was able to overcome this host antiviral response. We did not include the full complexity of the host-response, however the simulation qualitatively matched experimental observations, supporting the simplification we used to model the host response to VSV infections.

Additionally, our model does not predict the decrease in delay time as MOI increases, observed in infections in Chapter 2. The simulation delay to genome replication is essentially determined by the rate of the polymerase to produce the anti-genome and subsequent genome. In the experimental system, we suspect the delay is more controlled by early protein expression, and perhaps even the establishment of localized replication structures as observed in experimental systems (28). The qualitative agreement of experiment and model support the overall model structure and parameter combinations chosen for the VSV-host system.

3.3.8 Predicting gene insertion at different locations

The above model formulation section describes the determination of parameters for the wild-type virus genome. Using the parameters and reactions described above we can begin to simulate gene-insertion mutants. Because VSV gene order greatly affects fitness, the addition of new genes changes the normal infection process. Fluorescent protein genes can be added relatively easily via the reverse genetics system and have been shown to attenuate the virus growth, so we chose to predict viral growth after the addition of fluorescent protein genes. We assumed that the fluorescent proteins do not interact directly with host or viral intermediates; the only inhibition of viral function should be the result of transcriptional expression changes and the competition for ribosomes. Figure 3.3-7 shows the model prediction for two VSV mutants, RFP1 and RFP5, which include the RFP gene at position 1 or 5 in the genome.

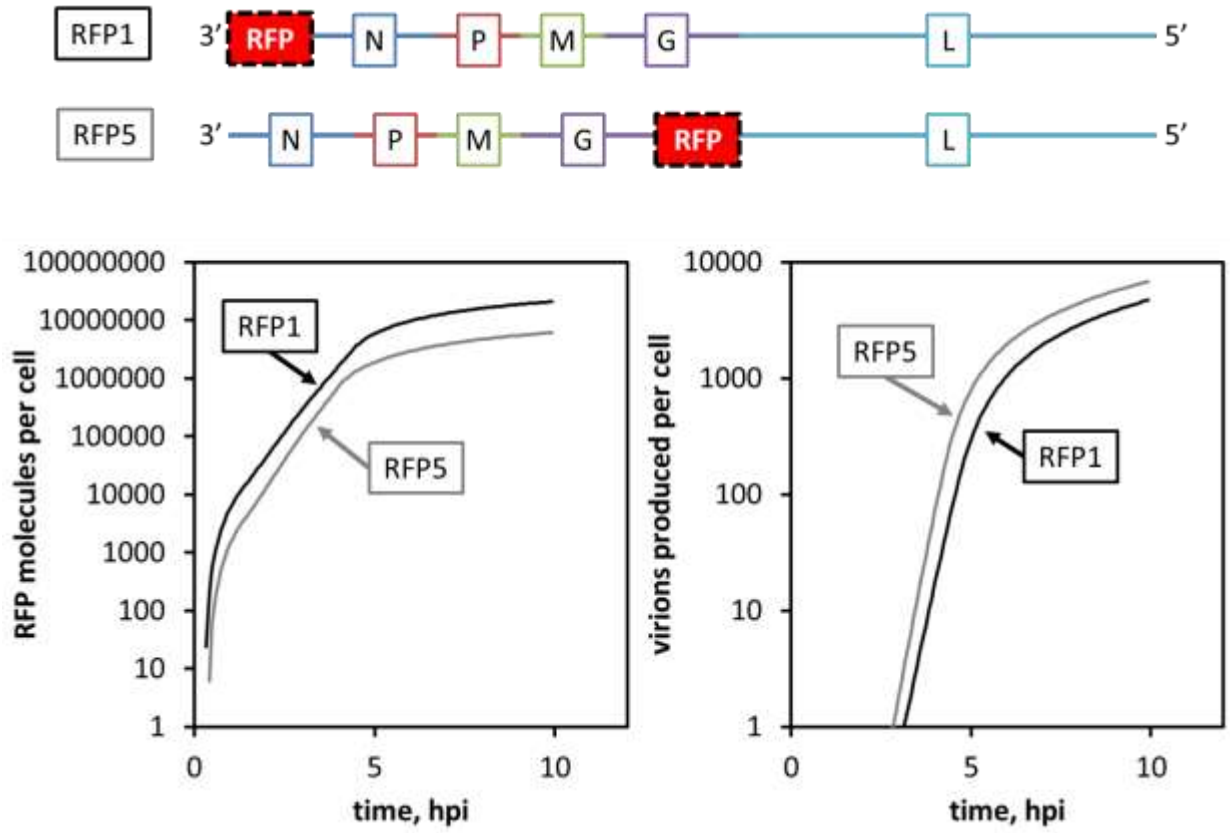


Figure 3.3-7 Model predicts RFP1 virus will be brighter than RFP5, but produce fewer infectious virions.

As shown, the model predicts that the virus with fluorescent protein in the first position will express more RFP molecules per cell, but the cost is reduced particle production. The RFP5 virus produces less RFP signal, but grows better than RFP1 virus. Current data suggests that this is true in wet bench experiments, though it is difficult to compare available strains because the fluorescent proteins are not in the exact same genetic background.

3.4 Discussion

In this work we generated a mechanistic model of VSV infection based on known VSV mechanisms and experimental observations. The computational model is genome based, meaning that adding new genes to the viral genome is very simple and requires little knowledge of the

viral replication cycle, assuming that the gene products do not interact with viral components. This model is complementary to VSV experiments and allows for new mechanisms to be incorporated. The model allows for rapid computational predictions of VSV infections to help design hypothesis-driven experiments.

In order to enable easy modification of the model we developed a unique simulation environment. The simulation allows items such as mRNA and genomes to carry information as well as act as chemical reactants in mass-action kinetics. This modeling style lays the groundwork for future virus simulations, which require flexibility as new mechanisms are discovered. There are some limitations for the current model, notably the solution method should be more accurate and more stable. The current model uses Euler forward integration. Further updates should use a more stable and accurate numerical integration strategy. An obvious and easy next step is to use Runge-Kutta integration.

A previous model (19) included a mechanism to convert host ribosomes to viral ribosomes by interaction with viral M protein (29). However, when we tried to impose this constraint the simulation was unable to produce sufficient amounts of N protein early enough to begin replication. We expect that this mechanism is important later in the simulation but until that time VSV mRNAs are translated at a lower rate. Future experiments may investigate the quantification of viral translation rates in the presence or absence of M protein mediated ribosomal conversion.

The true host response system is much more complex and is beyond the scope of this work. However, the innate response network would benefit from a similar flexible simulation environment as new mechanisms are often discovered. Further, the VSV model may act as a way to interrogate an innate response model and help determine potential mechanisms or hypothesis.

Comparing VSV infections on immune stimulated cells to model simulations would help identify the quantitative effect of the immune response on VSV infection.

By introducing a host response to genome numbers in the cell and choosing the response functions appropriately we are able ensure that the simulation is self-regulated, meaning optimal virus genome production is between “too fast” or “too slow” production. If genomes are produced “too fast”, then the host will shut down translation early and little virus will be produced. If genomes are produced “too slow” then they will not reach a sufficient level before the host responds and shuts down replication. This distinction is important because the model could be optimized with the intent of studying viral evolution. While the wild-type VSV is likely near its evolutionary optimum (for cell culture systems), other viral mutants may have more options to improve their performance. It is difficult to determine the evolutionary optimal or selective pressure that should be used to evaluate virus growth, even in simple cell culture systems. However, we feel that by exploring evolution with self-regulated models will help design and troubleshoot the engineering of the virus.

The lack of quantitative protein data for VSV infections leads to difficulty in estimating some parameters. In this case the VSV protein production rate is not well measured. Our model uses quantitative information such as the number of protein molecules in a virion, and to accurately describe the infection it is necessary to know how many proteins are produced at any time. As described above, the model was in some cases limited by the number of available proteins. It is not yet known whether proteins or other factors are growth limiting. Combining mechanistic models with quantitative data help us to determine limitations of the infection, which can be used to generate new hypotheses and design new experiments.

3.5 Supplemental Figures

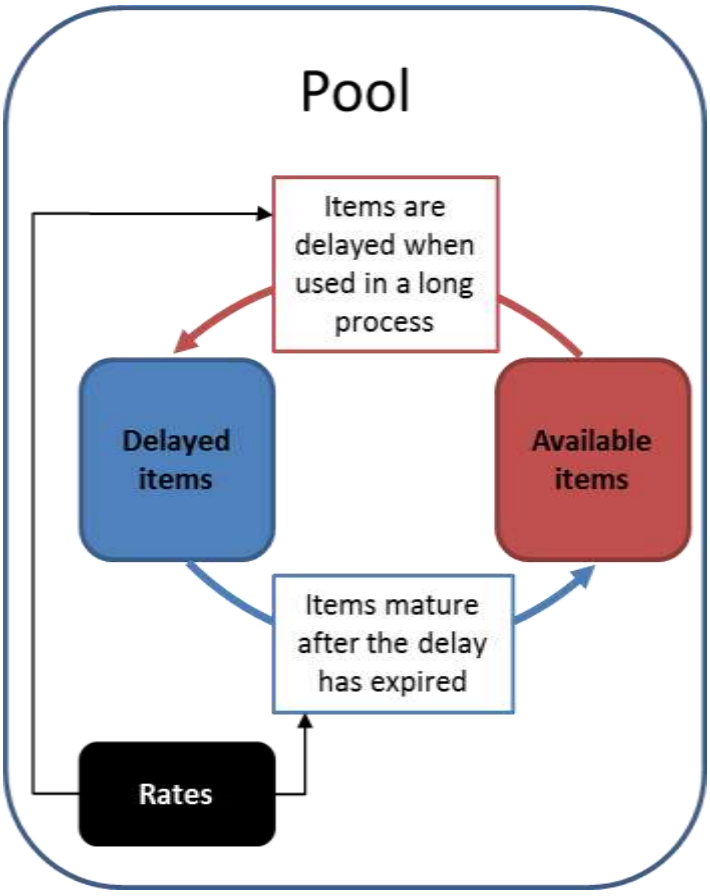


Figure 3.5-1Generic pool structure used to account for infection intermediates. Items that are visible to the simulation are included in “Available Items”. Newly produced items may be delayed and are stored in “Delayed Items”. Finally, the “Rates” structure is used to account for rates of change of intermediates.

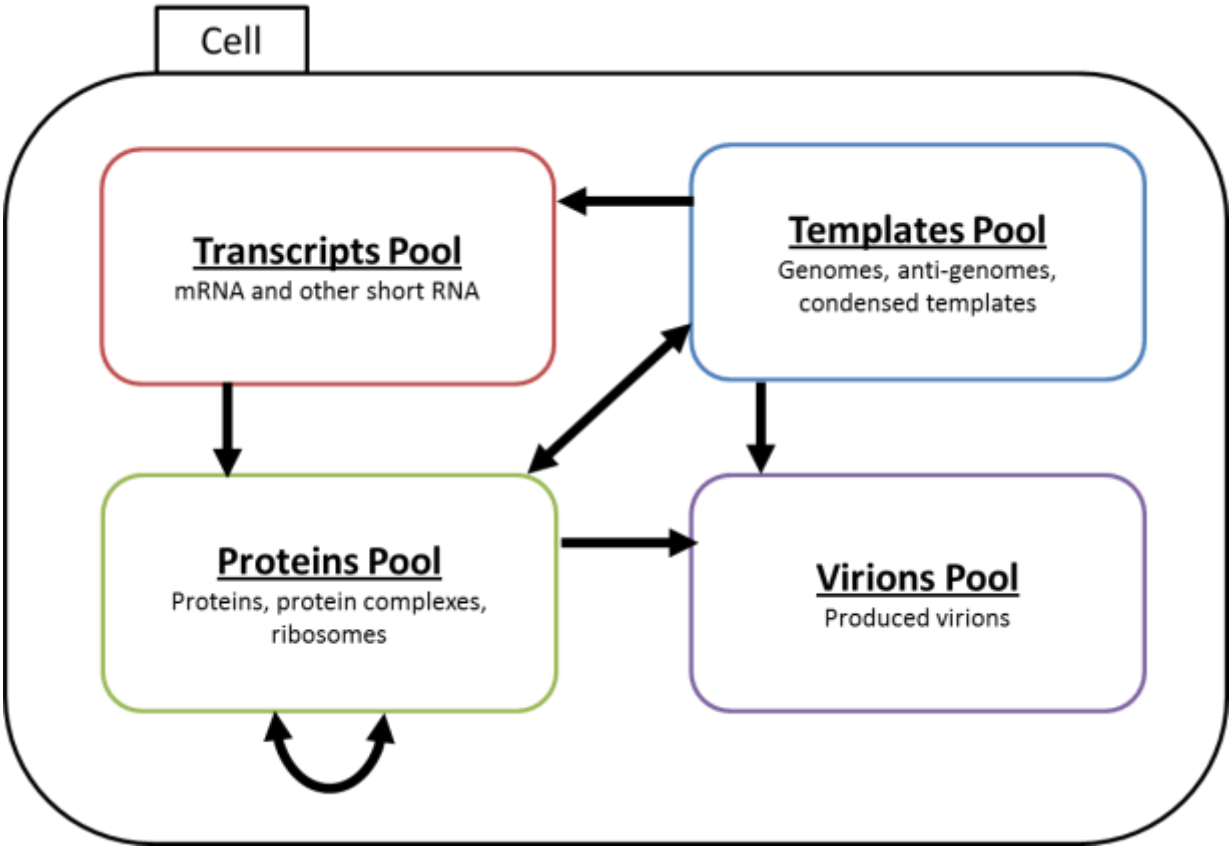


Figure 3.5-2 We simulate a cell containing distinct pools with specific contents. At every time step, the defined reactions are used to calculate rates of changes based on concentrations defined by the rate calculation. The arrows show the interactions that occur in a simulation of VSV infection. For example, templates are used to make transcripts or can be combined with proteins to produce virions. Virions are created in the simulation but not used to modify other pools.

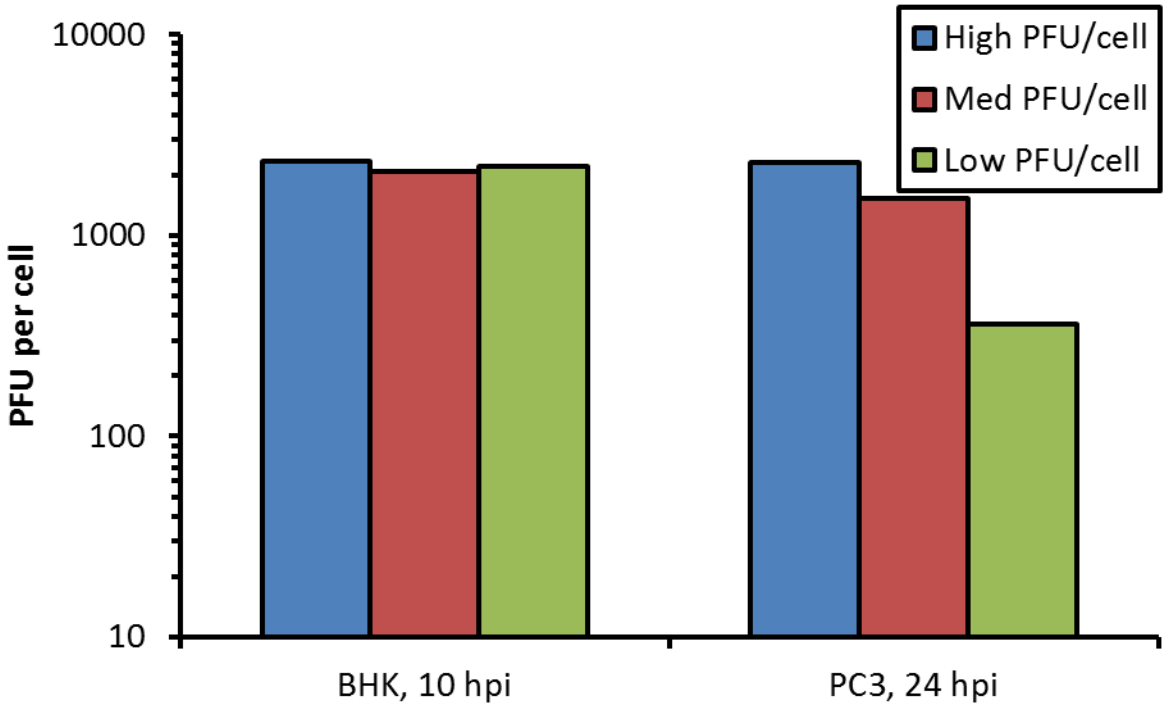


Figure 3.5-3Production of VSV at the end of infection. In BHK cells the production is independent of MOI, while in PC3 higher MOI is able to produce more particles at 24 hours.

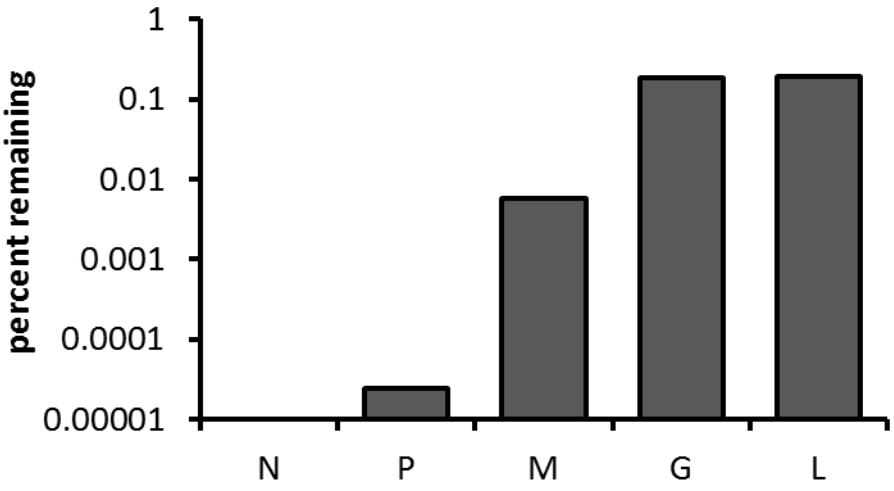


Figure 3.5-4 Simulated data used to calculate predicted percent of viral protein remaining in cell. We calculated how much protein was left after all particels were produced. Note that N protein is essentially 0.0. After that, P proteins (466 molecules per virion) are the most used proteins for producing viriosn, relative to the production level.

3.6 Acknowledgements

Collin Timm acknowledges his co-author Jay Warrick (JW) who contributed equally to this work. CT and JW developed the VSV model based on previous models. JW developed framework for model. CT developed plots and tests in this document.

3.7 Methods

3.7.1 Reaction formulation

Viral processes and intermediates are modeled as chemical reactions. The reactions are defined with the following generic form:

$$w \cdot A, (\alpha, pool_A) + x \cdot B, (\beta, pool_B) \rightarrow y \cdot C, (t_C, pool_C) + z \cdot D, (t_D, pool_D) \ , \quad k_{rxn}, \ r = f_x$$

In the standard reaction A, B, C, and D are potential reactants and products. The stoichiometry of the reaction is denoted by w, x, y, and z. The species in the reaction are followed by two quantities in parenthesis. For reactants, the quantities α and β represent the order of the reaction, to be used in calculating a rate based on mass action kinetics.

A list of reactions used for the simulations described here:

3.7.2 Maximal loading rate determined by spacing on template

The processes of transcription, translation and replication are similar in that a polymerase molecule (either viral RdRP or ribosome) interacts with a template to produce new items. The template cannot be infinitely loaded, rather there is a maximal association rate which is controlled by the concentration of available polymerase and the spacing of the polymerase on the

template. We define the rate of these reactions to approach a maximum value defined by the minimum spacing of a single polymerase molecule:

$$r = \frac{1}{\frac{s}{\mu} + \frac{1}{k \cdot f}} n_t$$

Where s is the spacing between polymerases, μ is the processivity rate, k is the reaction rate constant, f is a function to describe the concentration dependence of the reaction, and n_t is the number of available templates. As shown, the reaction rate is thus dependent on concentration but only approaches a maximum defined by the maximal loading of templates.

3.7.3 Delayed ODE solution

For each reaction in the simulation the reaction rate is calculated based on the specified function. The simulator modifies the rate based on stoichiometry, and the rate of change is added to the "rates" portion of the pool (Figure S1). In this way we can treat each reaction independently at each time step to update the overall rate of change for a particular intermediate. Items in the rate pool have units of item/time, so the value is multiplied by the simulation time step to perform Euler integration. The integrated change in number is added to the delay pool with an appropriate delay calculated for each item. We keep an absolute time counter on each delayed item. After the specified time has passed in the simulation, the item has matured and may be moved to the "Available Items" structure and used for calculations in the simulation.

3.7.4 Model structure

We developed our model in Java to take advantage of the string processing strengths and the object-oriented functions of the language. This format allows for easy introduction of non-interacting genes, or duplication of genes in a genome. For example, the VSV N protein at the

first position of the genome is equivalent to N protein in the last position, only the expression level and time of expression changes.

We keep track of different but related intermediates in Pools (see Figure S1). There is a pool for templates, mRNA, proteins and virions. Because mRNA are used to produce protein with the same name, the pool structure allows us to keep a constant naming scheme and simplifies coding and syntax. For example, VSV N can be used to refer to either N mRNA or N protein. By defining both a name and pool to look in we only have to query information about N.

We group viral and host mechanisms into small sets of related reactions: there are modules for transcription, translation, replication, protein interactions, viral detection, and packaging. Within each module there are multiple chemical reactions. The reactions are defined such that the simulation can identify what to look for and where to find it, and to generate products inot the correct pools. This modular grouping strategy allows us to quickly investigate different virus conditions and troubleshoot problems in the defined network. An example of this is shown in Figure 2 where we turn off everything except transcription to investigate the effects of cycloheximide treatment.

3.7.5 Reaction List

The following tables summarize the reactions that were simulated in this work.

Transcription			
1.0*VSVN1_Genome(1.0,Templates)	+ 1*Pol(1,Proteins)	-->	1.0*VSVN1_Genome(0.0,Templates)
		+	1*Le(3,Transcripts)
		+	0.75*N(93.0,Transcripts)
		+	0.56*P(148,Transcripts)
		+	0.42*M(205,Transcripts)
		+	0.31*G(318,Transcripts)
		+	0.23*L(744,Transcripts)
	+		[(0.25,3.0),(0.18,93.0),(0.14,148.0),(0.1,205.0),(0.07,318.0),(0.05,744.0),(0.17,749.0)]*Pol(Proteins)

Translation								
1.0*N(1.0,Transcripts)	+	1*Rib(1,Proteins)	-->	1.0*N(1011.0,Proteins)	+	1*Rib(1011,Proteins)		
			+	1.0*N(0.0,Transcripts)				
1.0*P(1.0,Transcripts)	+	1*Rib(1,Proteins)	-->	1.0*P(968.0,Proteins)	+	1*Rib(968,Proteins)		
			+	1*P(0,Transcripts)				
1.0*M(1.0,Transcripts)	+	1*Rib(1,Proteins)	-->	1.0*M(969.0,Proteins)	+	1*Rib(969,Proteins)		
			+	1*M(0,Transcripts)				
1.0*G(1.0,Transcripts)	+	1*Rib(1,Proteins)	-->	1.0*G(2839.0,Proteins)	+	1*Rib(2839,Proteins)		
			+	1*G(0,Transcripts)				
1.0*L(1.0,Transcripts)	+	1*Rib(1,Proteins)	-->	1.0*L(1431.0,Proteins)	+	1*Rib(1431,Proteins)		
			+	1*L(0,Transcripts)				

Replication								
1.0*VSVN1_Genome(1.0,Templates)	+							
1258.0*N(1.0,Proteins)	+	1*Pol(1,Proteins)	-->	1.0*VSVN1_Genome(0.0,Templates)	+	1*Pol(749,Proteins)		
			+	1*VSVN1_AntiGenome(749,Templates)	+	0*VSVN1_Genome_Error(749,Templates)		
1.0*VSVN1_AntiGenome(1.0,Templates)	+							
1258.0*N(1.0,Proteins)	+	1*Pol(1,Proteins)	-->	1.0*VSVN1_AntiGenome(0.0,Templates)	+	1*Pol(749,Proteins)		
			+	1*VSVN1_Genome(749,Templates)	+	0*VSVN1_AntiGenome_Error(749,Templates)		

Packaging				
1826.0*M(2.0,Proteins)	+	1*LeIG1NIG2PIG3MIG4GIG5LIG6TrC_Genome(1,Templates)	-->	1.0*LeIG1NIG2PIG3MIG4GIG5LIG6TrC_Genome_Condensed(0.0,Templates)
50.0*Pol(1.0,Proteins)	+			
1205*G(1,Proteins)	+	1*LeIG1NIG2PIG3MIG4GIG5LIG6TrC_Genome_Condensed(1,Templates)	-->	1.0*VSVN1(0.0,Virions)

Degradation		
1.0*Le(1.0,Transcripts)	-->	0
1.0*N(1.0,Transcripts)	-->	0
1.0*P(1.0,Transcripts)	-->	0
1.0*M(1.0,Transcripts)	-->	0
1.0*G(1.0,Transcripts)	-->	0
1.0*L(1.0,Transcripts)	-->	0
1.0*N(1.0,Proteins)	-->	0
1.0*P(1.0,Proteins)	-->	0
1.0*M(1.0,Proteins)	-->	0
1.0*G(1.0,Proteins)	-->	0
1.0*L(1.0,Proteins)	-->	0
1.0*VSVN1_Genome(1.0,Templates)	-->	0
1.0*VSVN1_AntiGenome(1.0,Templates)	-->	0
1.0*VSVN1_Genome_Condensed(1.0,Templates)	-->	0

Innate Response			
1.0*LeIG1NIG2PIG3MIG4GIG5LIG6TrC_Genome(1.0,Templates)	-->	1.0*Detection(0.0,Proteins)	
	+	1*LeIG1NIG2PIG3MIG4GIG5LIG6TrC_Genome_Error(0,Templates)	
1.0*Rib(1.0,Proteins) + 1*Detection(1,Proteins)	-->	1.0*RibDead(0.0,Proteins)	
	+	1*Detection(0,Proteins)	

3.7.6 Algorithm pseudocode

At each time step

Move matured items from “Delayed items” to “Available Items”

For each reaction

Calculate reaction rate based on visible (Available) items in the cell

For each reactant/product

Multiply rate by stoichiometry

Calculate delay (may be 0)

Add to "Rates" with determined delay

Multiple item quantities in "Rates" by time step (Euler integration) and move to "Delayed Items" with calculated delay

Repeat until simulation is finished ($t=t_{END}$)

3.7.7 Cell Culture

Baby hamster kidney (BHK-21) cells were maintained in 10% fetal bovine serum (FBS, Atlanta Biologicals), 1% Glutamax I (Gibco) in in Eagle's minimum essential medium (MEM; CellGro). Cells were passaged every 2 or 3 days when they reached confluency. Prostate cancer cells (PC3) were maintained in 10% fetal bovine serum (FBS, Atlanta Biologicals) in RPMI 1640 medium (Gibco). Cells were passaged every 2 or 3 days when they reached confluency.

3.7.8 Virus Culture

Vesicular stomatitis virus (30) was passaged on BHK cells at MOI =0.001 to prevent formation of DIPs. Stocks were centrifuged and filtered to remove cell debris. Stock concentration was ~10^9 PFU/mL.

3.7.9 Virus Infections

BHK or PC3 cells were plated 24 hours before infection at densities of 2.5x10^5 cells per mL or 3.5x10^5 cells per mL, respectively. Cells were infected by adding an appropriate amount of virus (MOI=1) for 1 hr of adsorption. After one hour, cells were rinsed with warm DPBS and fresh media was added. Samples were collect at specified times.

For plaque assays the same method was used, except 0.6% (w/v) agar in media was added instead of liquid media. Plaques were measured on BHK cell monolayers and counted by crystal violet staining after 24 hour incubation.

3.7.10 RNA Extractions

Prior to RNA extraction infections were halted by placing the culture plates on ice. RNA extractions were performed using the RNEasy 96 kit (QIAGEN) with the vacuum protocol according to manufacturer's instructions. The 96-well extraction format allows for rapid parallel processing of the samples. Output RNA concentrations ranged from 2-20 ng/uL.

Reverse Transcription

Reverse transcription was performed using the GoScript Reverse Transcriptase system (Promega, Madison, WI, USA). Gene specific primers for either N mRNA or the VSV genome were used in parallel reactions from RNA samples. 3.15 uL of undiluted RNA (see above) was mixed with 2 uL of 10 µM gene specific primer and incubated at 70°C for 5 min, then immediately transferred to an EtOH/ice bath for 5 min. Then 1.85 µL of 25 mM MgCl2, 0.5 µL nucleotides (10 µM each), and 0.5 µL GoScript Reverse Transcriptase was added to a final volume of 10 µL. Reactions were incubated according to manufacturer's instructions.

3.7.11 qPCR

qPCR was performed with a forward primer and reverse primer (see below) from RT reaction using the SsoFast Supermix (BioRad, Hercules, CA, USA) on a C1000 thermal cycler (BIORAD). 2 µL of 1:5 diluted cDNA or plasmid standard was added to an 8 µL reaction mix containing 5 µL SsoFast Supermix, 1.2 µL of 5 µM each primer mix, and 1.8 µL RNAse free water. qPCR reactions were performed in duplicate. The standard was a purified plasmid containing the VSV genome was used as a known standard to measure copies per mL. Primers

were optimized to detect viral RNA by dilutions of extracted viral genomes or infected cell RNA.

Target	**Primer**	**Sequence**	**Amplicon Length**
N mRNA	Gene specific RT and qPCR Reverse	ttccggatttgaggccttggtaga	119
	qPCR Forward	atccagtggaatacccggcagatt	
Genome	Gene specific RT and qPCR Reverse	atcctgctcggcctgagatacaaa	193
	qPCR Forward	gggtggtgcgatccctaatttctt	

Amplification protocol: 1) 95°C for 45s, 2) 95°C for 5s, 3) 62.5°C for 10s + plate read, 3) Goto 2 45x. Melt curve conditions were 65°C-95°C in 0.5 °C increments with 5 s step times and a plate read at each step.

3.7.12 qRT-PCR Assay Optimization

Primers were initially tested for single product amplicons by running amplified reactions on an agarose gel. Priemrs were temperature and concentration optimized using a dilution of a plasmid with a full genome VSV insert as a PCR standard. After primer and amplification conditions were optimized we tested the ability of gene specific primers to detect dilutions of viral RNA over a background of cellular RNA to determine limit of detection and linearity of the reverse transcription step.

3.7.13 Conversion to number per cell

The starting quantity of the plasmid standard was determined by measured DNA concentration on the NanoDrop and dividing by the molecular weight of the plasmid to convert concentration to copy number. Infection samples were measured using the qRT-PCR assay described above. Starting quantity per well was converted to copy number per cell using the following equation:

$$\frac{\text{copies}}{\text{cell}} = \frac{\text{Starting Quantity}}{\text{qPCR well}} \cdot \frac{\text{qPCR well}}{\text{V uL cDNA}} \cdot \frac{\text{W uLcDNA}}{\text{X uLRNA}} \cdot \text{Y uL RNA} \cdot \frac{\text{mock RNA conc}}{\text{this RNA conc}} \cdot \frac{\text{well}}{\text{Z cells}}$$

Variable	**Value**	**Description**
V	2	volume of cDNA added to each qPCR well
W	50	total volume of cDNA (after 1:5 dilution) used for qPCR
X	3.15	volume of undiluted RNA added to each RT reaction
Y	93	eluate volume from the RNEasy 96 RNA extraction protocol and includes all of the RNA extracted from cells
Z	Variable ~10^5	cell count from a given infection experiment (see infection protocol above)
this RNA conc	Variable 2-20 ng/uL	The RNA concentration of each sample was used to normalize counts relative to the mock RNA concentration

3.8 References

1. **Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD**. 1996. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. Science **271**:1582–1586.

2. **Dahari H, Layden–Almer JE, Kallwitz E, Ribeiro RM, Cotler SJ, Layden TJ, Perelson AS**. 2009. A Mathematical Model of Hepatitis C Virus Dynamics in Patients With High Baseline Viral Loads or Advanced Liver Disease. Gastroenterology **136**:1402–1409.

3. **Barricelli NA, Toombs R**. 1971. Virus-genetic Theory Testing by Data Processing Machines III . The Interpretation of Radiation Genetic and Partial Phage Experiments in T4 . General Remarks on the Theory-Testing Results. Journal of Theoretical Biology 603–623.

4. **Dixit NM, Layden-Almer JE, Layden TJ, Perelson AS**. 2004. Modelling how ribavirin improves interferon response rates in hepatitis C virus infection. Nature **432**:922–4.

5. **Adiwijaya BS, Kieffer TL, Henshaw J, Eisenhauer K, Kimko H, Alam JJ, Kauffman RS, Garg V**. 2012. A Viral Dynamic Model for Treatment Regimens with Direct-acting Antivirals for Chronic Hepatitis C Infection. PLoS Computational Biology **8**:e1002339.

6. **Reddy B, Yin J**. 1999. Quantitative intracellular kinetics of HIV type 1. AIDS Research and Human Retroviruses **15**:273–83.

7. **Sidorenko Y, Reichl U**. 2004. Structured model of influenza virus replication in MDCK cells. Biotechnology and Bioengineering **88**:1–14.

8. **Regoes RR, Crotty S, Antia R, Tanaka MM**. 2005. Optimal replication of poliovirus within cells. The American Naturalist **165**:364–73.

9. **Heinrich R, Rapoport TA**. 1980. Mathematical Modelling of Translation of mRNA in Eucaryotes ; Steady States, Time-dependent Processes and Applications to Reticulocytes. Journal of Theoretical Biology 279–313.

10. **Heyd A, Drew D a**. 2003. A mathematical model for elongation of a peptide chain. Bulletin of Mathematical Biology **65**:1095–109.

11. **Mier-y-Terán-Romero L, Silber M, Hatzimanikatis V**. 2010. The origins of time-delay in template biopolymerization processes. PLoS Computational Biology **6**:e1000726.

12. **Turner PE, Elena SF**. 2000. Cost of host radiation in an RNA virus. Genetics **156**:1465–70.

13. **Smith-Tsurkan SD, Wilke CO, Novella IS**. 2010. Incongruent fitness landscapes, not tradeoffs, dominate the adaptation of vesicular stomatitis virus to novel host types. The Journal of General Virology **91**:1484–93.

14. **Sardanyés J, Solé R V, Elena SF**. 2009. Replication mode and landscape topology differentially affect RNA virus mutational load and robustness. Journal of Virology **83**:12579–89.

15. **Holland JJ, De la Torre JC, Clarke DK, Duarte E**. 1991. Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. Journal of Virology **65**:2960–7.

16. **Lichty BD, Power AT, Stojdl DF, Bell JC**. 2004. Vesicular stomatitis virus: re-inventing the bullet. Trends in Molecular Medicine **10**:210–6.

17. **Marozin S, De Toni EN, Rizzani A, Altomonte J, Junger A, Schneider G, Thasler WE, Kato N, Schmid RM, Ebert O**. 2010. Cell cycle progression or translation control is not essential for vesicular stomatitis virus oncolysis of hepatocellular carcinoma. PloS One **5**:e10988.

18. **Roberts a, Buonocore L, Price R, Forman J, Rose JK**. 1999. Attenuated vesicular stomatitis viruses as vaccine vectors. Journal of Virology **73**:3723–32.

19. **Lim K, Lang T, Lam V, Yin J**. 2006. Model-Based Design of Growth-Attenuated Viruses. PLoS Computational Biology **2**.

20. **Lim K, Yin J**. 2009. Computational fitness landscape for all gene-order permutations of an RNA virus. PLoS Computational Biology **5**:e1000283.

21. **Cureton DK, Massol RH, Saffarian S, Kirchhausen TL, Whelan SPJ**. 2009. Vesicular Stomatitis Virus Enters Cells through Vesicles Incompletely Coated with Clathrin That Depend upon Actin for Internalization. PLoS Pathogens **5**:e1000394.

22. **Iverson LE, Rose JK**. 1981. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. Cell **23**:477–84.

23. **Alberts B, A J, J L, Al E**. 2002. Molecular Biology of the Cell, 4th ed. New York: Garland Science.

24. **Lyles DS, Rupprecht CE**. 2007. Rhabdoviridae, p. 1363–1408. *In* Knipe, DM, Howley, PM (eds.), Fields Virology, 5th ed. Lippincott Williams & Wilkins, Philadelphia, PA.

25. **Faul EJ, Lyles DS, Schnell MJ**. 2009. Interferon response and viral evasion by members of the family rhabdoviridae. Viruses **1**:832–51.

26. **Chong LD, Rose JK**. 1994. Interactions of normal and mutant vesicular stomatitis virus matrix proteins with the plasma membrane and nucleocapsids. Journal of Virology **68** :441–447.

27. **De BP, Thornton GB, Luk D, Banerjee AK**. 1982. Purified Matrix Protein of Vesicular Stomatitis Virus Blocks Viral Transcription in vitro. Proceedings of the National Academy of Sciences of the United States of America **79**:7137–7141.

28. **Heinrich BS, Cureton DK, Rahmeh A a, Whelan SPJ**. 2010. Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions. PLoS pathogens **6**:e1000958.

29. **Black BL, Brewer G, Lyles DS**. 1994. Effect of Vesicular Stomatitis Virus Matrix Protein on Host-Directed Translation In Vivo **68**.

30. **Wertz GW, Perepelitsa VP, Ball L a**. 1998. Gene rearrangement attenuates expression and lethality of a nonsegmented negative strand RNA virus. Proceedings of the National Academy of Sciences of the United States of America **95**:3501–6.

Chapter 4: Quantitative characterization of defective virus emergence by deep sequencing

4.1 Abstract

Populations of RNA viruses can spontaneously produce variants that differ in genome size, sequence and biological activity. Defective variants that lack essential genes can nevertheless reproduce by infecting cells along with viable virus. In such co-infections the defective particle production interferes with virus growth. How such defective interfering particles (DIPs) change in abundance and biological activity within a virus population is not known. To address this question, a prototype RNA virus, vesicular stomatitis virus (VSV), was serially cultured for three passages on BHK host cells, and each population was subjected to Illumina sequencing. Reads from the initial population, when aligned to the full-length viral sequence (11,161 nt), distributed uniformly across the genome. However, as the population was passaged two plateaus in read counts appeared toward the 5' end of the negative-sense viral genome. Analysis by normalization and a simple sliding-window approach revealed plateau boundaries that suggested the emergence and enrichment of two truncated species with genome sizes of medium (~5900 nt) and short (~4000 nt) length. Relative measures of full-length and truncated species based on read counts were validated by quantitative reverse transcriptase PCR. Further, particle counts from transmission electron microscopy, coupled with infectivity assays, linked the rise in smaller genomes with an increase in truncated particles and interference activity. In summary, variation in deep sequencing coverage can simultaneously reveal the size and relative abundance of different truncated genome-size variants, reflecting how virus populations can change and adapt.

4.2 Introduction

Defective interfering particles (DIPs) are truncated forms of viruses that emerge during viral infections. DIPs cannot replicate on their own, but upon co-infection with intact virus, they can divert the viral replication machinery to make defective genomes and particles, ultimately at the expense of intact virus production (1, 2). In culture, DIPs have been observed for most classes of DNA and RNA viruses (2–4). In nature, DIPs have been discovered in an outbreak of influenza A virus in chickens (5) and in West Nile virus isolates from avian hosts (6). In human hosts, clinical isolates from patients infected with Dengue virus (7) and influenza A virus (8) provide evidence that DIPs and their genomes can be transmitted between patients.

DIPs have been characterized by their genome length and sequences (9–13), their particle sizes and morphologies (14–18), and their biological activities (19, 20). Few studies have combined genome, particle and activity measurements of mixed DIP and virus populations (21, 22), though such measures can be useful for guiding computational models that aim to elucidate how such populations may grow, change and transmit (23–26). To advance models that not only elucidate but also enable forecasting of how virus populations grow and spread, new measures that facilitate the rapid quantitative characterization of viruses and their DIPs will need to be developed.

High-throughput (deep) sequencing has been widely used to characterize the diversity of virus variants in natural infections as well as their sequence-level adaptations to antiviral drugs (27–35). Deep sequencing has also revealed central roles for DIPs derived from influenza A, Sendai, measles and parainfluenza virus 5 in binding to the cellular pathogen recognition receptor,

retinoic acid inducible gene I (RIG-I), as an initial step toward establishing an antiviral state within the host cell (36–38). In deep sequencing of viral RNA, the samples of genomes or subgenomes are reverse transcribed, amplified and sequenced. When these reads are aligned to a reference genome, their match to a region on the reference genome (coverage) can approximate the level of that region in the analyzed genome population (37). In this way, deep sequencing can potentially be used to identify genome or sub-genomic sizes, sequences and relative levels in a biological sample.

Here we employed Illumina sequencing, quantitative RT-PCR (qRT-PCR), infectivity and interference assays, and transmission electron microscopy (TEM) to track the emergence of DIPs relative to intact particles during serial transfer culture of the non-segmented negative sense RNA virus, vesicular stomatitis virus (VSV), a prototypical virus for the study of DIPs (2, 20, 23, 39). Extents of template coverage by deep sequencing revealed changing relative levels of genomic and sub-genomic templates which corresponded to changes in physical particle sizes and biological activities within the virus population.

4.3 Results

4.3.1 Deep sequencing can detect differential coverage profiles in related populations

Vesicular stomatitis virus was cultured at MOI 10 on baby hamster kidney cells (BHK-21) for three passages (P1, P2, P3), starting with an initial population (P0), as shown in Figure 4.3-1. RNA extracted from supernatants of each passage was sequenced using Illumina technology with 35 base pair long reads. More than half of the total reads from the initial population mapped to

the initial passage zero (P0) VSV consensus sequence. Unmapped reads could not be directly attributed to the host (Syrian golden hamster) for which a genomic sequence is not yet available. However, *de novo* alignment on all reads from passages showed multiple contigs that aligned to ribosomal RNA of other species including Chinese Hamster (data not shown).

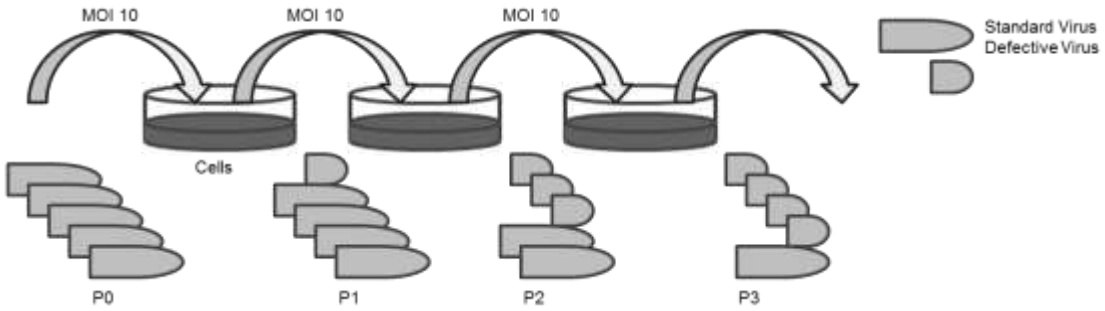


Figure 4.3-1 **Serial passage culture of VSV.** Virus was cultured on BHK-21 cells at MOI 10 for 3 passages, established conditions that promote the emergence of DIPs. RNA extracted from each passage was submitted for RNA sequencing.

In our initial alignment of P0 to the VSV Indiana reference 55.6% (508,249 of 913,387) of the total reads were mapped to the viral genome. When the extracted consensus of this alignment was used as the reference the mapped reads increased to 56.2%. For each passage total reads of $9x10^5$ to $1.7x10^6$ were obtained, with a maximum of 56% and a minimum of 1.8% for later passages mapping to the P0 consensus. Details of the mapping are in Table S1.

Base-by-base coverages for each of the 11,161 nucleotide positions of the VSV genome mapped to the P0 consensus, as shown in Figure 4.3-2, and the coverage for P0 exhibited a relatively flat profile across the VSV genome. A sharp drop in coverage of P0 between bases 1622-1625 is attributed to zero coverage at those four bases. In passage 2 different plateaus of coverage appeared, with 5-10 fold lower coverage toward the 3' end, relative to the 5' end. By passage 3 these differences were enhanced with 100-fold differences in coverage between the lowest and highest plateaus.

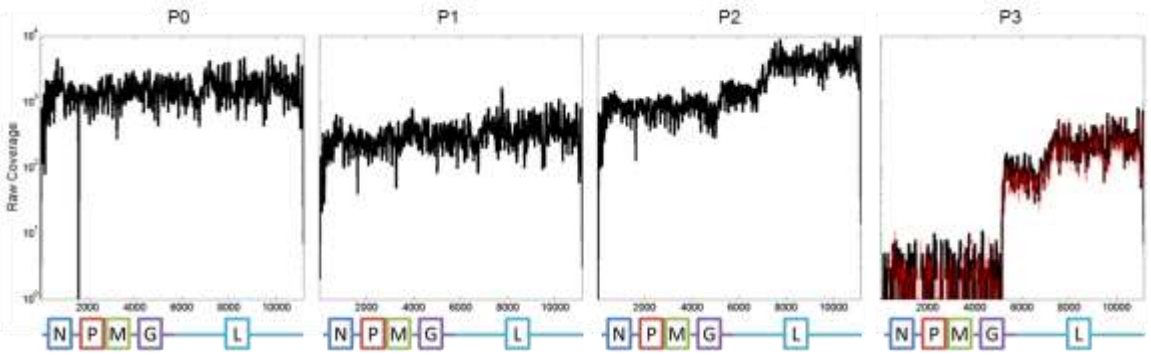


Figure 4.3-2 **Raw coverage data.** Coverage is plotted on y axis vs. position on the viral genome on the x-axis (genome cartoon below x-axis). Genome is from 3' to 5' with base 1 representing the first base on the 3' end. The P3 plot has two sequencing replicates, one in red and one in black. Notice the relatively flat coverage in P0 and the distinct plateaus that are seen in later passages.

4.3.2 Normalizing coverage to the initial population coverage reduced the noise in coverage profiles

To test whether fluctuations in coverage at the base level reflected sequence-dependent biases we normalized the number of reads at each base in P1 by its corresponding value in P0, as shown in Figure 4.3-3a. As an unbiased control we also normalized P1 to shifted P0, where P0 coverage was offset by a fixed number of nucleotide positions prior to normalization. Normalization of P1 by P0 (offset = 0) significantly reduced the bias (Figure 4.3-3a), while increasing the offset produced similar fluctuations when P1 or P0 were normalized by P0 (offset = 100 nt), suggesting sequence similarity between P1 and P0. By normalizing the base-level coverage for all passages by P0 (offset = 0) the plateaus of differential coverage became more visible (Figure 4.3-4 A).

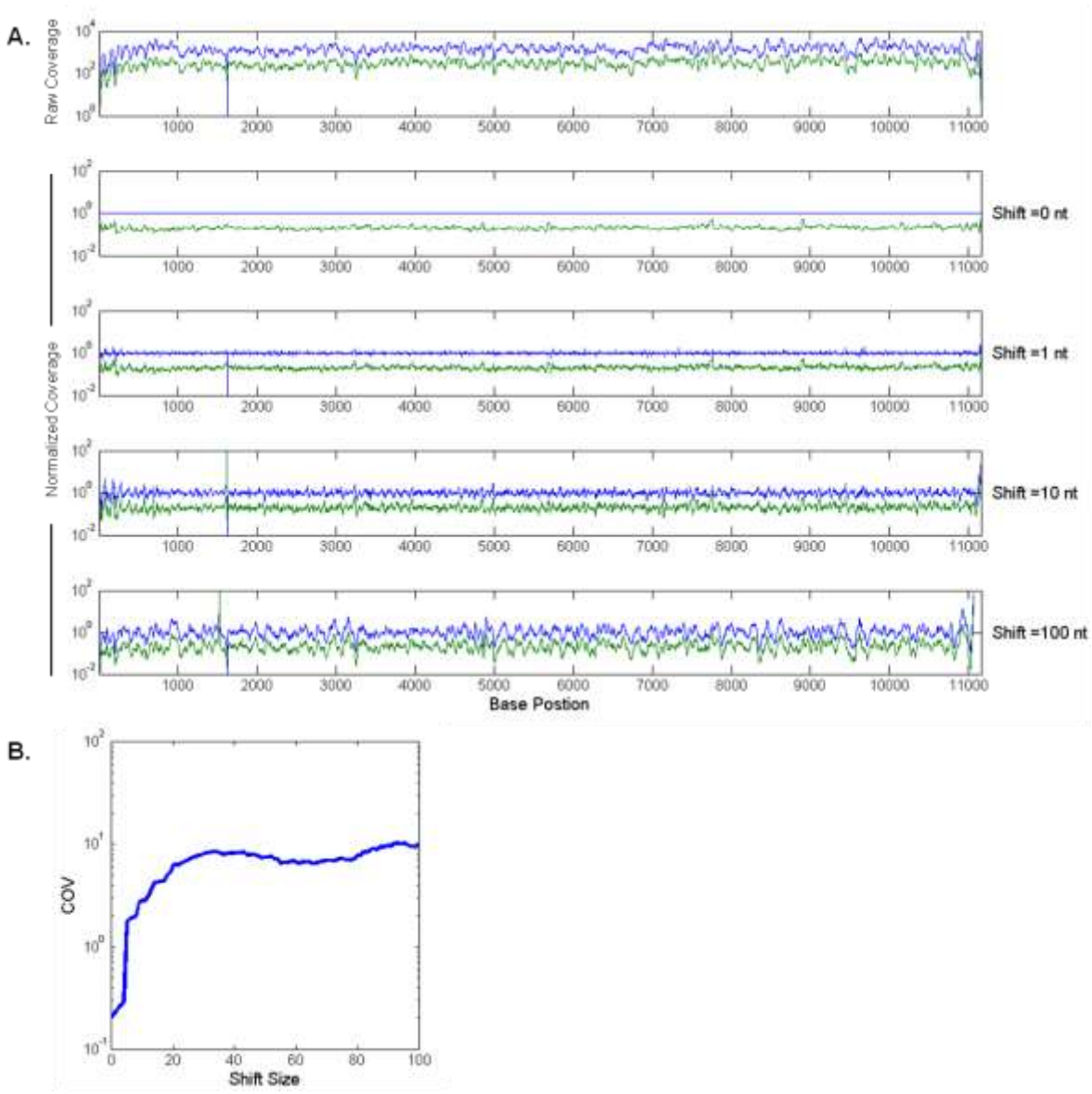


Figure 4.3-3 **Normalization by P0 coverage reduces variability** P1 coverage values are divided to P0 coverages at the corresponding positions shifted by an offset. **A)** The plots show the normalized P0 (blue) and P1(green) coverages with increasing offset from top to bottom. The x-axis represents the nucleotide positions on VSV genome. **B)** As the reference P0 coverages shifted by a higher offset, the noise in data increases, which can be seen in the increase of coefficient of variance (COV, standard deviation of normalized coverages divided by the coverage mean across genome) of P1. The COV approaches its maximum after about 30 bases. Similar behavior was observed when this analysis was performed on P2 and P3.

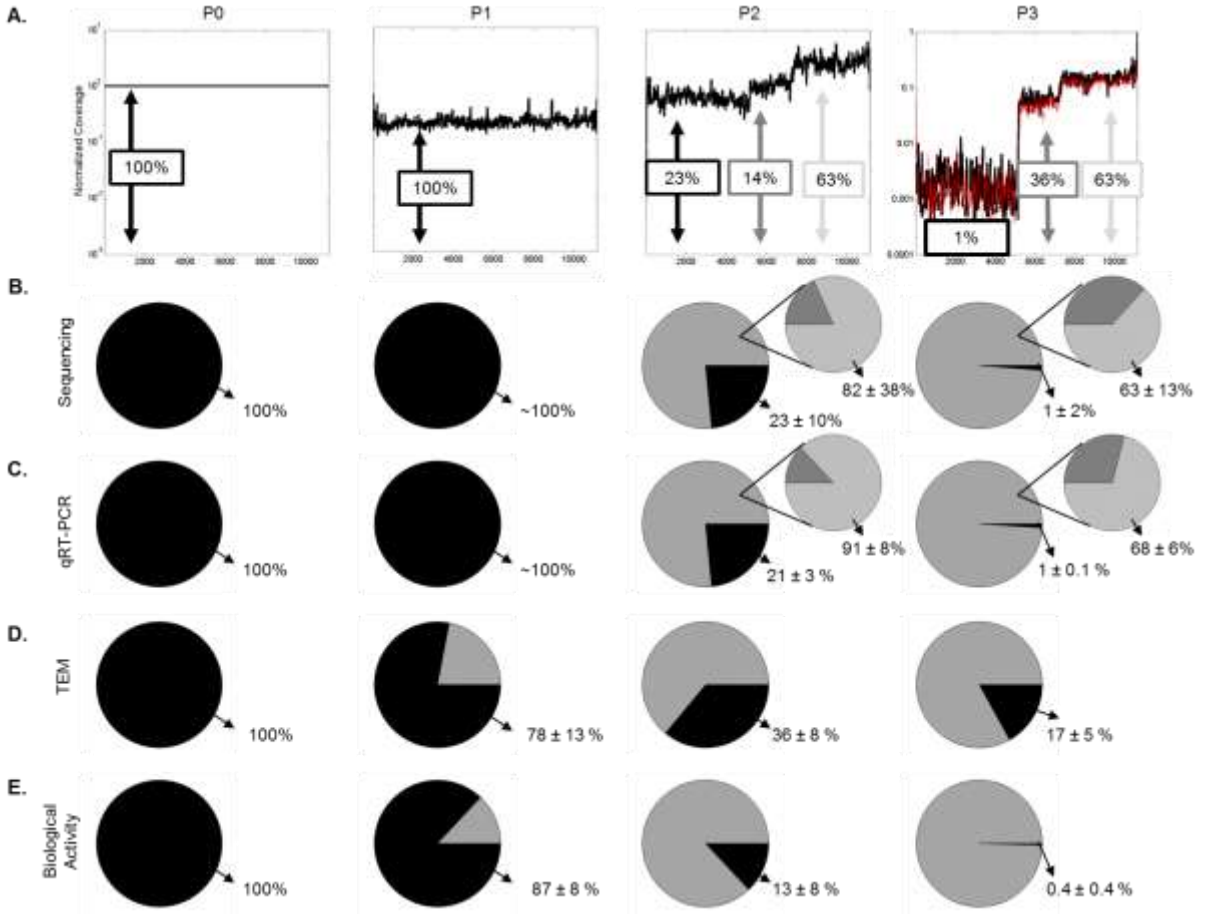


Figure 4.3-4 **Comparison of all quantification techniques. A)** Coverage values from deep sequencing were normalized to the corresponding base in P0, ignoring any zero values. The technical replicate of P3 sample is shown as the red curve. **B-E)** Population fractions are calculated for each method. The black segment corresponds to full size species, the gray in the large pie represent all defective particles. In **B** and **C** a second pie chart shows the fractions of medium and short genomes in the defective population of P2 and P3. Values reported are percent of population ± standard deviation. **B)** Fraction calculated from sequencing coverage. The standard deviation was calculated by coverage noise and propagation of uncertainty. **C)** Fractions by qRT-PCR. Standard deviation calculated by 2 replicates of extraction – qPCR workflow. **D)** Fraction calculated from TEM images. Standard deviation calculated from counts of two sets of ~15 images from each passage. **E)** Fractions calculated from plaque forming units and interfering units measured by plaque and interference assays. Standard deviation calculated from duplicate measures.

4.3.3 Coverage analysis reveals two distinct genome deletion mutants in the population

Different coverage plateaus suggested the formation and enrichment of truncated genomes associated with defective particles in the virus population. Boundaries of the plateaus, estimated as described in the Methods, highlight a 100 nucleotide long window centered at 5250, the second a 100 nucleotide long window around 6450, and finally a 200 nucleotide long window centered at 7200 (Figure 4.5-1). The boundaries at 5250 and 7200 occurred in both P2 and P3. However, the boundary at 6450 appeared only in P2. Therefore, we focused on conserved boundaries at 5250 and 7200 corresponding to defective genomes of lengths 5911 and 3961 nucleotides. Analysis of the coverage profile from P1 showed no detectable changes in coverage. It is important to note that the changes in coverage level were not abrupt; the transition between observed plateaus takes place over hundreds of bases.

4.3.4 *De novo* read assembly suggests a genome deletion mutant

Using *de novo* read assembly algorithms available in CLC Genomics Workbench (CLC bio, Denmark), 8 to 20 contigs were obtained at each passage. One of the contigs in P3 was a 6025 nucleotide sequence containing the 5’ end of the VSV genome, including most of the L gene (Figure 4.5-2). The 3’ end was complimentary to the 5’ end across a 33 nucleotide long section, as previously observed for other DI particles (12). This *de novo* sequence is likely a DI genome and corresponds to additional coverage in P2 and P3 as described above. However, we were unable to find by *de novo* alignment of a ~4000 nt length species (short DIP) corresponding to the high coverage at the 3’ end.

4.3.5 Relative coverage levels were used to quantify population fraction of full length and genome deletion mutants

The emergence of different plateaus (Figure 4.3-4a) is consistent with the enrichment of two truncated-genome populations, one medium length (5911 nucleotides) and the other of shorter length (3961 nucleotides). Relative levels of each species were estimated by average read coverages across each of three regions (bases 1 – 5250, 5251 – 6450, and 6451 – 11161), and this average was used to calculate population fractions (Figure 4.5-3). In Figure 4.3-4 b, the larger pie-charts represent the fractions of full-length genomes (black) and truncated genomes (grey), and the smaller pie charts show how the truncated genomes were distributed between short genomes (light grey) and medium-length genomes (dark grey). Over the course of three passages, the fraction of species carrying full-length genomes declined as the species carrying the shorter length genome emerged to become the dominant species (Figure 4.3-4 b).

In addition to P0 – P3 samples, a technical replicate of the P3 sample (shown in red in Figure 4.3-2 and Figure 4.3-4 a) was sequenced. The technical replicates show good reproducibility of the sequencing to measure differential coverage and both replicates yielded similar population fractions (Table S3).

4.3.6 Complementary measures agree with quantification by relative coverage

We employed qRT-PCR, TEM and activity assays (plaque and yield reduction assays) as complementary measures of how the virus populations changed with passage. A qRT-PCR assay was designed to measure the relative abundance for genomic species corresponding to each plateau region, with primer sets targeting the M mRNA region of the full-length genome, the medium-length species, and the L mRNA region associated with the shortest sub-genomic species. To compare levels of truncated and full-length genomes estimated from sequencing with

an independent measure we normalized all qRT-PCR values to their measures in P0, which served as our base case pure population of full-length genomes. Figure 4.3-4c shows the distribution of the genome population measured by qRT-PCR. The qRT-PCR results were consistent with estimates from sequencing in P2 and P3. Measures of relative abundance for genomic species from the same locations in P1, did not show a significant change in levels between locations.

TEM of populations enabled visualization and approximation of relative levels of full size and truncated virus particles (Figure 4.5-4). Here, truncated particles were treated as a single class, without attempting to distinguish between ones carrying medium-length or short genomes. The fraction of truncated VSV particles relative to full size particles increased with passage (Figure 4.3-4d), consistent with the overall trend observed by sequencing and qRT-PCR. However, the rate of enrichment of truncated-particles appeared to lag behind the genomic measures, potentially reflecting the uncertainty in distinguishing particles that carry full-sized versus medium-length genomes.

Over the course of the three passages levels of infectious virus declined more than 5000-fold, from 7×10^9 to 1×10^6 PFU/ml (Table S4), suggesting that the emergence of DIPs interfered with infectious virus production. Measures of interference activity by the yield-reduction assay indicated a significant rise in interfering units, with a maximum at P2 (Table S4). These infection and interference results were combined to provide estimates of DIP fractions, which increased with passage, consistent with genome and particle measures (Figure 4.3-4e).

The correlation between virus population distributions obtained by deep sequencing and genome, particle and activity distributions highlights the accuracy of deep sequencing to quantify virus population distributions and their dynamics. Moreover, the similarity between genome and

activity distributions suggests that all intact genomes are infectious while all defective genomes are interfering with infection. In order to test this, the absolute particle, genome and activity levels were compared.

4.3.7 Quantification by relative coverage is limited by statistical and empirical measurement noise

In contrast to P2 and P3 samples, we did not find evidence for different coverages in the P1 sample. Assuming normally distributed noise we calculated a statistical lower detection limit of ~1.5% of additional reads attributed to genome deletion mutants (Figure 4.5-5a). However, when compared to the complementary measures described above, the empirical lower detection limit in this analysis was determined as a 10% representation of genome deletions (Figure 4.5-5b). The discrepancy between statistical and empirical limits can be attributed to non-normally distributed noise in sequencing data, the inherent measurement variation of the complementary measures (qRT-PCR, TEM and biological activity), or the gradual change (over 10s to 100s of bases) in coverage at the plateau boundaries. Based on this analysis we concluded that the level of defective genomes in P1 was less than 10%.

4.3.8 Absolute quantification of population suggests a strong correlation between genome structure and viral activity

In addition to measuring relative levels of different species with a population passage, TEM and qRT-PCR measures allowed us to estimate the absolute particle and genome concentrations in passaged samples, while the plaque and interference assays provided infectious and interference measures. Concentrations of full-size particles, full-length genomes, and infectious units were within a factor of 10 from P0 to P2 (Figure 4.3-5a), and defective virus particles, truncated

genomes, and interference units were within a factor of 10 from P1 to P3, all with a maximum at P2 (Figure 4.3-5 b).

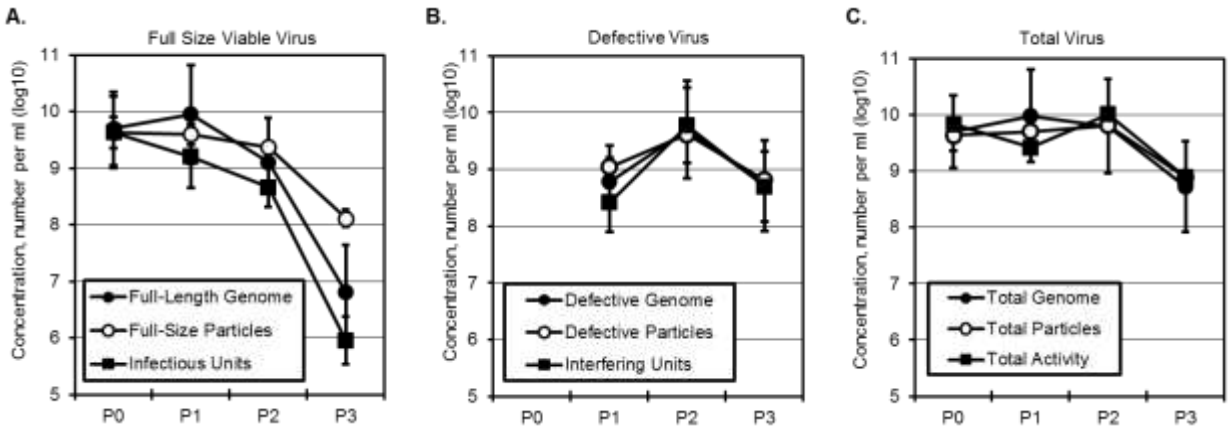


Figure 4.3-5 **Absolute virus and DIP concentrations using different methods** Genomes were measured by qRT-PCR, particle by TEM and activity by plaque assay or interference assay. A) Full length genomes, particles and infectious units decline at a similar rate. B) Defective genomes, particles and interfering units follow a similar trend. C) Total genomes, particles and activity follow similar trends. Note that in P2 and P3 the counts are dominated by defective particles.

The level of infectious particles followed an 8-fold lower but parallel trend with full-length genome levels (Figure 4.3-5 b), and full-length particle levels dropped to a lesser extent across passages than viral genomes and infectious virus. This may suggest that some full-length particles contain complete or mostly complete genomes that are not infectious. For defective particles there was greater agreement between particle, genome and activity levels (Figure 4.3-5b), evidence that most defective particles were able to interfere with infectious virus production. Interestingly, following their initial appearance and amplification, defective virus levels dropped from P2 to P3. Likewise, total virus particle levels dropped from P2 to P3 after being stable from P0 to P2 (Figure 4.3-5c). These parallel drops in total and defective virus levels reflect that DIPs can interfere with production of infectious virus, and thereby indirectly and adversely affect their own production.

Other studies (27, 32, 40) suggest a correlation between absolute measure and mapped reads, but when we compare our mapped reads with other absolute measures we did not see a correlation (Figure 4.5-6). Therefore, we were unable to use mapped reads as an absolute measure of genomes in the population.

4.4 Discussion

The growth dynamics of virus populations are important for understanding how viruses adapt to different selection environments. Deep sequencing provides a high-throughput platform to track the virus population dynamics. In this study, we tested the accuracy of deep sequencing in the quantitative characterization of virus population dynamics compared to other virus quantification methods.

The normalization of the coverage profile to that of initial population (P0) reduced the noise and allowed for the identification of genomic subpopulations and the estimation of population distributions. The good correlation between virus population distributions by sequencing and by other independent methods highlights that deep sequencing can accurately quantify the relative levels of genomic variants. While we were unable to resolve the presence of DIPs in P1 by deep sequencing, TEM and interference assays suggest their presence in P1 at a low level. Based on these other measures we determined that the lower detection limit of deep sequencing to distinguish between two genome subpopulations (intact and defective genomes) was about 10%, meaning that genome deletion mutants must comprise at least 10% of the population in order to be detectable based on sequence coverage.

Previous studies have tested the quantification accuracy of deep sequencing by comparing it with other genomic techniques including qRT-PCR and microarrays (27, 32, 41–44). However, these

techniques have similar limitations to deep sequencing, such as the variability in primer efficiency and the sensitivity to the cDNA amplification step. Here we compared deep sequencing with independent activity and particle measures and linked biological activity and particle size (phenotype) to genomic structure (genotype).

The coverage changes between observed plateaus were not discrete: the change occurred gradually over ~200 nt stretch of the genome. This gradual change may be due to the presence of multiple genome lengths of similar size, as observed in polio virus defective interfering particles (45, 46). The existence of a distribution of lengths suggests that multiple defective particles can arise in culture and have similar fitness.

We observed good agreement between biological activity and physical particle measurements for early passages and significant disagreement in later passages. However, the total particle counts produced did not change significantly until P3, which we attributed to the interference of DIPs on their own replication (23). These results suggest that there is a finite resource pool that is available for particle production. Further work to elucidate limiting resources in viral particle production can be facilitated by studying defective particle interactions with infectious virus.

While deep sequencing has been used primarily to detect and identify viral genome sequences, its ability to quantify the population distributions has been in part limited because of the noise in the coverage profiles. Here, we observed a similarity in the noise patterns of related samples and found that normalizing coverage measurements to a common ancestor reduced much of the sequence-dependent noise, allowing for unbiased species identification and more precise quantification. Such a normalization strategy can be applied to reduce the noise in chronologically linked biological samples, which may aid in the quantitative analysis of the population dynamics.

While we were able to quantify bulk population fractions, further advancement is required to identify individual genomic segments in the population. As sequencing technology develops and the length of the base reads increases we can study the molecular sequence makeup of the population. The discovery of unique templates in the population will help identify the contribution of individuals to observed population dynamics.

4.5 Supplemental Figures and Tables

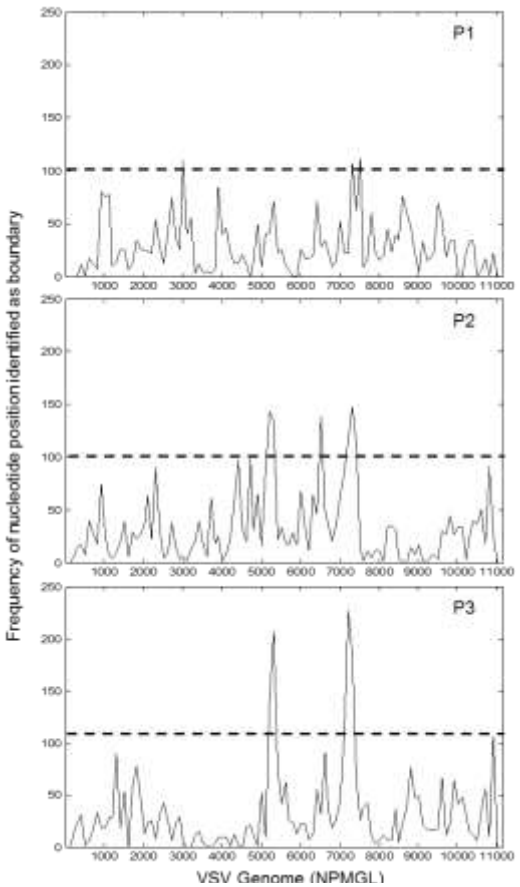


Figure 4.5-1 **Plateau boundary determination**

Figure 4.5- **1 Plateau boundary determination**

The plateau boundaries were determined by analyzing significant changes in coverage across the genome. A t-test was used to determine significant changes in coverage upon expansion of a window centered around each base. When the change in coverage was significant, the base was recorded as a boundary.

The frequencies are shown for P1 to P3. Using a frequency cut-off of 100 counts we determined bases that were likely the boundaries of plateaus. The boundaries in P2 occurs at nucleotide positions in the ranges of 5150-5300, 6400-6500, and 7000-7400. In P3 high frequency bases are observed at around 5200-5300 and 7150-7250. In P1, the peaks are around 3000 and 7400, but the frequencies are relatively lower than the ones in P2 and P3. Therefore, we conclude that we cannot confidently determine plateau boundaries in P1.

Specific parameters

Alpha = 1e-11, Expansion size = 10, Window size = 400, Step size = 5, Histogram bin size = 100

Algorithm Pseudocode

*algorithm available upon request OR attached as supplemental

For every fifth base

local coverage (Cov) = average coverage in 400 base neighborhood

test coverage (testCov) =average coverage in 400 base neighborhood + 10 bases

if testCov != Cov (not equal)

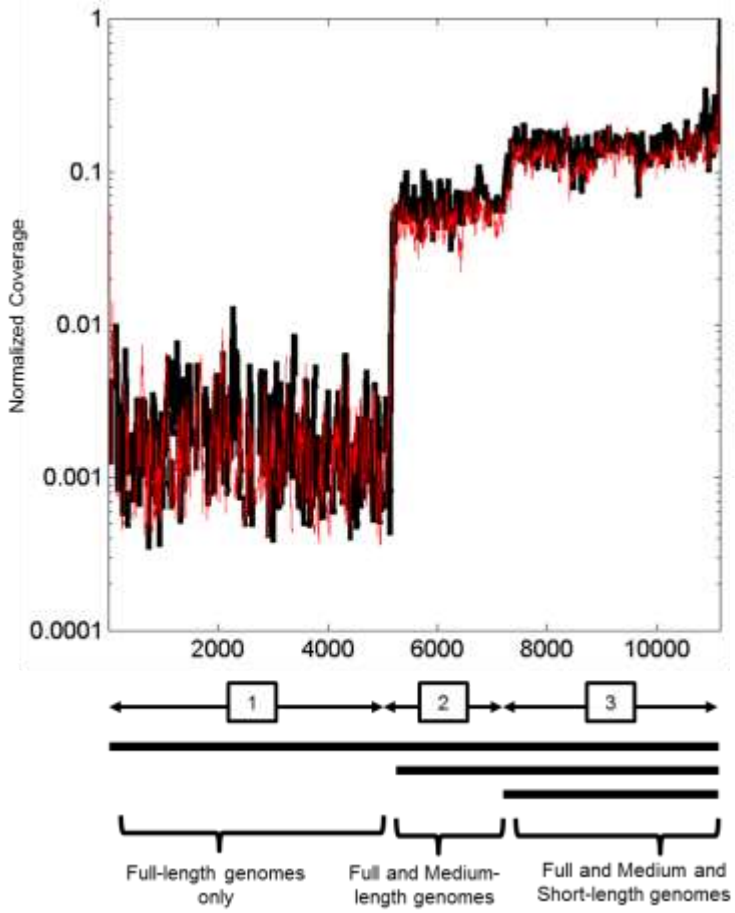
record base added as plateau boundary

Count the times each base recorded as a boundary

Repeat by expanding to the left (add 10 bases to the left of initial position)

Plot histogram

Figure 4.5-2 **Contig detected by *de novo* alignment has complementary ends and similarity to full length virus.** The *de novo* alignment tool in CLCBio was used with default parameters on P3 reads to generate this contig. The first base of the contig aligns with the 5,116 base of the parental genome (P0). There is a short segment where there is no alignment, then a long stretch of perfect match. Finally, the DIP sequence aligns to the end but is truncated before the final base of the parent sequence.



$$f_{full} = \frac{Coverage[1]}{Coverage[3]}$$

$$f_{med} = \frac{Coverage[2]}{Coverage[3]} - f_{full}$$

$$f_{short} = 1 - f_{full} - f_{med}$$

Figure 4.5-3 **Calculation of population fractions from coverage data.** Data from P3 is shown. The coverage at the right contains coverage from short, medium and full length genomes. We used the equations above to determine the fraction (f) of the population corresponding to full, medium, or short length genomes. These equations were also used with propagation of uncertainty rules (assuming no covariance) to determine standard deviations on the population fraction estimates.

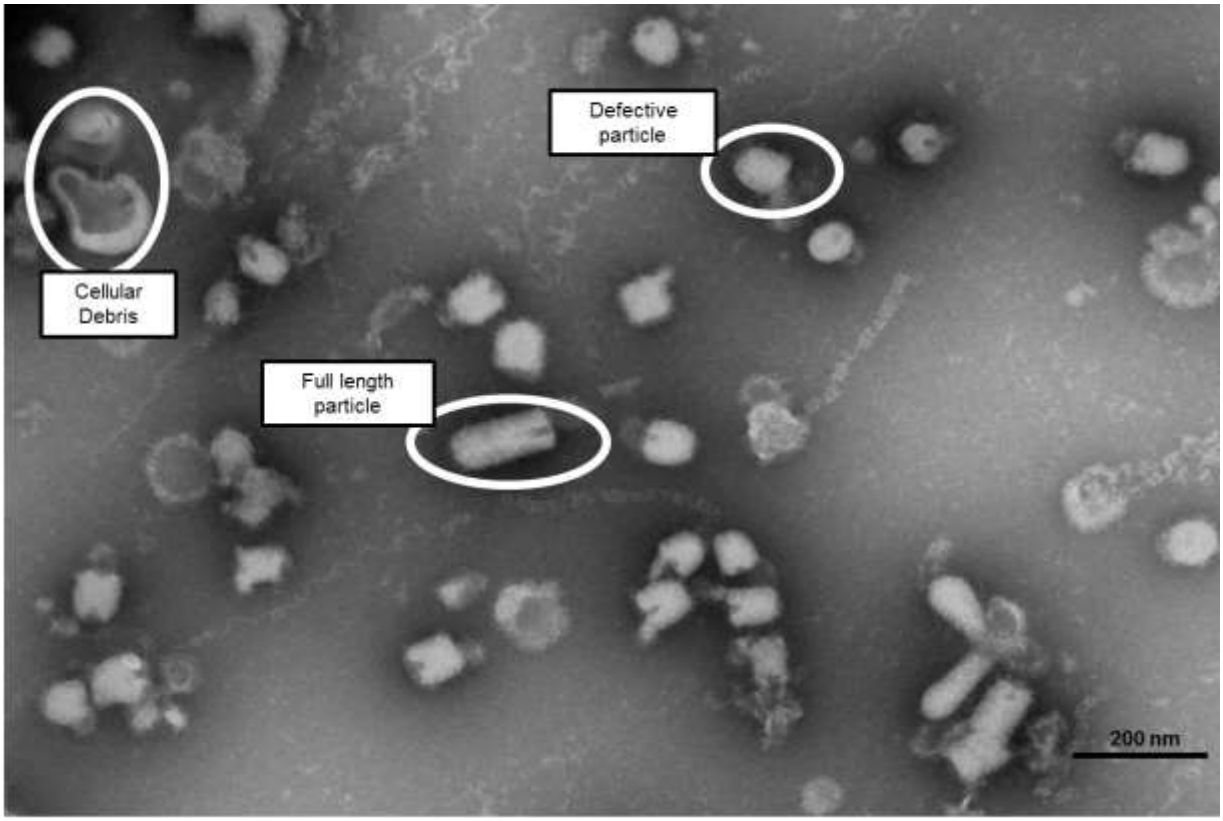


Figure 4.5-4 **TEM Image of P3 Virus and DIPs.** In this image we see cellular debris, full-length VSV particles and short defective particles. The scale bar is 200nm in the lower right corner of the image.

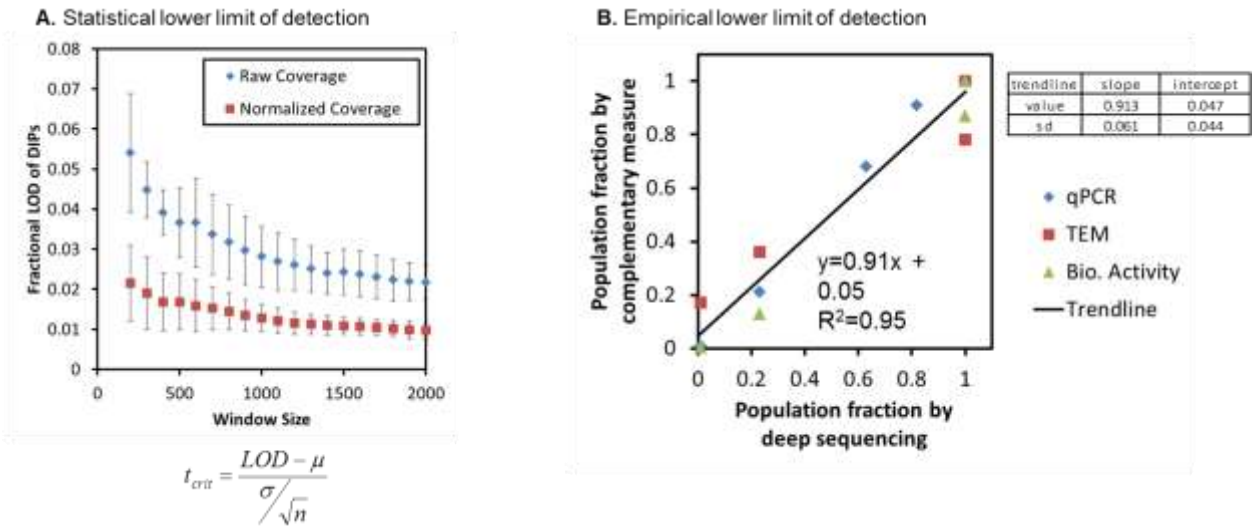


Figure 4.5-5 **Limit of detection calculated two ways.**

A. The statistical limit of detection was determined by measuring the mean (μ) and standard deviation (σ) across given window size then solving for LOD in the equation at a confidence limit α=0.001 and n = (windowSize – 1). The equation is the standard equation for a t-test comparison. All coverages were extracted from P1 raw coverage or coverage normalized to P0. Points with error bars are the average and standard deviation of 9 calculations of LOD for the given window size, spread at 1000 nt intervals across the genome. **B.** The x-axis is the population fraction as measured by deep sequencing, and the y-axis for the 3 complementary measures (see Figure 4.3-3). A trendline was fitted through all data points to determine slope and intercept. The empirical limit of detection was calculated by determining the x-intercept (predicted sequence measure when other measures are 0). Propagation of error was used to determine a value of 0.05±0.05, thus we concluded the empirical limit of detection is about 0.1, or 10%.

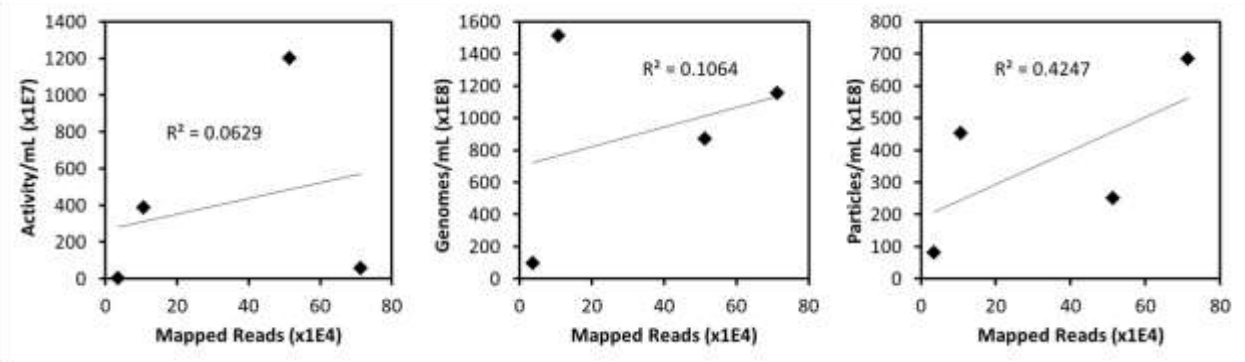


Figure 4.5-6 **Mapped reads do not correlate with other absolute measures**. The total number of mapped reads (x-axis) were plotted against biological activity per mL, genomes per mL (by qRT-PCR), or particles per mL. We do not see a strong correlation between mapped reads and any of the absolute measurements.

Table 4.5-1 **Read mapping statistics.** Total reads and mapped reads for P0-initial (P0 aligned to VSV reference genome), P0 (P0 aligned to the consensus of P0-initial), P1, P2, P3 and P3b are shown.

Passage	P0-Initial	P0	P1	P2	P3	P3b
Total Reads	913387	913387	1044166	1624786	1706897	1726404
Mapped Reads	508249	513612	106302	713200	35581	30877
% Mapped	55.64%	56.23%	10.18%	43.90%	2.08%	1.79%

Table 4.5-2 **Fixed mutations in P0 as compared to the original ancestor VSV N1 strain (Wertz et al. 1998). Noise comparison as calculated from replicates or coverage noise.** Using technical replicates of P3 we calculate the fractional distribution of the population, averaged and calculated the standard deviation (n=2). We also used the average and standard deviation of the normalized coverages for each plateau to calculate fractions and propagation of uncertainty to estimate standard deviation of the fractions. The second method was used for P1 and P2 because we did not have technical replicates for those samples.

Location from 3'end	Wertz N1 Base	P0 Consensus Base
1626	G	A
1725	A	G
1902	G	A
1903	G	A
2207	C	G
2350	T	A
2580	C	T
2804	G	A

Table 4.5-3 **Noise comparison as calculated from replicates or coverage noise.** Using technical replicates of P3 we calculate the fractional distribution of the population, averaged, and calculated the standard deviation (n=2). We also used the average and standard deviation of the normalized coverages for each plateau to calculate fractions and propagation of uncertainty to estimate standard deviated of the fractions. The second method was used for P1 and P2 because we did not have technical replicates for those samples.

Species	Method			
	Technical Replicates		Region Noise	
	Percent	Std Dev	Percent	Std Dev
Long	1.20%	0.01%	1.21%	1.21%
Medium	34.95%	1.85%	36.26%	7.17%
Short	63.85%	1.85%	62.53%	2.60%

Table 4.5-4**Infectious and interfering unit concentrations by passage**

Sample	PFU/ml	Interfering Units/ml
P0	6.78E+09	0
P1	2.26E+09	3.93E+08
P2	4.64E+08	9.85E+09
P3	1.12E+06	7.78E+08

4.6 Methods

4.6.1 Cell culture

Baby hamster kidney (BHK-21) cells were maintained in 10% fetal bovine serum (FBS, Atlanta Biologicals), 1% Glutamax I (Gibco) in in Eagle's minimum essential medium (MEM; CellGro). Cells were passaged every 2 or 3 days when they reached confluency. For virus passaging infections, cells were plated in a T-75 flask, grown for 1 day and then infected.

4.6.2 Preparation of viral passages at fixed MOI

Viral passages were prepared by serial infection of BHK-21 cells at fixed MOI (20). For virus passaging, we used the N1 virus arrangement (47) which has all genes from the San Juan isolate of the Indiana VSV strain except the G protein which originated from the Orsay isolate of the Indiana strain. The sequencing of our strain revealed 8 fixed mutations as shown in Table S2, likely accumulated during passaging of the original strain. Viral infections were performed in 2%FBS (50 mL, Atlanta Biologicals), 1% GlutaMAX (5mL, LifeTechnologies/Invitrogen) in MEM (500mL, Gibco). BHK-21 cells were grown to 90% confluency and infected at MOI 10. Virus was adsorbed for one hour at 37°C. Cells were rinsed with Dulbecco's Phosphate-Buffered Saline (DPBS, Gibco) and 10mL media was added. The viral infection was allowed to progress for 24 hours and the supernatant containing virus particles was centrifuged (300 g) to remove cell debris. Samples were stored at -80°C until analysis. The sample was measured using a plaque assay to determine viral yield, and the passaging was continued until maintenance of MOI of 10 was not possible. The initial population was named P0. After the first MOI 10 passage the population was named P1, then P2, then P3. The virus titer after each passage was measured and diluted appropriate to sustain a MOI of 10.

4.6.3 Plaque Assay

Virus was quantified by plaque assay. The day prior to infection, cells were removed from T-75 flasks by treating with Trypsin EDTA (CellGro). The cells were diluted in 10% FBS media to roughly 10^5 cells/ml. Two ml of the cell solution were pipetted into each well of six-well tissue culture treated polystyrene plates (Corning). The virus samples were thawed at room temperature and serially diluted ten-fold in MEM. The media above the cell monolayer was removed, rinsed with DPBS and the cells were infected with 200 µl of the virus suspension and incubated for one hour. At the end of the one hour, the inoculum was removed and the cells were overlaid with 2 ml of 0.6%(w/v) agar (Difco) diluted in 2% FBS infection medium. After 20 hours of incubation, the agar overlay was removed and cells were fixed in 4%(w/v) paraformaldahyde (PFA, MP Biomedicals) and 5%(w/v) sucrose (Fisher Scientific) solution. The fixative remained on the cells for 20 minutes and then the cells were rinsed twice with DPBS and then stained with 2.5% (v/v) crystal violet (CV, PML Microbiological) diluted in 20 %ethanol (Fisher, Certified ACS grade) to aid in visualizing plaques. After the CV dried plaques were counted and virus titers were calculated as PFU/ml.

4.6.4 Interference Assay

DIP-mediated interference of infection in samples was quantified by the yield-reduction assay of Bellett & Cooper (19). Prior to all infections the cells were counted, enabling control of the MOI and quantification of virus and DIP yields on a per-cell basis. Serial dilutions of samples (1:10 or 1:2) were performed and then used to initiate infections. After 1 hour adsorption, stock virus was added at a MOI of 20, allowed to adsorb for 1 h, rinsed with DPBS, and replaced with MEM containing 2% FBS. Infections were incubated for 24 h, and stored at -80° C until quantification

by plaque assay. The plaque assay results were used to calculate the interfering unit concentration as described by Bellett and Cooper (19).

4.6.5 RNA Extraction

Viral RNA was collected using the QIAGEN Viral RNA extraction kit according to manufacturer's instruction, except carrier RNA was not added to the lysis buffer. Carrier RNA would interfere with downstream sequencing. The VSV associated RNAs were extracted from a full lineage including two technical replicates on supernatant from P3 (P0, P1, P2 P3 and P3b).

4.6.6 RNA Sequencing

Viral RNA was sequenced by the UW sequencing center using Illumina sequencing. RNA preparation was performed by the University of Wisconsin Biotechnology Center Genome Center. They provided us with the protocol they used for processing: RNA was fragmented using multi-indexing kit, first and second strand cDNA synthesis and end repair, followed by ligation of PE Index adapter, cDNA purification on E-gel Sie select for ~225bp and 300bp, cDNA enrichment via 18 cycles of PCR using multiplex primers, PCR product purification using ZymoResearch DNA Clean & Concentrator columns and quantified at ~1-9 ng/uL then submitted to UW Sequencing facility. The Illumina Sequencing technology provided $1x10^6$ to $1x10^7$ reads per sample with a read length of 35 nucleotides.

4.6.7 Sequence alignment and P0 consensus determination

Sequence alignment, and *de novo* sequence identification was performed in CLCBio Genomics Workbench using default settings. To generate a P0 consensus sequence the reads from P0 were initially aligned to a published VSV sequence (NCBI Reference Sequence: NC00160.1). We

extracted the consensus of P0 and used it as a reference to determine coverage in P0, P1, P2, P3 and P3b. The second alignment had an increased number of reads compared to the first.

4.6.8 Quantification and analysis of coverage data

The sequence coverage was extracted from CLCBio and analysis performed in MATLAB. The coverage in each passage was normalized by dividing the coverage at each base by its corresponding coverage in P0. Following normalization, locations of plateau boundaries were estimated by defining local mean coverage, systematically applying these measures across the genome, and identifying where significant changes occurred. More specifically, 400-nucleotide long windows, centered at every 5 nucleotides along genome, were expanded toward the 3' or 5' end of the genome by 10 nucleotides. Mean coverages for each window before and after expansion were compared by t-test. When the means were found to be significantly different (alpha = 10^{-11}) the added bases were recorded as potential plateau boundaries. The frequency of bases recorded as potential boundaries was plotted as a histogram with a bin size of 100 nucleotides (Figure 4.5-1). The bases identified as boundary more than 100 times were chosen as plateau boundaries. After plateau locations were determined, we calculated the average and standard deviation over the entire range of each plateau, and used those values to determine population fractions (Figure 4.5-2). Confidence intervals for population fractions were calculated using propagation of uncertainty and the equations in Figure 4.5-2.

4.6.9 qRT-PCR

Viral RNA was extracted from cell culture supernatant using the QIAAmp Viral RNA Mini Kit (QIAGEN) according to manufacturer's instructions. Reverse transcription was performed with VSV genome specific primer (see below) using the GoScript Reverse Transcriptase (Promega, Madison, USA) according to manufacturer's instructions. qPCR was performed with a forward

primer and reverse primer (see below) from RT reaction using the SsoFast Supermix (BioRad, Hercules, CA, USA) on a C1000 thermal cycler (BIORAD) according to manufacturer's instructions. A purified plasmid containing the VSV genome was used as a known standard to measure copies per mL. Primers were optimized to detect viral RNA by dilutions of extracted viral genomes. The primer sequences for full length genomes spanned the NP gene junction with sequences cagtcatgtcactgcaaggcctaa and taccgcctgatccagacgagaata. The medium length genomes were detected with primers tgctgtctctgaggtggaattgct and aagctgggaaccctaatcctgcat. The short genomes were detected in the L region of the genome with cagccttgggaaagggaaagcaaa and aacctatgaagggcagaccctgtt. All primers were synthesized by Integrated DNA Technologies (Coralville, IA, USA) and were re-suspended in RNAse free water.

4.6.10 Particle Quantification by Transmission Electron Microscopy

The virus samples were mixed with a NIST Traceable size standard (115 nm in diameter, 10^9 particles/ml) in 1:1 ratio. The prepared mixture was diluted 1:1 with methylamine tungstate (Nanoprobes) and a drop was loaded on a pioloform coated grid. The sample was then imaged using a transmission electron microscope (CM120, Philips). For each sample 15 TEM images were acquired, and VSV particles and microspheres were counted and averaged. The VSV particle concentration was estimated relative to the concentration of NIST Traceable size standard beads.

4.7 Acknowledgements

We thank Dr. Ming Yuan (University of Wisconsin – Madison) for helpful discussions. We are grateful for support from the National Institutes of Health (AI071197). The authors thank the University of Wisconsin DNA Sequencing Facility (UWBC) for their help in preparing and analyzing sequencing samples. We also thank the University of Wisconsin Medical School

Electron Microscope Facility. C. Timm acknowledges support from the National Library of Medicine training grant 5T15LM007359.

4.8 References

1. **Von Magnus P**. 1954. Incomplete forms of influenza virus. Advances in virus research **2**:59–79.

2. **Huang AS, Baltimore D**. 1970. Defective Viral Particles and Viral Disease Processes. Nature **226**:325–327.

3. **Lazzarini RA, Keene JD, Schubert M**. 1981. The origins of defective interfering particles of the negative-strand RNA viruses. Cell **26**:145–154.

4. **Roux L, Simon AE, Holland JJ**. 1991. Effects of defective interfering viruses on virus replication and pathogenesis in vitro and in vivo. Advances in Virus Research **40**:181–211.

5. **Bean WJ, Kawaoka Y, Wood JM, Pearson JE, Webster RG**. 1985. Characterization of virulent and avirulent A/chicken/Pennsylvania/83 influenza A viruses: potential role of defective interfering RNAs in nature. Journal of Virology **54** :151–160.

6. **Pesko KN, Fitzpatrick KA, Ryan EM, Shi P-Y, Zhang B, Lennon NJ, Newman RM, Henn MR, Ebel GD**. 2012. Internally deleted WNV genomes isolated from exotic birds in New Mexico: function in cells, mosquitoes, and mice. Virology **427**:10–7.

7. **Li D, Lott WB, Lowry K, Jones A, Thu HM, Aaskov J**. 2011. Defective interfering viral particles in acute dengue infections. PloS One **6**:e19447.

8. **Saira K, Lin X, DePasse J V, Halpin R, Twaddle A, Stockwell T, Angus B, Cozzi-Lepri A, Delfino M, Dugan V, Dwyer DE, Freiberg M, Horban A, Losso M, Lynfield R, Wentworth DN, Holmes EC, Davey R, Wentworth DE, Ghedin E**. 2013. Sequence analysis of in vivo defective-interfering (DI)-like RNA of influenza A H1N1 pandemic virus. Journal of Virology .

9. **Epstein D a, Herman RC, Chien I, Lazzarini R a**. 1980. Defective interfering particle generated by internal deletion of the vesicular stomatitis virus genome. Journal of Virology **33**:818–29.

10. **Re GG, Gupta KC, Kingsbury DW**. 1983. Genomic and copy-back 3’ termini in Sendai virus defective interfering RNA species. Journal of Virology **45**:659–64.

11. **Davis AR, Hiti AL, Nayak DP**. 1980. Influenza defective interfering viral RNA is formed by internal deletion of genomic RNA. Biochemistry **77**:215–219.

12. **Meier E, Harmison GG, Keene JD, Schubert M**. 1984. Sites of copy choice replication involved in generation of vesicular stomatitis virus defective-interfering particle RNAs. Journal of Virology **51**:515–21.

13. **O'Hara PJ, Horodyski FM, Nichol ST, Holland JJ**. 1984. Vesicular stomatitis virus mutants resistant to defective-interfering particles accumulate stable 5'-terminal and fewer 3'-terminal mutations in a stepwise manner. Journal of Virology **49**:793–798.

14. **Huang AS, Greenawalt JW, Wagner RoR**. 1966. Defective T Particles of Vesicular Stomatitis Virus I. Preparation, Morphology, and Some Biologic Properties. Virology **30**:161–172.

15. **Reichmann ME, Pringle CR**. 1971. Defective particles in BHK cells infected with temperature-sensitive mutants of vesicular stomatitis virus. Microbiology.

16. **Hartford SL**. 1975. Physical properties of New Jersey serotype of vesicular stomatitis virus and its defective particles. Proceedings of the National Academy of Sciences **72**:1202–1205.

17. **Nonoyama M, Watanabe Y, Graham AF**. 1970. Defective virons of reovirus. Journal of Virology **6**:226–236.

18. **Petric M**. 1970. Vesicular stomatitis virus- A new interfering particle, intracellular structures and virus-specific RNA. Virology **41**:615–630.

19. **Bellett AJD, Cooper PD**. 1959. Some Properties of the Transmissible Interfering Component of Vesicular Stomatitis Virus Preparations. Microbiology **21**:498–509.

20. **Stauffer Thompson KA, Rempala GA, Yin J**. 2009. Multiple-hit inhibition of infection by defective interfering particles. Journal of General Virology **90** :888–899.

21. **Janda JM, Davis AR, Nayak DP, De BK**. 1979. Diversity and generation of defective interfering influenza virus particles. Virology **95**:48–58.

22. **Marcus PI, Sekellick MJ**. 1974. Cell killing by viruses. I. Comparison of cell-killing, plaque-forming, and defective-interfering particles of vesicular stomatitis virus. Virology **57**:321–38.

23. **Thompson K, Yin J**. 2010. Population dynamics of an RNA virus and its defective interfering particles in passage cultures. Virology Journal **7**:257.

24. **Kirkwood TB, Bangham CR**. 1994. Cycles, chaos, and evolution in virus cultures: a model of defective interfering particles. Proceedings of the National Academy of Sciences **91** :8685–8689.

25. **Zwart M, Pijlman G, Sardanyés J, Duarte J, Januário C, Elena S**. 2013. Complex dynamics of defective interfering baculoviruses during serial passage in insect cells. Journal of Biological Physics **39**:327–342.

26. **Ke R, Aaskov J, Holmes EC, Lloyd-Smith JO**. 2013. Phylodynamic Analysis of the Emergence and Epidemiological Impact of Transmissible Defective Dengue Viruses. PLoS Pathogens **9**:e1003193.

27. **Tsibris AMN, Korber B, Arnaout R, Russ C, Lo C-C, Leitner T, Gaschen B, Theiler J, Paredes R, Su Z, Hughes MD, Gulick RM, Greaves W, Coakley E, Flexner C, Nusbaum C, Kuritzkes DR**. 2009. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. PloS One **4**:e5683.

28. **Bimber BN, Dudley DM, Lauck M, Becker E a, Chin EN, Lank SM, Grunenwald HL, Caruccio NC, Maffitt M, Wilson N a, Reed JS, Sosman JM, Tarosso LF, Sanabani S, Kallas EG, Hughes AL, O'Connor DH**. 2010. Whole-genome characterization of human and simian immunodeficiency virus intrahost diversity by ultradeep pyrosequencing. Journal of Virology **84**:12087–92.

29. **Nasu A, Marusawa H, Ueda Y, Nishijima N, Takahashi K, Osaki Y, Yamashita Y, Inokuma T, Tamada T, Fujiwara T, Sato F, Shimizu K, Chiba T**. 2011. Genetic Heterogeneity of Hepatitis C Virus in Association with Antiviral Therapy Determined by Ultra-Deep Sequencing. PLoS One **6**:e24907.

30. **Verbinnen T, Van Marck H, Vandenbroucke I, Vijgen L, Claes M, Lin T-I, Simmen K, Neyts J, Fanning G, Lenz O**. 2010. Tracking the evolution of multiple in vitro hepatitis C virus replicon variants under protease inhibitor selection pressure by 454 deep sequencing. Journal of Virology **84**:11124–33.

31. **Kuroda M, Katano H, Nakajima N, Tobiume M, Ainai A, Sekizuka T, Hasegawa H, Tashiro M, Sasaki Y, Arakawa Y, Hata S, Watanabe M, Sata T**. 2010. Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. PloS One **5**:e10256.

32. **Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G, Kim E, Pillai DR, Guyard C, Mazzulli T, Isa P, Arias CF, Hackett J, Schochetman G, Miller S, Tang P, Chiu CY**. 2010. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. PloS One **5**:e13381.

33. **Görzer I, Guelly C, Trajanoski S, Puchhammer-Stöckl E**. 2010. Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. Journal of Virology **84**:7195–203.

34. **Fischer W, Ganusov V V, Giorgi EE, Hraber PT, Keele BF, Leitner T, Han CS, Gleasner CD, Green L, Lo C-C, Nag A, Wallstrom TC, Wang S, McMichael AJ, Haynes BF, Hahn BH, Perelson AS, Borrow P, Shaw GM, Bhattacharya T, Korber BT**. 2010. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. PloS One **5**:e12303.

35. **Wright CF, Morelli MJ, Thébaud G, Knowles NJ, Herzyk P, Paton DJ, Haydon DT, King DP**. 2011. Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. Journal of Virology **85**:2266–75.

36. **Komarova A V, Combredet C, Sismeiro O, Dillies M-A, Jagla B, Sanchez David RY, Vabret N, Coppee J-Y, Vidalain P-O, Tangy F**. 2013. Identification of RNA partners of viral proteins in infected cells. RNA Biology **10**:944–957.

37. **Killip MJ, Young DF, Gatherer D, Ross CS, Short JAL, Davison AJ, Goodbourn S, Randall RE**. 2013. Deep Sequencing Analysis of Defective Genomes of Parainfluenza Virus 5 and Their Role in Interferon Induction. Journal of Virology **87** :4798–4807.

38. **Baum A, Sachidanandam R, García-Sastre A**. 2010. Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. Proceedings of the National Academy of Sciences of the United States of America **107**:16303–8.

39. **Holland JJ**. 1987. Defective interfering rhabdoviruses, p. 297–360. *In* Wagner, RR (ed.), The Rhabdoviruses.

40. **Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B**. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods **5**:1–8.

41. **Bloom JS, Khan Z, Kruglyak L, Singh M, Caudy AA**. 2009. Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. BMC Genomics **10**:221.

42. **Willenbrock H, Salomon J, Søkilde R, Barken KB, Hansen TN, Nielsen FC, Møller S, Litman T**. 2009. Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing. RNA (New York, N.Y.) **15**:2028–34.

43. **Camarena L, Bruno V, Euskirchen G, Poggio S, Snyder M**. 2010. Molecular mechanisms of ethanol-induced pathogenesis revealed by RNA-sequencing. PLoS Pathogens **6**:e1000834.

44. **Oliver HF, Orsi RH, Ponnala L, Keich U, Wang W, Sun Q, Cartinhour SW, Filiatrault MJ, Wiedmann M, Boor KJ**. 2009. Deep RNA sequencing of L. monocytogenes reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs. BMC Genomics **10**:641.

45. **Cole CN, Smoler D, Wimmer E, Baltimore D**. 1971. Defective Interfering Particles of Poliovirus I. Isolation and Physical Properties. J. Virol. **7**:478–485.

46. **Kuge S, Saito I, Nomoto A**. 1986. Primary structure of poliovirus defective-interfering particle genomes and possible generation mechanisms of the particles. Journal of Molecular Biology **192**:473–457.

47. **Wertz GW, Perepelitsa VP, Ball L a**. 1998. Gene rearrangement attenuates expression and lethality of a nonsegmented negative strand RNA virus. Proceedings of the National Academy of Sciences of the United States of America **95**:3501–6.

Chapter 5: Summary and Conclusions

Viruses are important human pathogens. Because viruses are so diverse, it is difficult to design broad spectrum anti-viral drugs. Much of the current research on viruses is focused on determining specific mechanisms that may be potential targets for specific anti-viral treatments. To facilitate testing the potential impact of these drugs, quantitative mathematical models of virus infections can be used to complement experimental methods. The models will help identify sensitive aspects and thus potential targets in the viral infection network. In this work I developed a method to measure absolute numbers of mRNA and genomes in an infection and provide multiple data sets for parameter estimation. Using a combination of quantitative data and modeling approaches I showed a quantitative relationship between viral genomes and mRNA that suggests the mRNA production rate is independent of viral polymerase levels during infection.

5.1 Quantification of VSV processes reveals new biology

In this work I presented quantitative measurements of VSV mRNA and genomes during an infection. The quantitative measurements revealed that the rate of transcription was linearly related to genomes: independent of new protein synthesis, genome replication, cell type, or cell immune response. These observations required quantitative measurements of viral mRNA and genomes during an infection. These observations validate and complement knowledge of multiple VSV mechanisms.

5.1.1 VSV transcription is captured with a single reaction

While transcription is a complex process, I was able to model the process using only a single reaction. A very simple reaction

$$genome \rightarrow mRNA + genome$$

using mass-action kinetics was sufficient to describe VSV mRNA production for all tested cases. This highlights an important and conserved relationship between the number of genomes in the cell and the production rate of mRNA.

5.1.2 The rate constant for VSV transcription is similar between many conditions

The transcription rate constant for the above reaction was similar for all tested cases.

1) The rate constant was similar in the presence or absence of genome replication (Chapter 2, Figure 1). Even though the same polymerase is used for replication, we see a relatively constant level of mRNA production per genome early in the infection. This suggests that the transcription promoter is much stronger than replication, or that replication is initiating much slower than transcription during VSV replication.
2) The mRNA production rate from genomes was constant in permissive (BHK) and resistant (PC3) cells (Chapter 2, Figure 3). This suggests that host environment is not a major factor in the production rate of viral mRNA. However, it should be noted that the study only considers two cell types.
3) The mRNA production rate is constant in innate stimulated and non-stimulated resistant cells (PC3). While genome replication was greatly reduced, the production of mRNA was relatively constant. This suggests that the innate response does not have specific targets for preventing VSV transcription, rather that the inhibition of viral activity occurs in translation or genome replication.

While all of the mechanistic details are important for producing mRNA, the data and analysis presented in Chapter 2 suggests that the transcription rate is limited by a single mechanism. If there were multiple limitations we would see differential transcription rates and perhaps a different genome-mRNA relationship under different environmental conditions for the virus. It is likely that the limitation for VSV transcription is the processivity rate of the polymerase, but further testing is required to determine the exact limiting step in VSV transcription (see Chapter 5, Future Work).

5.1.3 A simple model captured VSV kinetic data

The data in Chapter 2 was fit using a simple model with only 6 parameters, for an infection that requires hundreds of unique mechanisms. The simplicity of the model relative to the details of the infection helps reveal conserved viral relationships and may represent simple strategies the virus uses to reprogram the host cell. In addition to allowing high level comparisons of viral intermediates, such simple models may be useful for high throughput modeling that requires some knowledge of the viral infection. Simple models for a single cell are computationally easy to solve given an initial condition, while a more complex model integrating all aspects of infection are computationally intensive. When only a summary is needed, these models may be used. Simple models are potentially useful in modeling the spread of viral infections between neighboring cells, tissues, or organism.

5.2 Simulation of VSV using quantitative data generates new hypotheses

In Chapter 3 a mechanistic model of VSV growth that incorporates many of the known mechanisms of the infection was presented. The model was solved using delayed ODEs to simulate interactions as well as delays inherent to the multiple polymerization processes of

transcription, translation and replication. In order to have a flexible model to complement wet-lab experimentation we developed a simulation environment which allows for very easy manipulation of the VSV genome and easy modification/addition/removal of reactions to the infection cycle. The described flexibility is important because as new mechanisms are discovered it is desirable to have an easy-to-modify model, as opposed to the alternative of building a new model from scratch.

5.2.1 Model predictions

The VSV model was parameterized to match quantitative RNA data and known VSV phenotypes. In the development of the model we used parameters from literature, known VSV mechanisms and mRNA and genome measurements similar to those collected in Chapter 2. The model predicts two interesting protein usages

1) VSV N protein is predicted to be used almost 100% efficiently in the replication of viral genomes.
2) After N protein, VSV P protein is the most efficiently packaged of the viral proteins. P protein is heavily packaged in the viral particle, but presumably produced at lower rate than other viral proteins.

Further experiments are required to test these quantitative hypotheses. Ongoing efforts to couple genome and protein measurements are described in Chapter 6, Future Work.

5.2.2 Model flexibility

We displayed the flexibility in the model by simulating gene insertions and multiple initial conditions. For multiple initial conditions, we observe model predictions similar to the results shown in Chapter 2, most notably genome saturation at about 2000 genomes per cell. We also demonstrated the genome flexibility by inserting a fluorescent protein into the viral genome. We

were able to reproduce trends seen in data of similar genome insertion mutants. It is important to note that inserting the gene is as easy as typing in a new genome. The model does the accounting necessary to calculate genome length, transcription attenuation, and ribosomal usage of the new protein.

5.3 Usable and modifiable virus models

Chapter 3 was devoted to developing a mechanistic model of VSV infection that is easy to modify. We used two approaches to make it modifiable: we generated a generic reaction which is used to modify the reaction network and we developed a simulation environment for complex, biological reactants. Our goal was to build a model that future researchers could easily modify as new mechanisms are discovered. The real challenge in developing modifiable models is making them flexible enough for new users to learn and modify quickly. The main advantage is that the model is a good approximation of the system that does not need to be built from scratch with each iteration or new discovery.

5.4 Quantification of VSV population changes

In Chapter 4 VSV populations were examined using the unbiased technique of RNA sequencing. Using coverage data and complementary measures it was shown that RNA sequencing can be used to detect, identify and quantify population dynamics of virus growth. This work introduces RNA sequencing as a quantitative tool that can be used to study viruses and genome deletion mutants in a population. While the technology is currently too expensive for detailed kinetics studies, the cost of sequencing is always dropping and it will prove to be a very useful high throughput tool for virus host interaction expression kinetics. As high throughput sequencing

kinetic studies develop models that can unite the dense expression data will be pivotal in interpreting how the gene expression of the virus interacts with the gene expression of the host.

5.5 Quantitative biology as a complement to molecular mechanisms

The overall theme of this work has been to quantify specific aspects of vesicular stomatitis virus infection. This work helped support previously identified mechanisms as well as prompted questions about other aspects of the infection. While much of the work was in the laboratory setting it is also feasible to apply quantitative biology in more complex systems. For example, in Chapter 2 we used a very simple model to explain the complex system of interactions that describe the VSV replication cycle. We identified a consistent relationship between the production rate of mRNA and genomes. The conclusions did not require knowledge of the detailed mechanisms of the system; rather it showed a high level overview of the kinetics and conserved quantitative relationships between two measured intermediates. Similar methods may be applied to unknown viral systems, either in laboratory, clinical, or natural settings. Do we observe any conserved quantitative trends in such systems? What does that mean for potential mechanisms? With quantitative biology and observation we can use conserved relationships to generate new hypothesis and predict the trajectories of systems.

Chapter 6: Future work

6.1 Introduction

In this section I describe in detail some of the implications of the results of previous chapters. I introduce questions and specific hypotheses and experimental methods that may help address the issues. The chapter is divided into two sections: Experimental work and Virus infection modeling work.

6.2 Experimental work

6.2.1 Longevity of linear transcription suggests active recycling of polymerases

In Chapter 2 Figure 1A and 2A I quantified the transcription rate in BHK or PC3 cells in the presence of cycloheximide. We observed a linear growth of mRNA that lasted for 6 or 8 hours (BHK or PC3, respectively). Based on the measured rate of polymerization of the VSV RdRP (3.7 nt/s (1)) and the genome length (11,161 nt), it should take about 50 minutes for a single polymerase to traverse the genome. We can also calculate a spacing time to describe how often a particular polymerase would cross a location on the genome using the rate and 200 nt spacing estimate (2). If the polymerases only make a single pass, we expect to see the rate of mRNA production decrease after all 50 polymerases have been loaded and traverse the genome: approximately 2 hours. As shown in Chapter 2, transcription proceeds at a constant rate for much longer suggesting that polymerases are actively recycled to the 3' transcriptional promoter of the VSV genome.

In Chapter 3 we model this 'active recycling' by applying a very high rate constant for the binding of the RdRp to the transcription and replication promoters, essentially forcing the genome to be fully loaded with polymerases when they are available.

A similar active recycling strategy is observed in translation. During translation, the 5' and 3' ends of the mRNA interact (mediated by translation initiation factors) forming a continuous loop to allow rapid reloading of ribosomes as they complete a round of translation (Figure 6.2-1) (3).

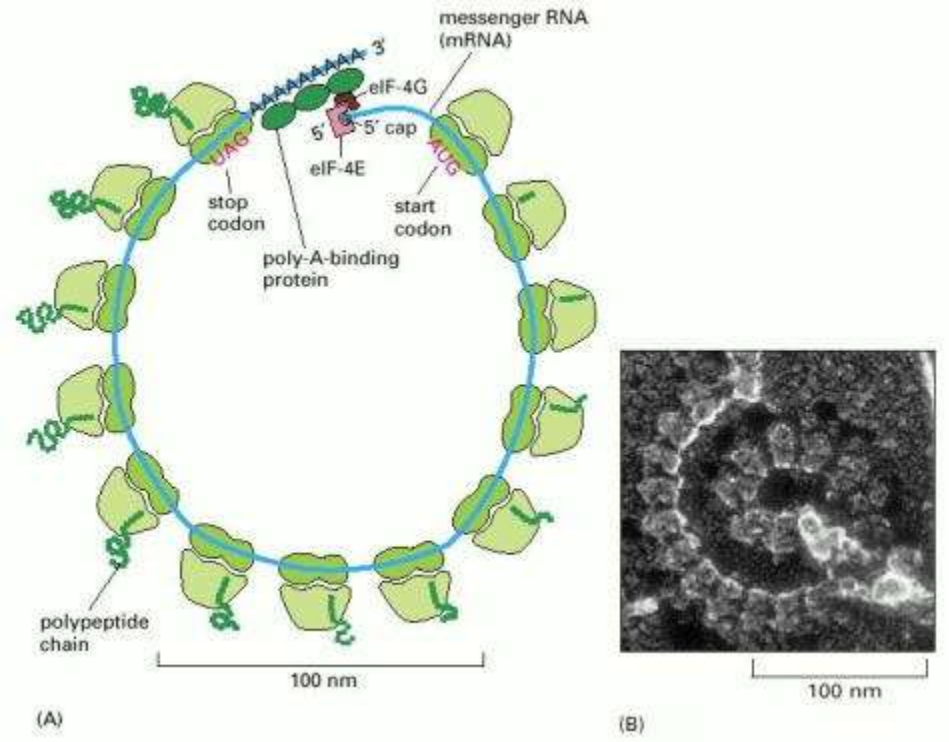


Figure 6.2-1 A polyribosome. (A) Schematic drawing showing how a series of ribosomes can simultaneously translate the same eucaryotic mRNA molecule. (B) Electron micrograph of a polyribosome from a eucaryotic cell. (B, courtesy of John Heuser.)
From: From RNA to Protein (3) [copied without permission]

We hypothesize that a similar process may be occurring with the RdRP in its interaction with the viral genome. The VSV genome ends are complementary (4) and may also interact during replication or transcription. However, there is a second mechanism that must be accounted for

during transcription which is intergenic attenuation. Between VSV genes, the polymerase has some probability of dissociating from the genome, leading to a transcriptional gradient. However, it is unclear what happens to the polymerases after dissociation. Using the same arguments as above (continued linear transcription) I hypothesize that these polymerases are also actively recycled to the 3' transcription promoter. This may also depend on proximity to the 3' promoter, suggesting that the VSV genome may take on a 'flower' shape to facilitate close proximity of the polymerase initiation site and exit sites (Figure 6.2-2). This structure may be influenced by RNA binding or N protein binding in the intracellular environment.

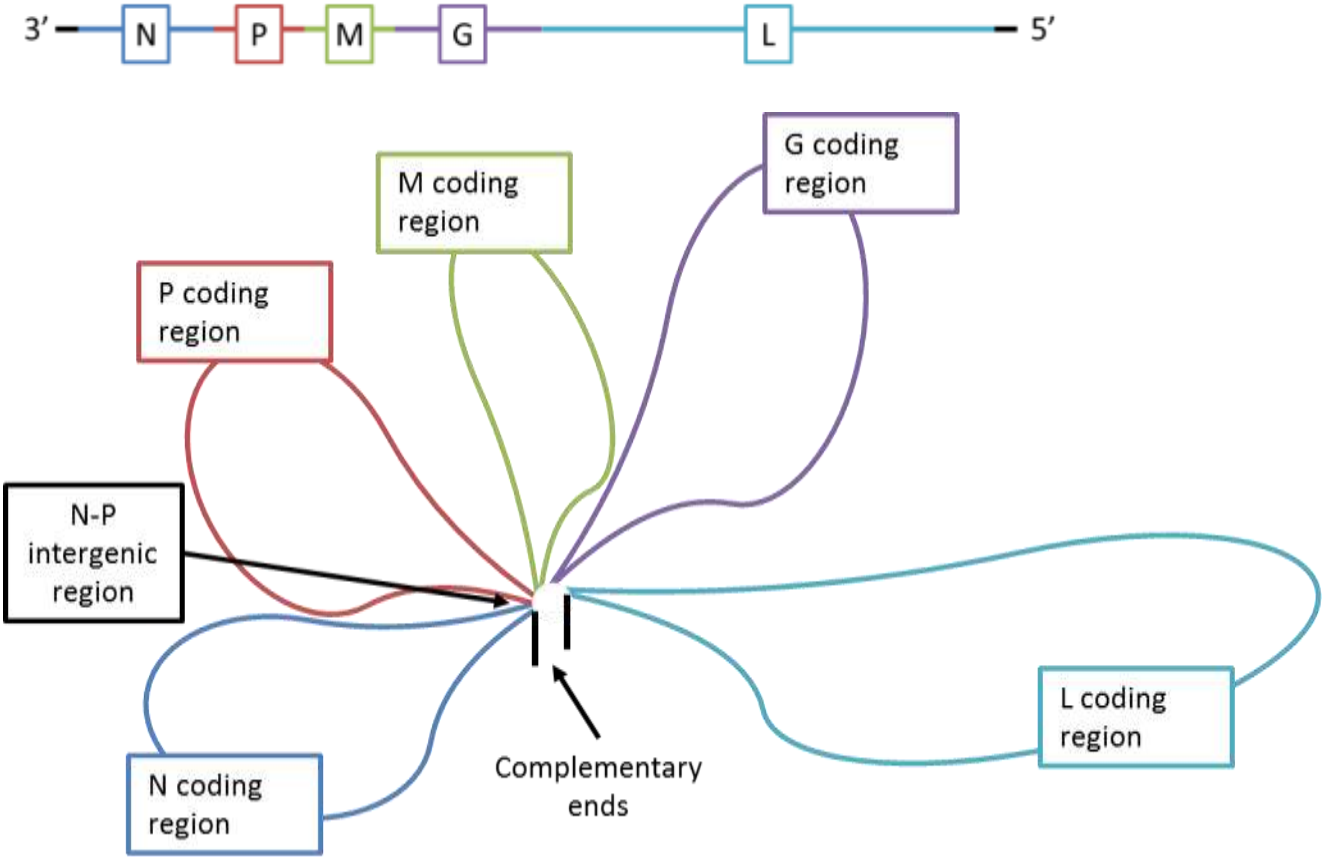


Figure 6.2-2 Potential structure of VSV genome during transcription. Top: unfolded genome. Bottom: This structure would allow for close proximity of polymerase 'exit sites' and the 3' transcriptional promoter, allowing for active recycling of polymerases.

However, this structure is quite speculative. Electron microscopy of fixed cells infected with VSV would allow us to visualize the genome structure during transcription or replication. I imagine a picture similar to the ribosome shown in Figure 1, except for VSV genomes and polymerases rather than mRNA and ribosomes.

Another approach to test the active recycling is to generate a VSV mutant that does not have a polymerase gene and monitor transcription in the presence of cycloheximide (no translation). The incoming particle would be produced such that it would carry active units of the RdRp, but be unable to produce more polymerase molecules (omitting L rules out "leaky translation" due to incomplete shutdown as a possibility for new polymerase). Then we can measure how long transcription is able to continue. If mRNA production continues linearly much longer than expected by the measured elongation rate, we know that polymerases must be efficiently transported to the transcriptional promoter.

6.2.2 VSV transcription is linear throughout replication, suggesting that the majority of genomes are used as templates for transcription rather than genome replication

In Chapter 2 we observed that the transcription rate does not change significantly in the presence of replication, there is no hard switch to a replication mode. This suggests that even during replication, there is a significant amount of transcription. In Chapter 3 we simulated this behavior by imposing a strong transcription promoter relative to replication. It is also possible that the VSV replication structures observed in (5) are both the source of replication and transcription. The authors observe intracellular structures that are dense with both viral proteins and viral RNA. The high density of viral intermediates suggests that these inclusion bodies are the major

site of genome replication (5). For example, the inner core of the cytoplasmic inclusions may be relatively poor in N protein, and transcripts are produced from genomes in the absence of N protein. If this is true, we might expect to see newly translated N protein accumulating at the outer edges of such replication structures. We may observe simultaneous transcription and replication within a single inclusion structure.

Another possible explanation is that the replication promoter truly is much weaker than the transcriptional promoter, independent of N protein presence. Thus, we should see replication and transcription from templates in the presence of high N protein. In vitro transcription and replication experiments would be useful in testing this hypothesis. In vitro transcription and replication for VSV have been described in the literature (1, 6) and could be adapted to address this specific hypothesis. For example, one could add dilutions of N protein to purified nucleocapsids in an *in vitro* transcription/replication buffer. Observing the growth of mRNA or genomes would help us learn about the replication promoter strength dependence on N protein.

An alternative explanation for linear transcription during replication is this: single cells do employ a 'hard switch' behavior, but because we measure the population average the effect is averaged out. However, if this were the case we would expect to see a decrease in transcription rate during replication. A decrease in transcription rate was not observed until after replication was nearly complete, suggesting a different mechanism for the decline in transcription. A likely mechanism to account for the stoppage of transcription is the inhibition of polymerase activity by viral M protein (7). Late transport processes or M protein accumulation late in the infection may be explanations for the late shutdown of polymerase activity. Quantitative protein measures would help determine the most likely mechanism, as well as visualization of live cell infections.

6.2.3 What limits VSV transcription?

In Chapter 2 the data showed that the rate constant for VSV mRNA production was similar for all conditions when modeled with a lumped model. Because the rate constant was similar, it suggests that in all these cases the viral transcription is limited by the same step. Transcription is a complex process, but determining the rate-limiting step may help identify a sensitive aspect of the VSV infection cycle that may be a potential drug target. VSV transcription requires multiple steps, any of which could be limiting the production rate of mRNA. The following table shows potential rate-limiting steps and ways to examine their mechanisms.

Potential Limiting Step	Variables	Tests	References
Polymerase interaction with genome	Sequence specific promoter Genome secondary structure Access to genome (nucleocapsid coated)	ChIP-Seq pull down assays Structure prediction	(8–12)
Initiation of transcription	Sequence specific promoter Complex formation (L and P) Access to genome		(1, 10, 13–15)
Production of mRNA through nucleotide polymerization	Processivity (speed of polymerization)	Polymerase mutations, nucleotide precursor starvation	(1, 16–18)
Poly-A tail addition by stuttering	Tail length Stop signals	Adenosine starvation Mutation studies	(16, 19–23)
Release of mRNA from the polymerase	Intergenic stop signal	Mutation studies	(20, 22)

While some of these methods focus on VSV, many of these mechanisms may also be important for other viruses. The approach of determining a rate-limiting step in transcriptional processes may be applied to other viruses that encode their own polymerase for transcription. Identifying the rate-limiting step can help in the design of new anti-viral drugs or treatments.

6.2.4 Quantification of VSV proteins by coupling qRT-PCR and mass spectrometry

One strategy to compare levels of proteins and RNA is to use a standard curve that relates the two. Because viral particles contain a single genome and a relatively constant number of the five

viral proteins, we chose to use virus adsorbed to cells to couple mass spectrometry and qRT-PCR. Further, this method allows us to examine cell-associated viral RNA and protein, as will be measured in later experiments. In Chapter 2 we observed that viral transcription begins early in infection, and we assumed that translation of viral mRNA would also begin early in infection. To prevent translation, we treated cells with cycloheximide prior to and during adsorption of particles. Sucrose gradient purified virus was diluted in a 1:5 dilution series and added to PC3 cells pretreated with cycloheximide. Parallel duplicates were collected for RNA measurements and protein measurements. Figure 6.2-3 shows the preliminary results of genome and protein measurements for this calibration.

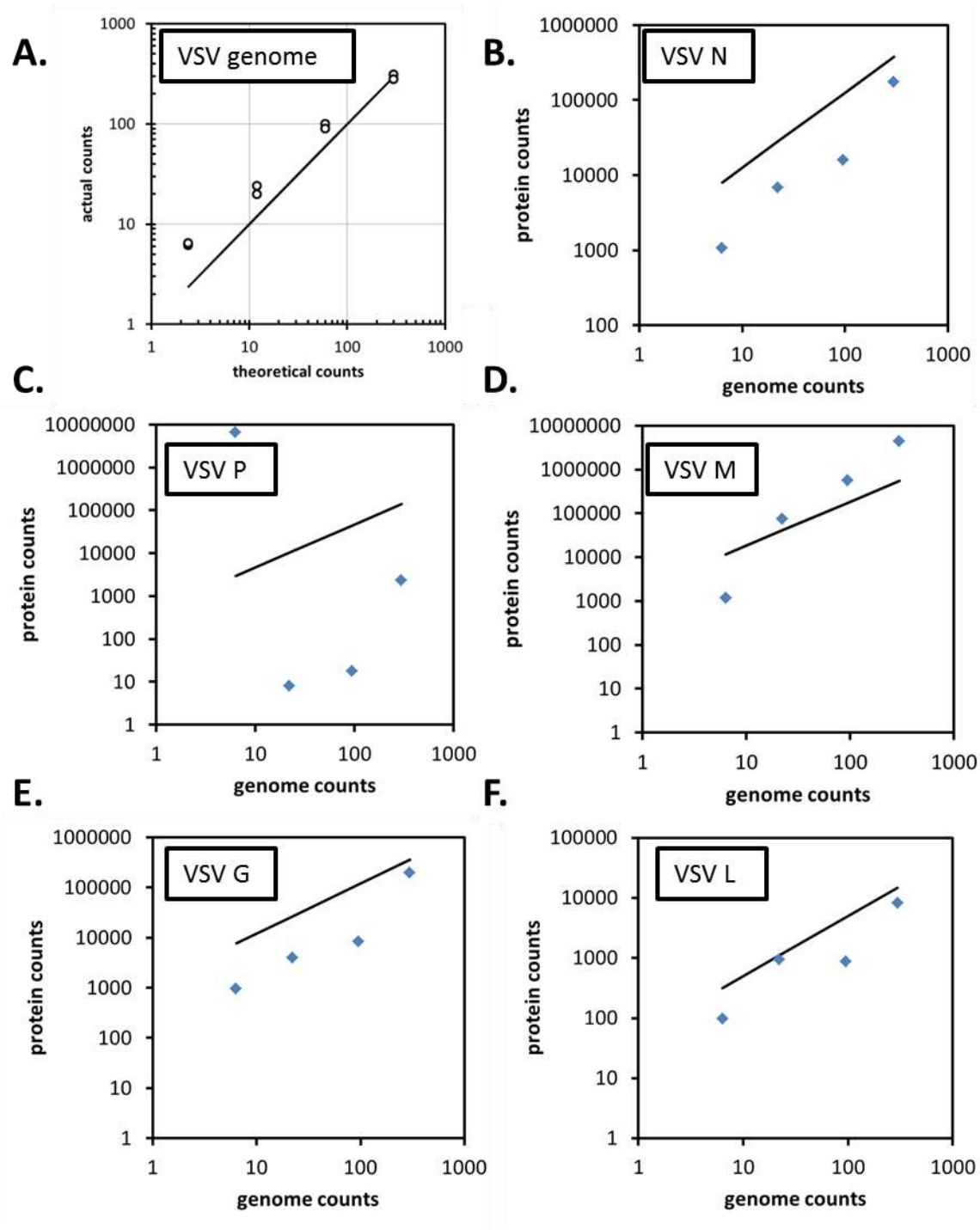


Figure 6.2-3 VSV particles were serially diluted 1:5 and adsorbed to PC3 cells. **A.** Genome measures. The points are the data, and the line is the theoretical 1:5 line from the most concentrated sample. **B.** VSV N protein counts plotted against VSV genomes at that dilution. **C.** VSV P protein vs. genome counts. **D.** VSV M **E.** VSV G **F.** VSV L

As shown in panel A, genome counts did not follow the expected 1:5 dilution. This is likely due to an error in the primary dilution of the samples, and this will be repeated. However, it is still worthwhile to compare proteins and genomes from this preliminary test. Panels B-F show the viral protein counts (y-axis) plotted against the genome counts for that sample. The line was determined by multiplying the genome counts by the number of proteins per particle as reported in Thomas et al. 1985 (24). Interestingly, the expected line is almost always higher than the measured counts, suggesting a disagreement with Thomas et al. It is possible that this disagreement is due to the very different methods used to quantify genomes and proteins. Nonetheless, these measures give us estimates of how many proteins we can expect to detect per genome in particles. This is summarized in the following Table.

	N per genome	P per genome	M per genome	G per genome	L per genome
this study	267	27	2731	197	20
Thomas et al	1258	466	1826	1205	50

As shown, the calibration constants shown here are significantly different than reported in literature. However, in this work we will use the calibration constants to compare protein levels during an infection to what is necessary to produce particles. We expect that the observed protein levels will be much higher than predicted by applying the calculations from calibration. This is because viral proteins are used for more than just building particles; they also have important functions in the cell. Also, it is unlikely that viral proteins are 100% efficiently used during replication, rather we expect to see an excess of proteins as compared to the number of particles we could produce.

We will use the same transcription model that was described in Chapter 2, that the mRNA production rate is linearly related to the number of genomes. The protein counts will be compared to mRNA, first assuming that translation is linearly related to the number of mRNA. If this model does not capture data, we will add more complexity as necessary. This analysis will provide absolute translation rates of protein relative to their respective mRNA. With this data we can draw conclusions about selective translation of VSV proteins.

In Chapter 3 our VSV simulation predicted that N protein was very rapidly consumed during an infection. In this section we examine the usage of proteins relative to what is observed in particles. Using absolute protein measures and the genome counts multiplied by the calibration constants we can compare the required protein for particles to the measured protein levels. We expect that N protein will be efficiently produced, while M protein will likely be overproduced. Because M protein is used for so many intracellular mechanisms, it is likely that it will not be as efficiently used relative to genomes as N protein.

Using the model from Chapter 2 (and potentially Chapter 3) we can examine the quantitative aspects of these measurements and how they change the model fitting or predictions. This will allow us to generate a more complete, quantitative view of the VSV replication cycle.

6.3 Virus infection modeling work

6.3.1 Evolution of virus reaction network and application to N1-N4 mutants

In Chapter 3 we introduced a VSV infection model based on known VSV mechanisms. A major biological improvement over other intracellular virus models is that this model also includes a mechanism for cell shutdown that depends on virus replication. Because of this interaction, the model is self-regulating: the simulated virus cannot replicate too quickly or too slowly because it

will not produce the essential proteins to produce new, infectious particles. Using this self-regulating structure we can potentially explore how the virus optimizes its production, which is related to its evolutionary dynamics.

In order to explore evolution one must define a fitness measure that is closest to the selection imposed on the system. If the virus is passaged in such a way as to give an advantage to high produces, then the fitness measure in the model should be total particles produced. If the virus is passaged to produce particles fastest, then the model selection coefficient should be related to how quickly the first virus can be produced from the cell. There are many options for parameters to optimize, each with its own hypotheses.

The model was parameterized to a system that is likely near its evolutionary optimum: VSV infection in cell culture. Because the model was parameterized to match intracellular data from this infection, we hypothesize that the model is already near its own evolutionary optimum. It would be interesting to explore how/if this simulation can be optimized, and likely would help identify which mathematical representations are farthest from the true behavior of the system. We already discussed that the 'immune response' reactions are far from from modeling the mechanistic response of the real system, so there may be much room for improvement in modeling the innate response.

Other non-wild-type VSV genome rearrangements are likely not at their evolutionary optimum. The rearrangement of VSV gene order changes the expression level of certain genes, and for a first pass we can assume that all protein interactions are conserved. Simulations are very good at exploring expression changes, and the model could be used to explore how such rearrangements would evolve over short time scales. Experimental evolution studies would complement and challenge the computational evolution predictions.

It is important to note that this model strategy is unable to explore major structural changes (such as novel interactions between proteins) but we can potentially predict changes in rate constants for an infection. For example, we may see a shift in the balance of transcription to replication dependent upon the relative promoter strengths. Such shifts have been observed in temperature sensitive mutants of VSV (25).

6.3.2 Localized models

The majority of intracellular virus models assume that the cell is well-mixed (2, 26–29). However, is well established that cells are not well-mixed; rather they are highly ordered structures with organelles that perform specific functions (3). While the virus may interact with and modify these structures, virus infections have also been shown to produce highly ordered and localized structures within the cell (30). VSV specifically has been shown to generate inclusion bodies that are believed to be the site of genome replication (5). While current models do a relatively good job at fitting data, localized models of infection should also be explored. A specific reason is the consideration of the VSV N protein. In Chapter 3 the well-mixed model predicts high consumption of N protein; however the real process should include the transportation of N protein to the inclusion body. With a localized model one could explore the effect of transport rate on the virus infection. Live-cell imaging on inclusion body growth would help to estimate transport rates of viral N protein. Further, the detailed mechanisms involved in transport processes may be potential targets for anti-viral drugs for more health relevant viruses.

6.3.3 Innate response models

The mechanistic VSV replication network involves many interactions; however we were able to capture much of the experimental data with a 3 reaction model using 6 parameters. The innate response network is similarly complex but may also have a structure that allows for a similar

lumped modeling approach. To identify the underlying structure it is important to use domain knowledge (31–34) to identify key nodes in the innate response. Then, assays can be developed to measure the species at or near these nodes.

Similar to the work presented in Chapters 2 and 3 it is likely that the system can be represented as only a few lumped equations, or the system can be modeled in more detail by including all known interactions. Both types of models will advance the understanding of the response.

Innate response models to VSV infection can be parameterized with multiple initial conditions of IFN treatments. Quantitative measurements such as viral output, viral gene expression, or host gene expression will likely be easiest to obtain to start developing models. qRT-PCR is relatively simple for gene expression, but RNA sequencing costs are dropping and also are an option for such studies.

Using the cells as a constant, it would be interesting to see how different viruses activate the response differently. In some cases we know the detection point (e.g. Toll-like receptors) but in others we may be able hypothesize which receptors are activated by the response of the network.

Finally, it is interesting to consider that the innate response may be very different between cell types or individuals. Multiple cell types should be analyzed in how they respond to a virus strain.

6.4 References

1. **Iverson LE, Rose JK**. 1981. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. Cell **23**:477–84.

2. **Lim K, Lang T, Lam V, Yin J**. 2006. Model-Based Design of Growth-Attenuated Viruses. PLoS Computational Biology **2**.

3. **Alberts B, A J, J L, Al E**. 2002. Molecular Biology of the Cell, 4th ed. New York: Garland Science.

4. **Lyles DS, Rupprecht CE**. 2007. Rhabdoviridae, p. 1363–1408. *In* Knipe, DM, Howley, PM (eds.), Fields Virology, 5th ed. Lippincott Williams & Wilkins, Philadelphia, PA.

5. **Heinrich BS, Cureton DK, Rahmeh A a, Whelan SPJ**. 2010. Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions. PLoS pathogens **6**:e1000958.

6. **Davis NL, Wertz GW**. 1982. Synthesis of Vesicular Stomatitis Virus Negative-Strand RNA In Vitro: Dependence on Viral Protein Synthesis. Journal of Virology **41** :821–832.

7. **Suryanarayana K, Baczko K, Ter Meulen V, Wagner RR**. 1994. Transcription inhibition and other properties of matrix proteins expressed by M genes cloned from measles viruses and diseased human brain tissue. Journal of Virology **68** :1532–1543.

8. **Galloway SE, Wertz GW**. 2009. A temperature sensitive VSV identifies L protein residues that affect transcription but not replication. Virology **388**:286–93.

9. **Green TJ, Macpherson S, Qiu S, Lebowitz J, Wertz GW, Luo M**. 2000. Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. Journal of virology **74**:9515–24.

10. **Green TJ, Rowse M, Tsao J, Kang J, Ge P, Zhou ZH, Luo M**. 2011. Access to RNA encapsidated in the nucleocapsid of vesicular stomatitis virus. Journal of virology **85**:2714–22.

11. **Moyer S a, Smallwood-Kentro S, Haddad A, Prevec L**. 1991. Assembly and transcription of synthetic vesicular stomatitis virus nucleocapsids. Journal of Virology **65**:2170–8.

12. **Hinzman EE, Barr JN, Wertz GW**. 2002. Identification of an Upstream Sequence Element Required for Vesicular Stomatitis Virus mRNA Transcription. Journal of Virology **76**:7632–7641.

13. **Rainsford EW, Harouaka D, Wertz GW**. 2010. Importance of hydrogen bond contacts between the N protein and RNA genome of vesicular stomatitis virus in encapsidation and RNA synthesis. Journal of virology **84**:1741–51.

14. **Ivanov I, Yabukarski F, Ruigrok RWH, Jamin M**. 2011. Structural insights into the rhabdovirus transcription/replication complex. Virus Research **162**:126–37.

15. **Whelan SP, Wertz GW**. 1999. Regulation of RNA synthesis by the genomic termini of vesicular stomatitis virus: identification of distinct sequences essential for transcription but not replication. Journal of Virology **73**:297–306.

16. **Stillman E a, Whitt M a**. 1997. Mutational analyses of the intergenic dinucleotide and the transcriptional start sequence of vesicular stomatitis virus (VSV) define sequences required for efficient termination and initiation of VSV transcripts. Journal of Virology **71**:2127–37.

17. **Stillman E a, Whitt M a**. 1998. The length and sequence composition of vesicular stomatitis virus intergenic regions affect mRNA levels and the site of transcript initiation. Journal of Virology **72**:5565–72.

18. **Wang JT, McElvain LE, Whelan SPJ**. 2007. Vesicular stomatitis virus mRNA capping machinery requires specific cis-acting signals in the RNA. Journal of Virology **81**:11499–506.

19. **Grdzelishvili VZ, Smallwood S, Tower D, Hall RL, Hunt DM, Moyer SA**. 2005. A Single Amino Acid Change in the L-Polymerase Protein of Vesicular Stomatitis Virus Completely Abolishes Viral mRNA Cap Methylation. Journal of Virology **79**:7327–7337.

20. **Rahmeh A a, Schenk AD, Danek EI, Kranzusch PJ, Liang B, Walz T, Whelan SPJ**. 2010. Molecular architecture of the vesicular stomatitis virus RNA polymerase. Proceedings of the National Academy of Sciences of the United States of America **107**:20075–80.

21. **Whelan SP, Barr JN, Wertz GW**. 2000. Identification of a minimal size requirement for termination of vesicular stomatitis virus mRNA: implications for the mechanism of transcription. Journal of Virology **74**:8268–76.

22. **Barr JN, Whelan SPJ, Wertz GW**. 2002. Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus. Biochimica et biophysica acta **1577**:337–53.

23. **Barr JN, Whelan SP, Wertz GW**. 1997. cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. Journal of Virology **71**:8718–25.

24. **Thomas D, Newcomb WW, Brown JC, Wall JS, Hainfeld JF, Trus BL, Steven AC**. 1985. Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. Journal of Virology **54** :598–607.

25. **Wertz GW, Levine M**. 1973. RNA Synthesis by Vesicular Stomatitis Virus and a Small Plaque Mutant: Effects of Cycloheximide. Journal of Virology **12** :253–264.

26. **Heidtke KR, Schulzećkremer S**. 1998. Design and implementatio of a qualitative simulation model of λ phage infection. Bioinformatics **14**:81–91.

27. **Dahari H, Layden–Almer JE, Kallwitz E, Ribeiro RM, Cotler SJ, Layden TJ, Perelson AS**. 2009. A Mathematical Model of Hepatitis C Virus Dynamics in Patients With High Baseline Viral Loads or Advanced Liver Disease. Gastroenterology **136**:1402–1409.

28. **Regoes RR, Crotty S, Antia R, Tanaka MM**. 2005. Optimal replication of poliovirus within cells. The American Naturalist **165**:364–73.

29. **Sidorenko Y, Reichl U**. 2004. Structured model of influenza virus replication in MDCK cells. Biotechnology and Bioengineering **88**:1–14.

30. **Den Boon J a, Diaz A, Ahlquist P**. 2010. Cytoplasmic viral replication complexes. Cell Host & Microbe **8**:77–85.

31. **Faul EJ, Lyles DS, Schnell MJ**. 2009. Interferon response and viral evasion by members of the family rhabdoviridae. Viruses **1**:832–51.

32. **Otto Haller, Peter Staeheli GK**. 2007. Interferon-induced Mx proteins in antiviral host defens. Biochimie **89**:812–818.

33. **Gao Y, Whitaker-Dowling P, Watkins SC, Griffin J a, Bergman I**. 2006. Rapid adaptation of a recombinant vesicular stomatitis virus to a targeted cell line. Journal of Virology **80**:8603–12.

34. **Chen S, Short JAL, Young DF, Killip MJ, Schneider M, Goodbourn S, Randall RE**. 2010. Heterocellular induction of interferon by negative-sense RNA viruses. Virology **407**:247–255.