

RESEARCH ARTICLE

Stochastic model of vesicular stomatitis virus replication reveals mutational effects on virion production

Connor R. King , Casey-Tyler Berezin, Jean Peccoud *

Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, Colorado, United States of America

* Jean.Peccoud@colostate.edu

Abstract

We present the first complete stochastic model of vesicular stomatitis virus (VSV) intracellular replication. Previous models developed to capture VSV's intracellular replication have either been ODE-based or have not represented the complete replicative cycle, limiting our ability to understand the impact of the stochastic nature of early cellular infections on virion production between cells and how these dynamics change in response to mutations. Our model accurately predicts changes in mean virion production in gene-shuffled VSV variants and can capture the distribution of the number of viruses produced. This model has allowed us to enhance our understanding of intercellular variability in virion production, which appears to be influenced by the duration of the early phase of infection, and variation between variants, arising from balancing the time the genome spends in the active state, the speed of incorporating new genomes into virions, and the production of viral components. Being a stochastic model, we can also assess other effects of mutations beyond just the mean number of virions produced, including the probability of aborted infections and the standard deviation of the number of virions produced. Our model provides a biologically interpretable framework for studying the stochastic nature of VSV replication, shedding light on the mechanisms underlying variation in virion production. In the future, this model could enable the design of more complex viral phenotypes when attenuating VSV, moving beyond solely considering the mean number of virions produced.

OPEN ACCESS

Citation: King CR, Berezin C-T, Peccoud J (2024) Stochastic model of vesicular stomatitis virus replication reveals mutational effects on virion production. *PLoS Comput Biol* 20(2): e1011373. <https://doi.org/10.1371/journal.pcbi.1011373>

Editor: Jessica Conway, Penn State, UNITED STATES

Received: July 22, 2023

Accepted: January 24, 2024

Published: February 7, 2024

Copyright: © 2024 King et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All code and files containing models can be found on GitHub https://github.com/Peccoud-Lab/VSV_Stochastic_Model_2023.

Funding: This work was supported by the National Science Foundation Award MCB-2123367 to JP and the National Institute of General Medical Sciences Award T32GM132057 to CK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author summary

This study presents the first complete stochastic model of vesicular stomatitis virus (VSV) replication. Our model captures the dynamic process of VSV's replication within host cells, accounting for the stochastic nature of early cellular infections and how these dynamics change in response to mutations. By accurately predicting changes in mean virion production and the distribution of viruses in gene-shuffled VSV variants, our model enhances our understanding of viral replication and the variation we see in virion production. Importantly, our findings shed light on the mechanisms underlying the production of VSV virions, revealing the influence of factors such as the duration of the early

Competing interests: The authors have declared that no competing interests exist.

infection phase and the interplay between the genome's ability to switch into an inactive state and viral protein production. We go beyond assessing the mean number of virions produced and examine other effects of mutations, including the probability of aborted infections and the variability in virion production. This stochastic model provides a valuable framework for studying the complex nature of viral replication, contributing to our understanding of single-cell viral dynamics and variability. Ultimately, this knowledge could pave the way for designing more effective strategies to attenuate VSV.

Introduction

Vesicular stomatitis virus (VSV) is a non-segmented negative-strand RNA virus (NNSV) that has been studied for years as an agricultural pest. It is a model for negative-sense RNA viruses and a biomedical tool [1–3]. VSV's extensive characterization makes it an ideal system for model-based investigation of specific viral mechanisms of NNSVs and broader questions about viral behaviors. Since 2000, several mechanistic models have been developed to dissect different characteristics of VSV. Recently, a model of transcription in VSV and respiratory syncytial virus (RSV) revealed a novel theory for why the transcription of genes along the genome follows a predictable gradient dependent on position [4]. This model showed that the long-held understanding of how NNSV transcription works, assuming a single promoter at the 3' end of the genome and a constant rate of attenuation between genes [1,5], could not produce the data that we see in the literature and led to the development of an entirely new model consistent with these observations.

In this study, we seek to understand how the stochastic nature [6] of the reactions that occur during intracellular replication contributes to the variation we see in the number of viruses produced both between different cells and between different variants. Previous models of the dynamics of intracellular replication in VSV have fallen short of capturing the stochastic behavior of virion production while remaining biologically interpretable. The first model was an ordinary differential equation (ODE)-based model made to identify methods to attenuate VSV [7]. This model predicted how balancing transcription and genome synthesis rates might affect viral growth and was later built upon by predicting how all 120 possible gene-shuffled variants of VSV might behave [8]. The model, while thorough, is extremely complex, making it difficult to interpret and work with. In addition, the model is deterministic and, therefore, unable to capture cell-to-cell variation in virion production.

The next model of VSV intracellular replication sought to simulate the cell-to-cell variation of virion production [9]. This model introduces a new hybrid simulation algorithm that includes a new stochastic simulation algorithm called a delayed stochastic simulation algorithm (DSSA). The model of the VSV life cycle has been limited to transcription, translation, and replication reactions; it lacks reactions corresponding to the assembly of the viral particles. In addition, while the model does make a distinction between active and inactive RNA genomes, which is biologically accurate, the active genome is assumed to be the one that is not bound to any proteins. This is inaccurate as the nucleoprotein (N protein) is an essential participant in the facilitation of polymerase activity and it is rare to find genomes that are not encapsidated to some degree [10,11]. Furthermore, the model assumed that the transcription rate of genes is a function of the number of genomes and L polymerase proteins, which was later disproven [12]. As a result of these simplifications, the model validation is limited to a qualitative comparison of the dynamics of the variables representing genome numbers with the known dynamics of virus production.

The third model was an ODE model that modeled the production of N protein encoding mRNA molecules as a function of the number of genomes present in the cell. Here it was shown that the number of mRNA molecules is purely a function of the number of genomes and that the L polymerase does not need to be considered. This is consistent with the observation that VSV forms phase-separated compartments [13,14], and that it is normal for viruses to form factories which may be uninhibited by fluctuations in the concentration of replicase [15,16]. However, for our purposes, this model is too simplistic and does not consider stochasticity. A fourth stochastic model was also developed that only sought to replicate extremely early infection dynamics (less than 20 seconds into the infection) purely to compare parameter estimation methods [17]. This model was too limited for our purposes and did not incorporate the creation of new virions as well.

Here we present the first stochastic model of VSV that incorporates the entire replication cycle. The model relies only on mass action reactions, meaning the only factors considered when calculating the reaction rate are the reactant concentrations, and incorporates the synthesis, degradation, and assembly reactions of VSV and can easily be simulated with the standard Gillespie stochastic simulation algorithms. The model also incorporates a transition of the viral genomes from an active to an inactive state facilitated by the M protein [1,2,18]. The model is the most accurate to date for predicting changes in mean virion production in gene shuffled VSV variants [19] and can additionally capture the distribution of the number of viruses produced by VSV. While we began working with a model that incorporated the individual binding of every single protein that binds to VSV, we ultimately developed a stochastic model that treats the amount of protein that binds to the virus as a single unit. This is consistent with observations that the number of proteins incorporated into a virion varies, such that having an exact number of proteins bind to the genome is not necessarily biologically accurate [20]. We also know that some proteins that make up the virion, like the G protein, assemble independently of the other components before being incorporated into the virion [21]. We believe this to be the best approach to modeling the dynamics of VSV replication, as we capture the stochastic behavior of virion assembly in this model while producing the most accurate predictions to date for the influence of mutations on virion production. This model allowed us to improve our understanding of both intercellular variation and variation between variants in virion production. Intercellular variability appears to be the result of how long it takes for the virus to exit the early phase of infection. However, the differences between variants instead appear to arise from the complexity of balancing the time the (-) genome spends in the active state and the speed of incorporating new (-) genomes into virions, which in turn is balanced by the production of viral components.

Results

Model of VSV replication

The full model can be found in Folder 2 on Github under the name Full_Model.xml.

Molecular species

All species abbreviations are defined in Table 1. A single VSV particle consists of 1 negative-sense strand of RNA genome (nsg), 1258 N, 466 P, 1826 M, 1,205 G, and 50 L [22]. When the RNA genome is released into the cell, the M protein travels to the nucleus while the N, P, and L proteins stay associated with the genome as an RNP complex. As such, we set the initial conditions as a single nsg_N molecule with 466 P, 1,205 G, and 50 L available, the quantity of N to zero since it is already bound to the genome, and M to zero as well since it is in the nucleus and not available in the cytoplasm.

Table 1. List of abbreviations used for each molecular species in the model.

Molecular Species	Abbreviation
Negative-Sense Genome (Free Genome)	nsg
Positive-Sense Genome (Free Antigenome)	psg
Nucleoprotein	N
N bound genome	nsg_N
N bound antigenome	psg_N
Phospho-protein cofactor	P
N and P bound genome	nsg_NP
RNA-dependent RNA polymerase	L
N, P and L bound genome	nsg_NPL
Matrix protein	M
N, P, L, and M bound genome	nsg_NPLM
Glycoprotein	G
N, P, L, M, and G bound genome	nsg_NPLMG
Nucleoprotein mRNA	mN
Phospho-protein cofactor mRNA	mP
Matrix protein mRNA	mM
Glycoprotein mRNA	mG
RNA-dependent RNA polymerase mRNA	mL
Assembled Viral Particle	V
Either nsg, nsg_N, nsg_NP, nsg_NPL	Active nRNP
Either nsg_NPLM, nsg_NPLMG, and V	Inactive nRNP
Either psg or psg_N	pRNP

<https://doi.org/10.1371/journal.pcbi.1011373.t001>

Assembly

In this model, assembly occurs sequentially, starting with the binding of the N proteins to the nsg, followed by subsequent binding of the P proteins, the L proteins, the M proteins, and finally the G proteins (Fig 1). This was inspired by the three layers of the VSV virion with the most internal layer consisting of a single molecule of nsg_NPL which is encased in an outer layer of M proteins and wrapped in a membrane studded with G proteins [1,18]. The binding of each of these factors is broken into several steps, with a single protein binding at each step and each step having the same rate constant listed next to the overall reaction in the table. In this model, psg only recruits the N protein. The literature-derived reaction rates were taken from the binding rate of the N protein from a previous model [9] where the binding rate is assumed to be 0.0461 s^{-1} for the binding of one molecule of N to the genome/antigenome. Here, we make this our initial guess for the binding rate of all proteins.

Transcription

The reactants in the transcription reactions are the nRNP species. This is based on the observation in the literature that the number of mN molecules produced is dependent on only the number of genomes present [12]. In previous models, the transcription of VSV genes has been modeled so that they are coupled to the transcription of any upstream genes in the viral genome [7,9]. Here, we model the transcription reactions as independent from one another. This decision was informed by the conclusions gained from the recently proposed novel model for transcription in VSV and RSV which suggest that transcription of each gene does not require the transcription of each upstream gene [4].

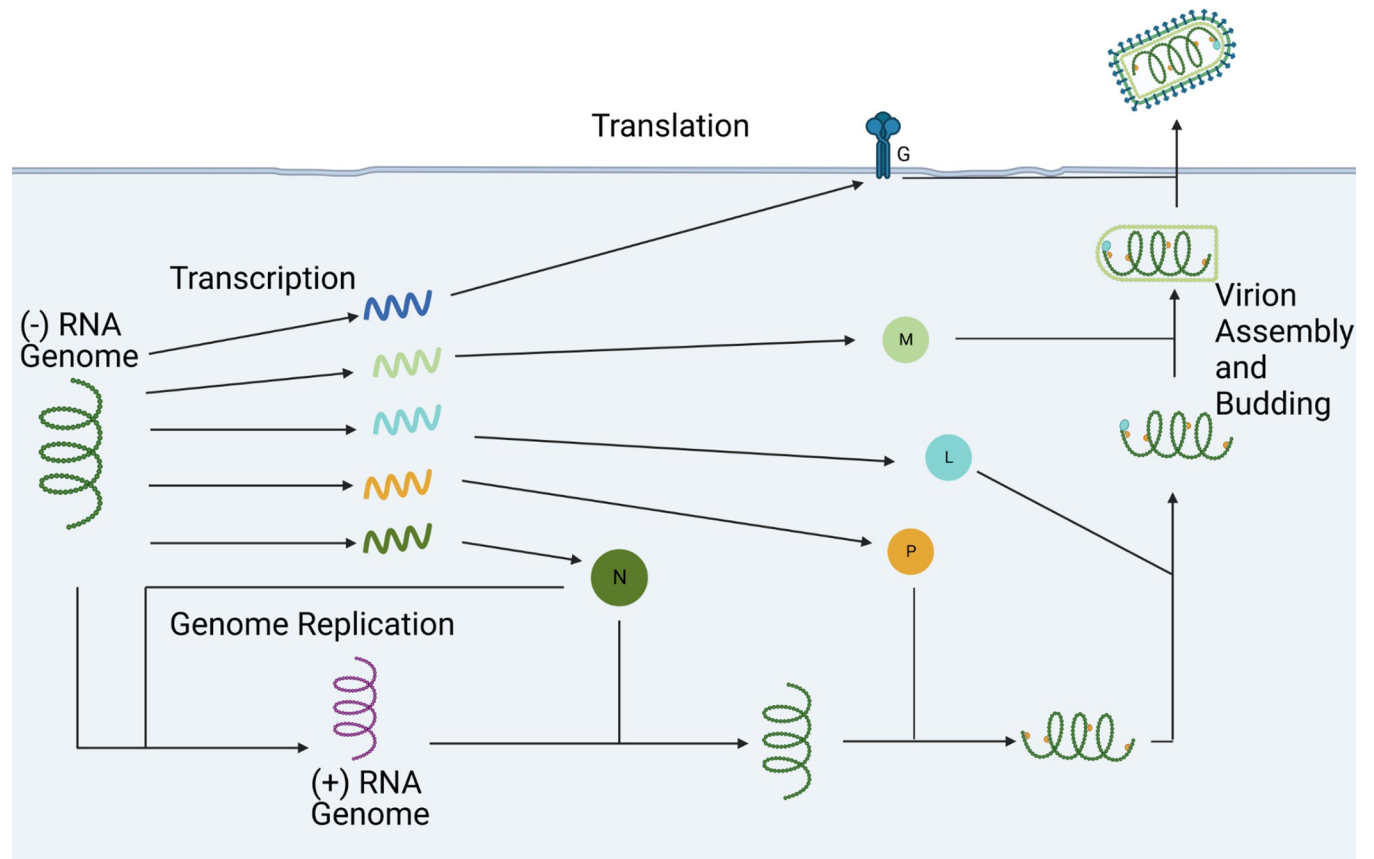


Fig 1. Simplified Model of Vesicular Stomatitis Virus. Visual representation of the simplified model used in this paper. Every molecule depicted has a degradation reaction associated with it that is not represented here.

<https://doi.org/10.1371/journal.pcbi.1011373.g001>

To calculate the transcription rates, we started by calculating the rate of transcribing the N gene. This was inspired from a previous stochastic model that split the transcription rate constants into the rate of polymerase recruitment and a delay to release the reactants [9]. These were combined into a single rate constant. We then calculated the rate constants of all downstream genes by reducing the rate constant of each gene by 30% for each gene-gene junction upstream of the gene, except for the L polymerase which was reduced by 90% for the final gene-gene junction. This is to create the ratio of 1:0.7:0.49:0.343:0.0343 (ratio of mN to mRNA encoded in position 2 to mRNA encoded in position 3 to mRNA encoded in position 4 to mL) that is observed in the literature [23,24].

Translation

Each translation reaction is carried out independently and uses the corresponding mRNA to produce the protein. The rate constant of these reactions we use here assumes uniform saturation of the transcripts by ribosomes and is derived from a previous model [9]. This uniformity of translation is also done to maintain the 1:0.7:0.49:0.343:0.0343 ratio, as it is also present at the protein level. While it is known that during the course of infection, ribosomes can become limiting and change the dynamics of the translation of transcripts [7], we do not incorporate this because we assume that this would influence the translation of each transcript equally.

Genome synthesis

The reactants in the genome synthesis reactions are the pRNP species along with the N protein. This was calculated in a similar fashion to the transcription of the N gene by combining the estimated rate of polymerase recruitment with the delay to produce the antigenome [9]. The decision to include the N protein in the reaction was because the N protein is known to play a pivotal role in facilitating and regulating genome/antigenome synthesis [10,11].

Antigenome synthesis

The reactants in the antigenome synthesis reactions are the nRNP species along with the N protein. Since the amount of pRNP present in a cell is roughly 50 times higher than the amount of nRNP, we multiplied the genome synthesis rate constant by 50 as was done in a previous model [9].

Degradation reactions

All molecular species except for V have a degradation reaction associated with them. All rates are taken from ones used in a previous model [7].

Model simplification

Initially we attempted to simulate the binding of every protein to the genome as a separate reaction, but quickly realized that even when simulating the ODE model took an unreasonable amount of time. We then performed a series of tests where we increased the number of reactions and timed how the length of loading and simulating the ODE model changed as the number of reactions increased (Python Notebook 1 on Github). We found that these times increased exponentially as we increased the reaction number (S1 Fig). This informed our ultimate decision that the best model to use is the simplest one. In this model, each molecule of protein represents a unit of the protein that makes up a virion. The reaction rate associated with these binding reactions represents the average amount of time it takes for the binding of all the proteins of that type to occur. So, we adjusted the rate constant of any reactions that involve the creation or consumption of proteins by dividing the rate constant by the number of proteins that is in a virion. For example, if the binding rate of protein P to the nsg is 0.0461 then it would be adjusted to be 0.0461/466 because 466 P proteins are in a single virion. While there may be some concern in how this might affect the variability of the aggregation rate of the virus, it is known that the rate of the release of the first virus in single cell infections is highly variable usually occurring within 2–6 hours of infection [33] suggesting that early viral aggregation is highly variable.

A visual representation of the reactions that make up the final model, not including the degradation reactions, can be seen in Fig 1. A table of all the reactions in the final model can also be seen in Table 2. Each reaction is mass action, so the reaction rate for each is the rate constant multiplied by the concentration of the reactants.

Model fitting

We then wanted to see if the parameters of the model could be adjusted to fit the mean number of virions produced by VSV gene shuffled variants that have been previously recorded in literature [19]. We chose the data from BHK-cl.13 cells because it was the dataset where the mean number of viruses produced by the wild-type was closest to the mean number of viruses recorded when infecting single cells with wild-type VSV [25]. The mean number of virions used for comparison was taken ~16 hours post infection (57,000 seconds). This time was

Table 2. Model Reactions with Reactants and Products.

Reaction Name	Reactants	Products
nsg synthesis	pRNP + N	pRNP + nsg + N
psg synthesis	Active nRNP + N	Active nRNP + psg + N
nsg encapsidation	nsg + N	nsg_N
psg encapsidation	psg + N	psg_N
P binding	nsg_N + P	nsg_NP
L binding	nsg_NP + L	nsg_NPL
M binding	nsg_NPL + M	nsg_NPLM
G binding	nsg_NPLM + G	nsg_NPLMG
Virion budding	nsg_NPLMG	V
Transcription N	Active nRNP	Active nRNP + mN
Transcription P	Active nRNP	Active nRNP + mP
Transcription M	nsg	nsg + mM
Transcription G	nsg	nsg + mG
Transcription L	nsg	nsg + mL
Translation N	mN	mN + N
Translation P	mP	mP + P
Translation M	mM	mM + M
Translation G	mG	mG + G
Translation L	mL	mL + L
N degradation	N	
P degradation	P	
M degradation	M	
G degradation	G	
L degradation	L	
mN degradation	mN	
mP degradation	mP	
mM degradation	mM	
mG degradation	mG	
mL degradation	mL	
nsg degradation	nRNP	
psg degradation	pRNP	

<https://doi.org/10.1371/journal.pcbi.1011373.t002>

chosen because there is not a consensus on how long VSV takes to kill a cell, but it is known that virions start emerging around 2–6 hours after infection [26] and that VSV's mechanism for killing cells is tied to the cell cycle [27]. So, if we assume the average cell cycle time in the cells we are modeling is about a day [28] and that the cells are asynchronous, then an average time until death of 16 hours should be reasonable.

We started by manually adjusting the parameters to get the wild-type model to be close to the mean number of virions produced by the wild-type VSV. We then used these parameters to make the models for the gene shuffled variants by assigning the transcription rate constants to each gene based on its position in the genome. This model, however, failed to reproduce the mean number of virions produced by the gene-shuffled variants and resulted in a very large MSE (Table 3). These models can be found in Folder 3 on Github. We then proceeded to fit our model, but due to the computational burden of fitting the stochastic model we had to fit our model using a combination of fitting the ODE model followed by fitting the stochastic model. This final stochastic model was able predict changes in virion production in our gene shuffled variant dataset (Table 3). The fit model as well as other intermediate models can be

Table 3. Model Predictions and Mean Square Error.

VSV Variant	Data From Literature	Manually Adjusted Param.	Stochastic Model
NMGPL	4.35×10^3	7.69×10^2	$4.27 \times 10^3 \pm 46$
NPMGL (WT)	2.09×10^3	1.89×10^3	$1.87 \times 10^3 \pm 20$
NMPGL	2.35×10^3	5.72×10^2	$2.55 \times 10^3 \pm 26$
NGPML	1.30×10^3	3.05×10^3	$1.33 \times 10^3 \pm 10$
NGMPL	3.96×10^3	1.90×10^3	$3.93 \times 10^3 \pm 41$
MSE		4.66×10^6	1.92×10^4

<https://doi.org/10.1371/journal.pcbi.1011373.t003>

found in Folder 5 on Github. This stochastic model allowed us to also visualize the distribution in the number of virions produced by the wild-type model and see that it is consistent with viral distributions observed in the literature (S2 Fig) [25,29,30].

Individual infections that exit the early phase of infection with more mRNA produce more virions

The individual trajectories of simulations using this model (Fig 2A) reveal heterogeneity in both the number of virions produced and the time at which the virions begin to be produced. Early on in infection, the random chance a reaction may or may not occur can have long lasting effects on the ultimate outcome of that infection, so we wanted to see if we could identify which early events contribute to the variation in the number of virions produced. We first set a distinction between the early phase of infection and the rest of the infection cycle by identifying when the first virion is produced. Here we define the time until the production of the first

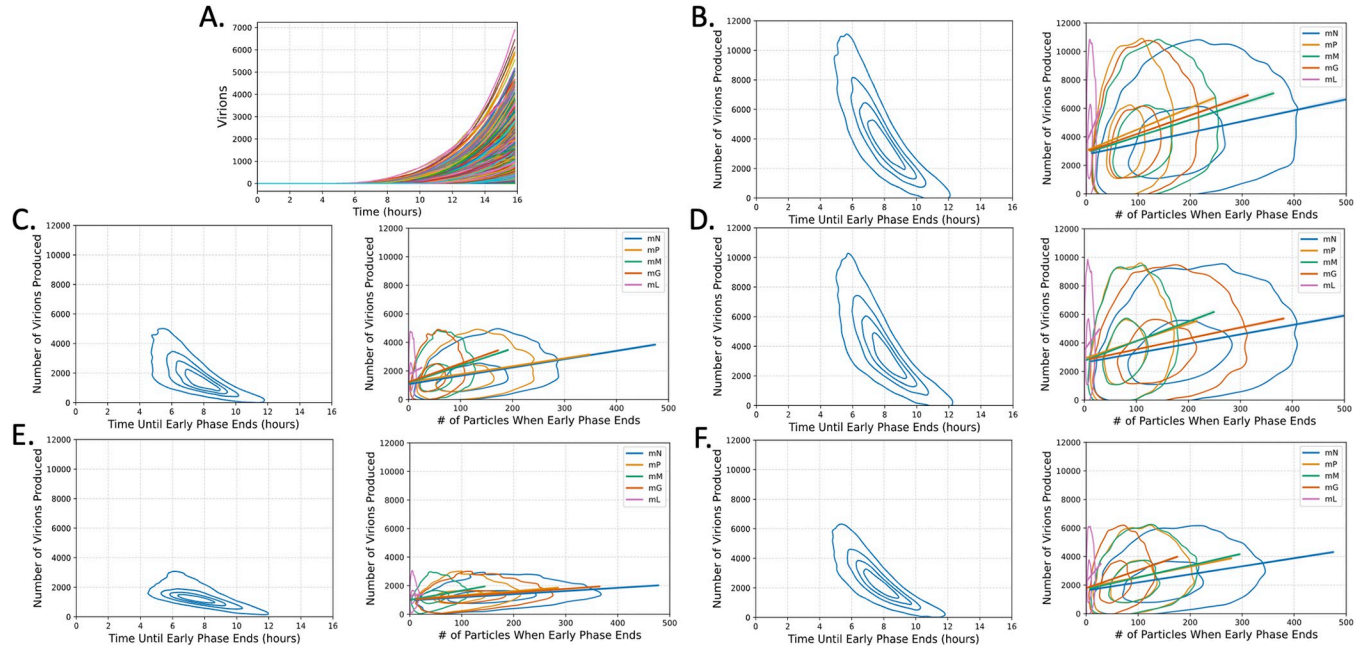


Fig 2. VSV infections that exit the early phase of infection earlier with more mRNA present produce more virions. (A) 1000 stochastic simulations of wild-type VSV showing the number of virions produced over time. (B-F) Left: Kernel density estimate (KDE) plots of the number of virions produced vs. the time until the early phase ends. Right: KDE plots of the quantity of each mRNA species present when the early phase ends, along with the 99% CI line of best fit, for variants (B) NMGPL, (C) NGMPL, (D) NMPGL, (E) NPMGL (wild-type), and (F) NGPML. All mRNA species show a weakly positive correlation ($0.2 < r < .39$) with the number of virions produced, except for mL which has a very weak correlation ($0 < r < .19$) in all variants.

<https://doi.org/10.1371/journal.pcbi.1011373.g002>

virion as the time it takes for the infection to leave the early phase. When we plot the time until the early phase is left against the number of virions produced (Fig 2B–2F) we can see that there is a visible trend that the longer it takes for the early phase to end, the fewer virions that are ultimately produced.

This naturally then led us to investigate what early molecular species could also be participating in these observations. We then turned our attention to the quantity of mRNA present when the infection leaves the early phase, as the transcription reactions are some of the earliest reactions that occur upon infection. We plotted the number of mRNA present at the time when the infection exits the early phase against the number of virions produced (Fig 2B–2F) and revealed a clear positive correlation between every mRNA and the number of virions produced. These observations together suggest that the variation we see in the number of virions produced within variants is likely due to variation in how long it takes for the infections to stockpile resources and actively start producing viruses. The script used to perform these analyses and generate the figures can be found in Python Notebook 2 on Github.

The positions of the N, P, and L genes are the most important for maximizing virion production

We next turned our focus to identifying what is influencing the variation in the mean number of virions produced between different variants. We started by generating all 120 possible gene-shuffled VSV variants to observe the fitness landscape of these variants (Fig 3A). It appears that most gene-shuffled variants are at least weakly viable (produce more than 1 virion per infection), but few produce virions on the scale of the wild-type virus (>1000). Even more intriguing, however, is that there are nine that are predicted to produce more virions than the wild type (Table 4).

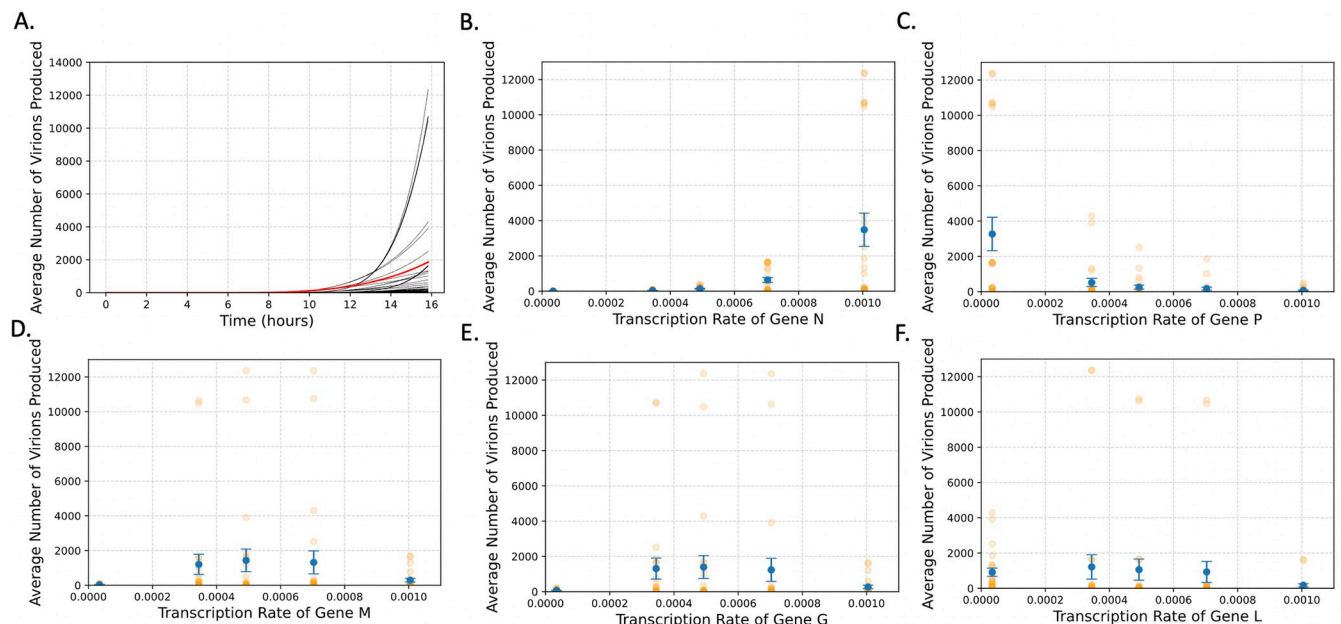


Fig 3. N and P gene position most strongly influence virion production in gene shuffled VSV variants. (A) The average number of virions produced over time for all 120 variants. The trajectory of the wild-type variant is in red. (B-F) The average number of virions produced as a function of gene position (which determines the transcription rate) for genes (B) N, (C) P, (D) M, (E) G, and (F) L. The error bars represent the standard error of the mean, and the orange dots represent the individual variants.

<https://doi.org/10.1371/journal.pcbi.1011373.g003>

Table 4. Highest Virion Producing Gene Shuffled Vesicular Stomatitis Virus Variants with Associated Statistics.

Variant	Mean Virions Produced	Standard Deviation	Percent of Infections Aborted	Mean/Standard Deviation
NMGLP	12365.23	11777.92	0.16	1.05
NGMLP	12355.23	11534.64	0.10	1.07
NMLGP	10742.46	9275.01	0.08	1.16
NLMGP	10673.82	9092.71	0.08	1.17
NGLMP	10629.27	8462.59	0.06	1.26
NLGMP	10481.59	8520.48	0.12	1.23
NMGPL	4294.93	2384.08	0.28	1.80
NGMPL	3914.32	2119.58	0.28	1.85
NMPGL	2516.32	1332.83	0.48	1.89
NPMGL (WT)	1863.50	995.86	0.52	1.87

<https://doi.org/10.1371/journal.pcbi.1011373.t004>

To identify what features made these variants so successful, we then looked at the average number of virions produced and the percent of infections aborted (infections producing < 1 virion) by every variant as a function of each transcription rate to look for any patterns that may explain this variation (Figs 3B–3F and 4). For vitality, this revealed a negative correlation between the transcription rate of the N gene and the number of infections that were aborted. Most notably, when the N gene is in position 1, 2, or 3 nearly all infections are fruitful, while in position 5 nearly all infections are aborted. There appeared to be no obvious relation between the rate of aborted infections with the other genes, however (Fig 4).

When looking at the number of virions produced, however, a much more complex behavior emerged. We observed a positive relationship with the transcription rate of N with the number of virions produced, but a negative correlation with the transcription rate of P. There

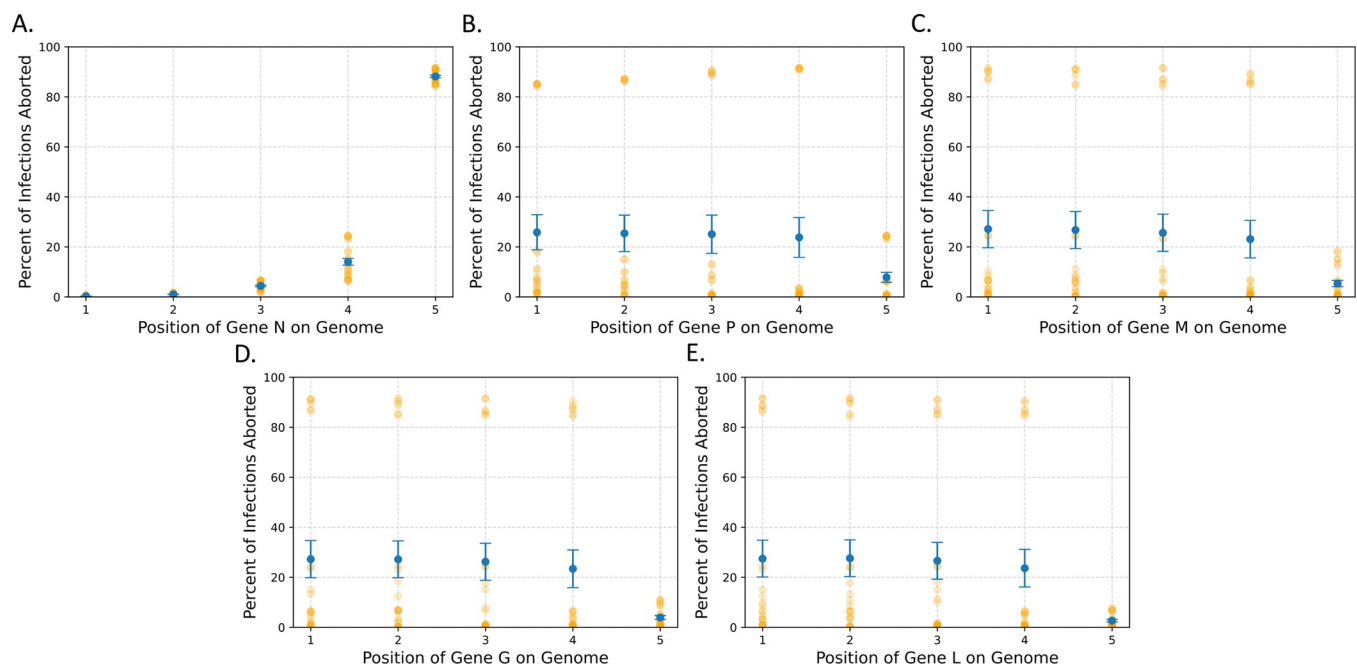


Fig 4. N gene position most strongly influences the percentage of infections that are aborted. The average percent of infections aborted as a function of the gene position of genes (A) N, (B) P, (C) M, (D) G, and (E) L for all 120 variants. The gene position is directly related to transcription rate of each gene. The error bars are the standard error of the mean and the orange dots represent the individual variants.

<https://doi.org/10.1371/journal.pcbi.1011373.g004>

also appeared to be a nonlinear relationship between the number of virions produced and the transcription rates of L, M, and G, all of which showed a small peak in the middle of the transcription rates. These observations are consistent with the gene orientation of the predicted most fruitful variant (NMGLP), as in this variant the N transcription rate is maximized, and the P transcription rate minimized with the other three in the middle.

These plots also revealed a few rules for what positions these different genes need to be in to produce a virus that yields more virions than the wild-type variant. The first is that the N gene must be in the first position of the genome. The next is the L and P genes must be in the 3rd, 4th, or 5th positions, which is consistent with the idea that the transcription rate of P is negatively associated with the number of virions produced. Finally, the G and M genes must be either in the 2nd, 3rd, or 4th positions, given the peak we observe for these middle rates.

These results suggest that the transcription rates of P and N appear to be the most influential on the number of virions each variant produce. This is further supported by the fact that all 6 variants that have N in the 1st position and P in the 5th position are the top 6 virion producing variants. However, it also appears that the position of L seems to be also important, as within these top 6 variants, if L is in the 4th gene position the average yield is ~12350 virions, while if it is in any other position, we only see ~10000 virions produced (S1 Table). In short, it looks like the best virion production occurs when the transcription of L and P are minimized and N is maximized. The script used to perform these analyses and generate the figures can be found in Python Notebook 3 on Github.

However, this only explains the variation in the most prolific phenotypes. It does not explain the wide variation in virion production of the other 14 variants that produce more than 1000 virions. This is because we are still left with only a fraction of an image of what the relationships between these transcription rates and the number of the virions produced are, because to alter one transcription rate in these models, you must alter at least one more. This is because when altering the position of a gene, the gene in the position that the first gene is moved into is moved into the original position of the first gene, altering the transcription rate of both genes.

The balance of all transcription rates plays a role in determining virion production

The next step was to see how altering individual transcription rates independently could alter the number of virions produced. For this section we looked at the best variant and the wild-type variant to see how altering each transcription rate by values between -25% and +25% would affect virion production (Fig 5). The wild-type variant reveals that altering every transcription rate has potential to alter the number of virions produced. Here, it appears that the transcription rates of the genes encoding mN, mM, and mG have a positive association with the while the rates for genes encoding mP and mL have a negative association. The most fruitful variant (NMGLP), however, is only strongly altered by changing the transcription rate of mN which also has a positive association, but interestingly, has a very slight positive association with increasing the transcription rate of mP as well. The script used to perform these analyses can be found in Python Notebook 3 on Github.

These two observations together suggest that for each variant, there is an optimal balance of these transcription rates that maximizes virion production that is not simply achieved by maximizing all transcription rates. It appears that there is a hierarchy where the best virion production is achieved by maximizing N transcription but also by balancing the production of G, M, L, and P transcription, and deviations from this balance can alter virion production. This suggests that producing more virions is not achieved by merely maximizing transcription rates, but also by considering the interplay between all species.

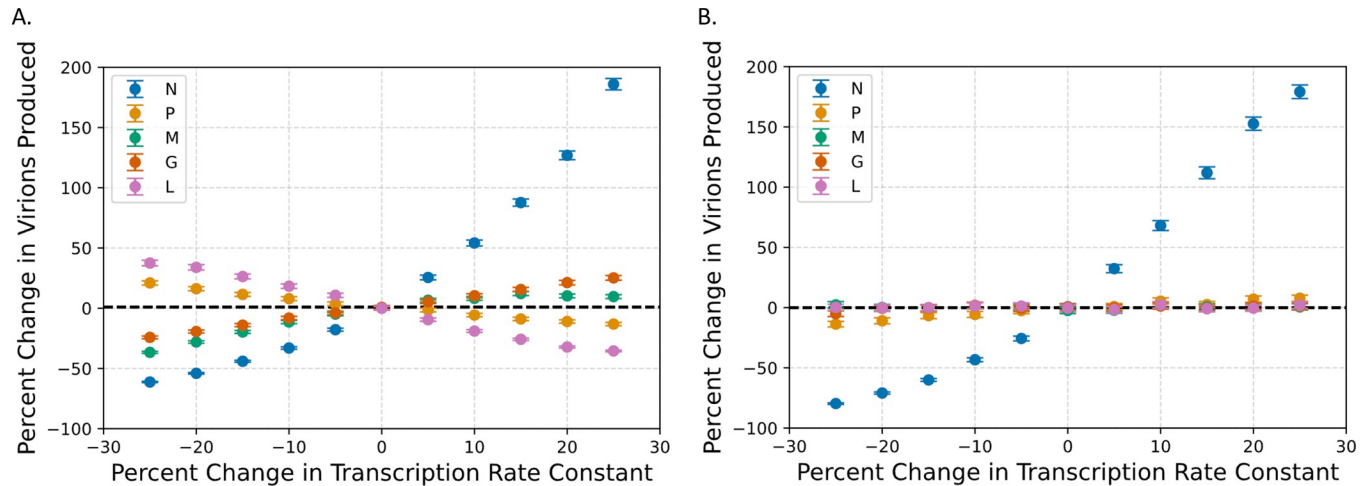


Fig 5. Parameter scans reveal the influence of L, M, and G genes on virion production. These graphs show how adjusting the transcription rates of each gene by a percentage of the value of the parameter in (A) WT and (B) the optimal variant (NMGLP) influences the number of virions produced. The black dotted line shows the mean number of virions produced by those variants.

<https://doi.org/10.1371/journal.pcbi.1011373.g005>

Discussion

In this study we sought to understand how the stochastic nature of the reactions in early cellular infections influence the variation we observe in the number of VSV virions produced by these cells. Previous models have attempted to investigate the cellular dynamics of VSV, but these models are either ODE-based [7,12] or do not represent the full replication cycle due to the computational burden required to investigate the model [9]. The model we present here is the first stochastic mechanistic model that captures the full replicative cycle of VSV. The model is also the most accurate one to date in predicting changes in virion production in response to gene shuffling. However, we want to acknowledge that our model contains the assumption that the kinetics of the reactions do not change during the course of the infection. Some of these reactions likely do change, but without this explicitly included in our model we are still able to achieve accurate predictions with the available data and we argue that this suggests our model is capturing some of these dynamics. In the case that our predictions do not hold up when new data emerges, it may be necessary to revisit our model and alter this assumption.

This model provides valuable insight into the sources of variation in virion production, both within a single variant as well as between variants. Within variants, the main determinant of the number of virions produced appears to be how long it takes to leave the early phase of infection (Fig 2B–2F). This is consistent with the observation that some of the variability in virion production can be explained by what phase in the cell cycle the cell is infected in [25]. VSV's mechanism for killing cells is tied to the cell cycle, and so the phase the cell is in could influence the ultimate number of virions that can be produced by determining how long VSV has to start replicating [27]. Here, the variation in the amplifying time of VSV is the result of delays in entering the replicative cycle, which may allow us to capture some of the variation that is due to infecting cells at different cell cycle stages. Nonetheless, this intrinsic feature we identified is likely to play a role in intercellular variation. What is also exciting is that we can capture these delays by implementing the traditional stochastic simulation algorithm without the implementation of a DSSA.

As for the variation between VSV variants, our model seems to suggest that the most important factor for determining the number of virions produced is the transcription rate of

the N gene. This is evident as all variants that produce more than 2000 virions have the N gene in the first position (Fig 3B). In addition, when we increased the transcription rate of each gene independently, N led to the largest increase in the number of virions produced by both the wild-type and most fruitful gene shuffled variant (NMGLP) (Fig 5). Furthermore, the transcription rate of the N gene appears also to play a major role in how frequently infections are aborted (Fig 4). These findings are consistent with observations in the literature that genome and antigenome replication are both strongly dependent on and regulated by the N protein [10,11]. As we increase the transcription rate of the N gene, we likely also increase the rate of replication in the cell, producing more genomes and ultimately, exponentially more virions.

However, we also see that the other genes also actively participate in determining the virion output, although to a lesser extent. There is clearly a complex balance of transcription rates between M, G, L, and P that allows for the optimal growth rate, and deviations from this optimal balance can alter virion production. This is likely due to the distinction between active vs. inactive nsg in the model. Since L and P both bind to the active genome, excess L and P might cause newly synthesized genomes to be converted into the inactive form too quickly. This may explain the negative correlation between these transcription rates and virion production in the wild-type. Indeed, there is evidence that the P protein may inhibit transcription [31]. M and G then convert those inactive genomes into virions, and as such, they can influence virion production by forming bottlenecks during virion assembly. L and P have the potential to do this as well, and so it may be that maximizing virion production is a matter of balancing how long nsgs stay in the active state with how quickly they are incorporated into virions.

The negative relationship of the transcription rate of the P and L genes and the positive relationship of the transcription rate of the N gene with regards to virion production are consistent with the results of a previous model that also sought to predict how all 120 variants of VSV may behave [8]. However, our predictions appear to put a heavier weight on the transcription rate of P, while the previous model put more emphasis on L. As for the M and G genes, both our and their results support the notion that maximum virion output is achieved when these are in gene positions 2, 3, or 4, and that when M and G are in position 5, almost no virions are produced. However, our results differ in that putting these genes in position 1 also is more detrimental than they predicted and is similar to putting them in position 5.

As for the overall fitness landscape of the variants, they also predicted that several variants produce as many virions as the wild-type as well as a few that produce more than the wild-type [8]. However, the different relationships between transcription rates we found alter which variants are in these positions. For example, their predicted best gene shuffled variants are the NMPGL and NMGPL variants, while our predicted best variant is the NMGLP variant. Likely these differences are due to differences in the datasets utilized for fitting the model, as we fit the mean number of virions produced to the mean number of virions produced per cell in the five VSV gene shuffled variants, while their model was fit to ranked variant data of six gene shuffled VSV variants from a previous paper [32] and their own one-step growth curve of the wild-type. Furthermore, the cell type of the ranked variant data was from BSC-1 cells, while the data we utilized was from BHK-cl.13 cells. This is extremely important as the number of virions produced and the ranking of variants is highly dependent on the cell line utilized to grow the virus [19]. For example, in our dataset of the five gene-swapped variants, the wild-type was the fourth highest producing variant, while in the dataset they used, the wild-type was the third highest producing variant of six gene shuffled variants.

However, since our model is stochastic and the previous model is an ODE model, our model has additional traits that we can observe in the variants that are not possible in the ODE model. We can observe the distribution of the number of virions produced by different variants (S2 Fig) as well as predict how frequently infections may be aborted (Fig 4). We can also

calculate additional statistics about the virion distribution, allowing us to select for more complex phenotypes when designing specific viral behaviors. For example, if we wanted to maximize the mean number of virions produced while minimizing the standard deviation, we could calculate the ratio of mean to standard deviation (Tables 4 and S1). Our model currently predicts that the best variant for achieving such a phenotype would be variant NPMLG. These attributes of our model allow for more in-depth analyses of the effects of different mutations on the viral life cycle than any previous model beforehand.

Considering the limited size of the datasets used to estimate the model parameters, the model is likely to be overfitted. Ongoing work in our lab aims at increasing the amount of data available to refine this model. This will be achieved by producing a large library of VSV variants and testing them on different cell lines [33]. The library will include all the gene-shuffled variants as well as codon-deoptimized versions [34] of the genes to control their translation independently of their transcription. We will also broaden the scope of phenotypic data beyond the number of virions at one time point by observing the dynamics of their replication by time-lapse microscopy [35,36] and estimating the number of mRNA molecules present in infected cells [37,38].

Methods

Model creation was formalized as previously described [39].

Model creation

The models were encoded in XML format. These files were generated by creating two Excel spreadsheets, one containing the reactions and their parameters and the other containing the initial number of molecular species. The models were then fed to the Python script provided in Folder 1 which can be found on Github.

Model simulation

Models were simulated using COPASI's Python implementation, COPASI Basico [40]. ODE models were simulated using the LSODA Method, while stochastic simulations were performed using the Gillespie Direct Method. Models were simulated for ~16 hours (57,000 seconds). The dataset used to analyze the stochastic models consisted of 10,000 trajectories of each variant.

Parameter optimization

Models were fit using the minimize function from SciPy [41]. The Nelder-mead method was utilized. ODE and stochastic Models were fit by calculating the mean squared error (MSE) of the number of virions present at the endpoint (57,000 s) and the average number of viruses produced recorded in the literature. The scripts used to fit the model can be found in Folder 4 on Github.

Cloud computing

When performing large numbers of optimizations or stochastic simulations, we utilized Amazon Web Service's Elastic Cloud Computing service to create an instance that could handle the computations. We utilized a single c5.18xlarge instance which is a computation optimized instance with 72 virtual CPUs and 144 Gigabytes of memory.

Supporting information

S1 Fig. Simulation and Loading time of ODE Models Increases as the Number of Reactions in the Model Increase. These two graphs show how the time to load each model (B) and simulate the model (A) increase drastically as the number of reactions in the model increases. (PDF)

S2 Fig. Distribution of the number of virions produced by each gene shuffled Vesicular Stomatitis Virus variant used to fit the model. These graphs show the distribution for the final number of virions produced for 10,000 simulations of each gene-shuffled variant in the training set. (PDF)

S1 Table. Gene Shuffled Vesicular Stomatitis Virus Variants with Associated Statistics. (PDF)

Author Contributions

Conceptualization: Jean Peccoud.

Funding acquisition: Jean Peccoud.

Methodology: Connor R. King.

Software: Connor R. King.

Visualization: Connor R. King.

Writing – original draft: Connor R. King.

Writing – review & editing: Casey-Tyler Berezin, Jean Peccoud.

References

1. Liu G, Cao W, Salawudeen A, Zhu W, Emeterio K, Safronetz D, et al. Vesicular Stomatitis Virus: From Agricultural Pathogen to Vaccine Vector. *Pathogens*. 2021; 10(9). Epub 20210827. <https://doi.org/10.3390/pathogens10091092> PMID: 34578125; PubMed Central PMCID: PMC8470541.
2. Lichty BD, Power AT, Stojdl DF, Bell JC. Vesicular stomatitis virus: re-inventing the bullet. *Trends Mol Med*. 2004; 10(5):210–6. <https://doi.org/10.1016/j.molmed.2004.03.003> PMID: 15121047.
3. Overend C, Yuan LJ, Peccoud J. The synthetic futures of vesicular stomatitis virus. *Trends in Biotechnology*. 2012; 30(10):497–8. <https://doi.org/10.1016/j.tibtech.2012.06.002> WOS:000309946600001. PMID: 22789133
4. Piedra FA, Henke D, Rajan A, Muzny DM, Doddapaneni H, Menon VK, et al. Modeling nonsegmented negative-strand RNA virus (NNSV) transcription with ejective polymerase collisions and biased diffusion. *Front Mol Biosci*. 2022; 9:1095193. Epub 20230109. <https://doi.org/10.3389/fmolb.2022.1095193> PMID: 36699700; PubMed Central PMCID: PMC9868645.
5. Banerjee AD, Abraham G, Colonno RJ. Vesicular stomatitis virus: mode of transcription. *J Gen Virol*. 1977; 34(1):1–8. <https://doi.org/10.1099/0022-1317-34-1-1> PMID: 188975
6. Peccoud J, Ycart B. Markovian Modeling of Gene-Product Synthesis. *Theoretical Population Biology*. 1995; 48(2):222–34. <https://doi.org/10.1006/tpbi.1995.1027> WOS:A1995RZ41300005.
7. Lim KI, Lang T, Lam V, Yin J. Model-based design of growth-attenuated viruses. *PLoS Comput Biol*. 2006; 2(9):e116. Epub 20060724. <https://doi.org/10.1371/journal.pcbi.0020116> PMID: 16948530; PubMed Central PMCID: PMC1557587.
8. Lim KI, Yin J. Computational fitness landscape for all gene-order permutations of an RNA virus. *PLoS Comput Biol*. 2009; 5(2):e1000283. Epub 20090206. <https://doi.org/10.1371/journal.pcbi.1000283> PMID: 19197345; PubMed Central PMCID: PMC2627932.
9. Hensel SC, Rawlings JB, Yin J. Stochastic kinetic modeling of vesicular stomatitis virus intracellular growth. *Bull Math Biol*. 2009; 71(7):1671–92. Epub 20090521. <https://doi.org/10.1007/s11538-009-9419-5> PMID: 19459014; PubMed Central PMCID: PMC3169382.

10. Hanke L, Schmidt FI, Knockenhauer KE, Morin B, Whelan SP, Schwartz TU, et al. Vesicular stomatitis virus N protein-specific single-domain antibody fragments inhibit replication. *EMBO Rep.* 2017; 18(6):1027–37. Epub 20170410. <https://doi.org/10.15252/embr.201643764> PMID: 28396572; PubMed Central PMCID: PMC5452021.
11. Arnheiter H, Davis NL, Wertz G, Schubert M, Lazzarini RA. Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell.* 1985; 41(1):259–67. [https://doi.org/10.1016/0092-8674\(85\)90079-0](https://doi.org/10.1016/0092-8674(85)90079-0) PMID: 2986844.
12. Timm C, Gupta A, Yin J. Robust kinetics of an RNA virus: Transcription rates are set by genome levels. *Biotechnol Bioeng.* 2015; 112(8):1655–62. Epub 20150520. <https://doi.org/10.1002/bit.25578> PMID: 25726926; PubMed Central PMCID: PMC5653219.
13. Heinrich BS, Maliga Z, Stein DA, Hyman AA, Whelan SPJ. Phase Transitions Drive the Formation of Vesicular Stomatitis Virus Replication Compartments. *mBio.* 2018; 9(5). Epub 20180904. <https://doi.org/10.1128/mBio.02290-17> PMID: 30181255; PubMed Central PMCID: PMC6123442.
14. Monette A, Moulard AJ. Zinc and Copper Ions Differentially Regulate Prion-Like Phase Separation Dynamics of Pan-Virus Nucleocapsid Biomolecular Condensates. *Viruses.* 2020; 12(10). Epub 20201018. <https://doi.org/10.3390/v12101179> PMID: 33081049; PubMed Central PMCID: PMC7589941.
15. Nikolic J, Le Bars R, Lama Z, Scrima N, Lagaudriere-Gesbert C, Gaudin Y, et al. Negri bodies are viral factories with properties of liquid organelles. *Nat Commun.* 2017; 8(1):58. Epub 20170705. <https://doi.org/10.1038/s41467-017-00102-9> PMID: 28680096; PubMed Central PMCID: PMC5498545.
16. Netherton CL, Wileman T. Virus factories, double membrane vesicles and viroplasm generated in animal cells. *Curr Opin Virol.* 2011; 1(5):381–7. Epub 20111012. <https://doi.org/10.1016/j.coviro.2011.09.008> PMID: 22440839; PubMed Central PMCID: PMC7102809.
17. Gupta A, Rawlings JB. Comparison of Parameter Estimation Methods in Stochastic Chemical Kinetic Models: Examples in Systems Biology. *AIChE J.* 2014; 60(4):1253–68. Epub 20140305. <https://doi.org/10.1002/aic.14409> PMID: 27429455; PubMed Central PMCID: PMC4946376.
18. Zhou K, Si Z, Ge P, Tsao J, Luo M, Zhou ZH. Atomic model of vesicular stomatitis virus and mechanism of assembly. *Nat Commun.* 2022; 13(1):5980. Epub 20221010. <https://doi.org/10.1038/s41467-022-33664-4> PMID: 36216930; PubMed Central PMCID: PMC9549855.
19. Novella IS, Ball LA, Wertz GW. Fitness analyses of vesicular stomatitis strains with rearranged genomes reveal replicative disadvantages. *J Virol.* 2004; 78(18):9837–41. <https://doi.org/10.1128/JVI.78.18.9837-9841.2004> PMID: 15331718; PubMed Central PMCID: PMC514966.
20. Hodges J, Tang X, Landesman MB, Ruedas JB, Ghimire A, Gudheti MV, et al. Asymmetric packaging of polymerases within vesicular stomatitis virus. *Biochem Biophys Res Commun.* 2013; 440(2):271–6. Epub 20130919. <https://doi.org/10.1016/j.bbrc.2013.09.064> PMID: 24055706; PubMed Central PMCID: PMC4350925.
21. Brown EL, Lyles DS. Organization of the vesicular stomatitis virus glycoprotein into membrane microdomains occurs independently of intracellular viral components. *J Virol.* 2003; 77(7):3985–92. <https://doi.org/10.1128/jvi.77.7.3985-3992.2003> PMID: 12634358; PubMed Central PMCID: PMC150637.
22. Thomas D, Newcomb WW, Brown JC, Wall JS, Hainfeld JF, Trus BL, et al. Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. *J Virol.* 1985; 54(2):598–607. <https://doi.org/10.1128/JVI.54.2.598-607.1985> PMID: 2985822; PubMed Central PMCID: PMC254833.
23. Iverson LE, Rose JK. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. *Cell.* 1981; 23(2):477–84. [https://doi.org/10.1016/0092-8674\(81\)90143-4](https://doi.org/10.1016/0092-8674(81)90143-4) PMID: 6258804.
24. Barr JN, Whelan SP, Wertz GW. Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Biochim Biophys Acta.* 2002; 1577(2):337–53. [https://doi.org/10.1016/s0167-4781\(02\)00462-1](https://doi.org/10.1016/s0167-4781(02)00462-1) PMID: 12213662.
25. Zhu Y, Yongky A, Yin J. Growth of an RNA virus in single cells reveals a broad fitness distribution. *Virology.* 2009; 385(1):39–46. Epub 20081213. <https://doi.org/10.1016/j.virol.2008.10.031> PMID: 19070881; PubMed Central PMCID: PMC2666790.
26. Holzwarth G, Bhandari A, Tommervik L, Macosko JC, Ornelles DA, Lyles DS. Vesicular stomatitis virus nucleocapsids diffuse through cytoplasm by hopping from trap to trap in random directions. *Sci Rep.* 2020; 10(1):10643. Epub 20200630. <https://doi.org/10.1038/s41598-020-66942-6> PMID: 32606395; PubMed Central PMCID: PMC7326962.
27. Chakraborty P, Seemann J, Mishra RK, Wei JH, Weil L, Nussenzveig DR, et al. Vesicular stomatitis virus inhibits mitotic progression and triggers cell death. *EMBO Rep.* 2009; 10(10):1154–60. Epub 20090911. <https://doi.org/10.1038/embor.2009.179> PMID: 19745842; PubMed Central PMCID: PMC2759734.

28. Wang Z. Regulation of Cell Cycle Progression by Growth Factor-Induced Cell Signaling. *Cells*. 2021; 10(12). Epub 20211126. <https://doi.org/10.3390/cells10123327> PMID: 34943835; PubMed Central PMCID: PMC8699227.
29. De Paepe M, De Monte S, Robert L, Lindner AB, Taddei F. Emergence of variability in isogenic *Escherichia coli* populations infected by a filamentous virus. *PLoS One*. 2010; 5(7):e11823. Epub 20100727. <https://doi.org/10.1371/journal.pone.0011823> PMID: 20676396; PubMed Central PMCID: PMC2910729.
30. Schulte MB, Andino R. Single-cell analysis uncovers extensive biological noise in poliovirus replication. *J Virol*. 2014; 88(11):6205–12. Epub 20140319. <https://doi.org/10.1128/JVI.03539-13> PMID: 24648454; PubMed Central PMCID: PMC4093869.
31. Bloyet LM, Morin B, Brusica V, Gardner E, Ross RA, Vadakkan T, et al. Oligomerization of the Vesicular Stomatitis Virus Phosphoprotein Is Dispensable for mRNA Synthesis but Facilitates RNA Replication. *J Virol*. 2020; 94(13). Epub 20200616. <https://doi.org/10.1128/JVI.00115-20> PMID: 32321813; PubMed Central PMCID: PMC7307139.
32. Ball LA, Pringle CR, Flanagan B, Perepelitsa VP, Wertz GW. Phenotypic consequences of rearranging the P, M, and G genes of vesicular stomatitis virus. *J Virol*. 1999; 73(6):4705–12. <https://doi.org/10.1128/JVI.73.6.4705-4712.1999> PMID: 10233930; PubMed Central PMCID: PMC112512.
33. Gallegos JE, Adames NR, Rogers MF, Kraikivski P, Ibele A, Nurzynski-Loth K, et al. Genetic interactions derived from high-throughput phenotyping of 6589 yeast cell cycle mutants. *NPJ systems biology and applications*. 2020; 6(1):1–15.
34. Sharma D, Baas T, Nogales A, Martinez-Sobrido L, Gromiha MM. CoDe: a web-based tool for codon deoptimization. *Bioinform Adv*. 2023; 3(1):vbac102. Epub 2023/01/27. <https://doi.org/10.1093/bioadv/vbac102> PMID: 36698765; PubMed Central PMCID: PMC9832946.
35. Ball DA, Lux MW, Adames NR, Peccoud J. Adaptive imaging cytometry to estimate parameters of gene networks models in systems and synthetic biology. *PLoS ONE*. 2014; 9(9):e107087. <https://doi.org/10.1371/journal.pone.0107087> PMID: 25210731; PubMed Central PMCID: PMC4161401.
36. Ball DA, Marchand J, Poulet M, Baumann WT, Chen KC, Tyson JJ, et al. Oscillatory Dynamics of Cell Cycle Proteins in Single Yeast Cells Analyzed by Imaging Cytometry. *PLoS ONE*. 2011; 6(10):e26272. <https://doi.org/10.1371/journal.pone.0026272> WOS:000296519600012. PMID: 22046265
37. Ball DA, Adames NR, Reischmann N, Barik D, Franck CT, Tyson JJ, et al. Measurement and modeling of transcriptional noise in the cell cycle regulatory network. *Cell cycle (Georgetown, Tex)*. 2013; 12(19):3203–18. <https://doi.org/10.4161/cc.26257> PMID: 24013422; PubMed Central PMCID: PMC3865016.
38. Barik D, Ball DA, Peccoud J, Tyson JJ. A Stochastic Model of the Yeast Cell Cycle Reveals Roles for Feedback Regulation in Limiting Cellular Variability. *PLoS Comput Biol*. 2016; 12(12):e1005230. Epub 2016/12/10. <https://doi.org/10.1371/journal.pcbi.1005230> PMID: 27935947; PubMed Central PMCID: PMC5147779.
39. Goss PJE, Peccoud J. Quantitative modeling of stochastic systems in molecular biology using stochastic Petri nets. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95:6750–5. 1104.
40. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, et al. COPASI—a COMplex PATHway SIMulator. *Bioinformatics*. 2006; 22(24):3067–74. <https://doi.org/10.1093/bioinformatics/btl485> PMID: 17032683
41. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020; 17(3):261–72. Epub 20200203. <https://doi.org/10.1038/s41592-019-0686-2> PMID: 32015543; PubMed Central PMCID: PMC7056644.