Mathematical Modeling and Machine Learning Guided Optimization to Characterize Immunoregulation during Respiratory Infection

by

Jordan John Alexander Weaver

Bachelor of Science in Chemical Engineering, South Dakota School of Mines and Technology, 2017

Submitted to the Graduate Faculty of the

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2022
UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Jordan John Alexander Weaver

It was defended on

July 5, 2022

and approved by

Dr. Christopher E. Wilmer, PhD, Associate Professor, Department of Chemical and Petroleum Engineering

Dr. Ipsita Banerjee, PhD, Associate Professor, Department of Chemical and Petroleum Engineering

Dr. James R. Faeder, PhD, Associate Professor, Department of Computational and Systems Biology

Dissertation Director: Dr. Jason E Shoemaker, PhD, Associate Professor, Department of Chemical and Petroleum Engineering

ii
Copyright © by Jordan John Alexander Weaver

2022
Respiratory viruses present major public health challenges, as evidenced by seasonal influenza’s 290,000 – 650,000 worldwide annual deaths, while the Severe Acute Respiratory Coronavirus 2 (SARS-CoV-2) has caused 6.31 million deaths worldwide. These viruses invoke excessive immune responses; however, the kinetics that regulate inflammatory responses within infected cells remain unresolved. Understanding the dynamics of the innate immune response and its manifestations at the cell and tissue levels is vital to understanding the mechanisms of immunopathology and to developing strain-independent treatments. Computational models of the innate immune response to respiratory infections are designed to provide greater insights into the regulation of the immune system, which will likely provide insights into clinical treatments and the pathological understandings of the disease. Efforts to develop these models have greatly increased as RNA and protein level data have become widely available.

Aim 1 incorporates viral replication, cell death, interferon stimulated genes’ effects on viral replication, and demonstrating that RIG-I is robust to viral antagonism. Aim 2’s model is a spatialized, multicellular representation of RNA virus infection and type-I interferon-mediated antiviral response that model suggests that modifying the activity of signaling molecules in the JAK/STAT pathway or altering the ratio of the diffusion lengths of interferon and virus leads to plaque growth arrest. Aim 3 compares low-pathogenic H1N1 and high-pathogenic H5N1 influenza virus infections, suggesting that the production rate of interferon is the major driver of strain-specific immune responses. This rate difference may arise from the degree of antagonism of RIG-
I by the invading virus. Aim 4 details an unbiased method to determine the minimum number of parameters which must vary to explain differences observed between two or more datasets using an extension of Aim 3.

A greater understanding of the contributors to strain-specific immunodynamics can be utilized in future efforts aimed at treatment development to improve clinical outcomes of high-pathogenic viral strains. As kinetics are host cell-specific, the model presented provides an important step to modeling the intracellular immune dynamics of many RNA viruses, including the viruses responsible for influenza and COVID-19. A visual summary of the work is given in Figure 1.

Figure 1. Visual abstract of work.
# Table of Contents

Preface ........................................................................................................................................ xvii

1.0 Introduction ............................................................................................................................. 1

1.1 Viral Respiratory Infections .......................................................................................... 1

1.2 Mathematical Modeling ................................................................................................. 2

1.2.1 Ordinary and Partial Differential Equations ....................................................3

1.2.2 Agent Based Model ..............................................................................................3

1.2.3 Parameter Exploration and Optimization .........................................................4

1.3 Definition of Terms ......................................................................................................... 5

2.0 Determining How Paracrine Signaling Impacts Epithelial Cells’ Response to Influenza Infection ................................................................................................................... 9

2.1 Introduction .................................................................................................................... 9

2.2 Materials and Methods ................................................................................................ 12

2.2.1 Data Sources .......................................................................................................13

2.2.2 ODE Simulation and Sensitivity Analysis ........................................................13

2.2.3 Model Parameterization ....................................................................................13

2.2.4 Structural Identifiability ...................................................................................14

2.2.5 Interparameter Correlation ..............................................................................15

2.3 Results ............................................................................................................................ 16

2.3.1 MCMC Parameterization..................................................................................16
2.3.2 Model Validation by Predicting Response to Infection Using a NSI Knockout Influenza Virus.........................................................................................................................................................17

2.3.3 Sensitivity Analysis Reveals IRF7 Phosphorylation as Critical Step..........19

2.3.4 Simulating Varying RIG-I Antagonism Reveals Robust Sensor Protein Action .................................................................................................................................................................................21

2.3.5 Sensor Protein and JAK/STAT Originated Interferon Production ...........23

2.4 Conclusion........................................................................................................................................................................................................................................................................................................24

3.0 Identifying the Major Regulators of Viral Plaque Growth ..................................................... 26

3.1 Introduction.................................................................................................................................................................................................26

3.2 Materials and Methods ................................................................................................................................................................................28

3.2.1 ODE Model ................................................................................................................................................................................................................................................................................28

3.2.2 Spatial Considerations of the MSIS model .................................................................................................................................................................................................30

3.2.3 Cell Types and Rationale.................................................................................................................................................................................................31

3.2.4 Plaque Growth Metrics .................................................................................................................................................................................................32

3.3 Results ...................................................................................................................................................................................................................................................................................33

3.3.1 Multicellular Spatial Model of RNA Virus Infection and IFN Signaling (MSIS model) Reproduces ODE Model Dynamics for High MOI infection ..................33

3.3.2 MSIS Model Recapitulates Experimentally Observed Plaque Formation and Growth Dynamics .........................................................................................................................................................................................36

3.3.3 Increased STAT Activity Leads to Arrested Plaque Growth and Reduces Final Plaque Diameter .........................................................................................................................................................................................40

3.3.4 Elevated RIG-I Activity Delays Cell Death and Increases IFN Production 43

3.3.5 Interferon Prestimulation Arrests Plaque Growth ..................................................45
3.3.6 Faster Interferon Diffusion Promotes Plaque Growth Arrest .......................... 47

3.3.7 Sensitivity Analysis Reveals that the Main Parameters Controlling Radial Plaque Growth Differ Between Regimes ............................................................... 50

3.4 Conclusion ............................................................................................................. 53

4.0 Comparing H5N1 and H1N1 infection dynamics in murine hosts ....................... 56

4.1 Introduction .......................................................................................................... 56

4.2 Materials and Methods ........................................................................................ 59

4.2.1 Model development rationale and equations ................................................. 59

4.2.2 Experimental data collected from literature and relating the data to the model .............................................................................................................................. 60

4.2.3 Parameter training ............................................................................................ 61

4.2.4 Model and scenario prioritization .................................................................... 62

4.2.5 Sensitivity analysis ............................................................................................ 63

4.3 Results .................................................................................................................... 64

4.3.1 In silico screenings of candidate innate immune models find that H5N1 and H1N1 viruses induce interferon production at different rates in vivo ...................... 64

4.3.2 Strain-specific interferon production is not an artifact of parameter sensitivity ......................................................................................................................... 68

4.3.3 Exploration of Model 4’s parameter space using PT MCMC ......................... 69

4.3.4 MCMC-based parameter exploration again finds that H5N1 and H1N1 viruses induce interferon production at different rates in vivo .................................. 71

4.3.5 Independent estimation of virus parameters per strain does not improve model AIC ................................................................................................................... 74
4.4 Conclusion ..................................................................................................................... 78

5.0 SPOT ...................................................................................................................................... 81

5.1 Introduction .................................................................................................................. 81

5.2 Materials and Methods .................................................................................................. 82

5.2.1 Methodology .......................................................................................................82

5.2.2 Software ..............................................................................................................83

5.2.3 Data Sources .......................................................................................................83

5.3 Results ............................................................................................................................ 84

5.3.1 Characteristic Curves from Brute Force Evaluation .....................................84

5.3.2 Categorical Linear Regression Highlights Strain – Specific Rates ...............86

5.3.3 Sensitivity of SPOT to Dataset-Specific Parameter Values............................89

5.3.4 SPOT with Reduced Samples............................................................................90

6.0 Discussion and Future Work ............................................................................................... 93

6.1 ODE Model of Interferon During Viral Respiratory Infection......................... 93

6.2 Spatial Effects of Interferon Regulation and Viral Infections ......................... 94

6.3 Strain-Specific Modeling of H5N1 and H1N1 Influenza...................................... 96

6.4 Minimizing Model Size While Fitting Multiple Data Sets with Parameter Sharing .............................................................................................................................................. 96

6.5 Publications Resulting from This Work................................................................. 98

Appendix A Innate Immune Model.................................................................................... 99

Appendix B Multiscale Immune Model.......................................................................... 107

Appendix C Macrophage Models.................................................................................... 132

Appendix D SPOT Supplement ....................................................................................... 145
List of Tables

Table 1. Intracellular chemical concentrations in cells 12 hours after in silico exposure to \([IFN_e]\)............................................................................................................................... 46

Table 2. The minimum energy, degrees of freedom (DOF), and AIC values achieved by Model 4..................................................................................................................................... 71

Table 3. The minimum energy, DOF, and AIC values for all seven Viral subset, \(\{V\}\)........ 75

Table 4. Python3 package versions and applications. ............................................................. 83

Table 5. Best-fit model parameters for the Innate Immune Model. ................................. 102

Table 6. Pearson Correlation Coefficients for ODE parameters found via MCMC. ......... 106

Table 7. Initial conditions for each cell type when present at the start of a simulated infection. ..................................................................................................................................... 112

Table 8. Baseline parameter values and sources....................................................................... 114

Table 9. AD Minimum energy parameter values and units for each dataset in Model 4. . 138
List of Figures

Figure 1. Visual abstract of work. ........................................................................................................... v

Figure 2. Schematic of intracellular innate immune signaling (Weaver model)......................... 12

Figure 3. Innate immune model simulations show excellent fit to biological data. .............. 16

Figure 4. NS1 knockout simulation shows much greater cell survival and sustained cytokine production............................................................................................................................... 18

Figure 5. System Sensitivity Analysis carried out on MCMC results reveals \(k_{42}\) and \(k_{51}\) as the most sensitive parameters. ........................................................................................................ 20

Figure 6. Simulations at varying levels of RIG-I knockdown show robust virus sensing capability........................................................................................................................................ 21

Figure 7. (A) Total, Sensor Protein, and Paracrine \([IFN]\) production simulations in dNS1PR8 influenza. (B) Total, Sensor Protein, and Paracrine \([IFN]\) production simulations in wild-type PR8 influenza. ........................................................................................................ 23

Figure 8. Conceptual diagram of the MSIS model. ................................................................. 29

Figure 9. Comparison of time series for key variables between the multicellular spatial interferon signaling (MSIS) model and Weaver model for high MOI demonstrates vaillidity in similar conditions........................................................................................................ 35

Figure 10. Comparison of an experimental plaque assay for influenza (H5N1; left) with an MSIS model plaque simulation (for H1N1; right). ................................................................. 37

Figure 11. Plaque growth simulations replicate experimentally observed linear radial plaque growth. ........................................................................................................................................ 39
Figure 12. Elevated $[STATP]$ activity (larger $k_{STATP,IFN}$) leads to arrested plaque growth. 42

Figure 13. Increased RIG-I activity ($k_{IFN,V(RIGI)}$) lowers plaque growth rates and viral titers, slows cell death, and increases interferon production. 44

Figure 14. Prestimulating cells with type-I interferon led to plaque growth arrest in simulations. 47

Figure 15. Dependence of plaque growth rate and arrest on viral and $[IFN]$ diffusion coefficients. 49

Figure 16. Local single-factor sensitivity analysis varies greatly for three simulation regimes. 51

Figure 17. 4 Structure Model Ensemble schematic. 60

Figure 18. Energy versus AIC values for all four model structures under different parameterization scenarios. 66

Figure 19. First-order indices of the eFAST sensitivity analysis of the Macrophage models. 68

Figure 20. The top 1,000 fits of Model 4 to the H1N1 (top row) and H5N1 data (bottom row) when using PT MCMC parameterization. 70

Figure 21. Model 4 output for the minimum energy parameter set (lines) for OSSD parameterizations and corresponding training data (markers) for H1N1 (top row) and H5N1 (bottom row). 72

Figure 22. Posterior density distributions for all parameters for Model 4 with $r_{LV}$ varying between strains. 74

Figure 23. Model 4 output for minimum energy parameter set (line) for virus-related parameter independent fits. 77
Figure 24. Models 1 and 4 Characteristic Curves for a SPOT-significant parameter ($r_I,V$) and a non-significant parameter ($r_V,V$)................................................................. 85

Figure 25. SPOT value for each parameter from Aim 3, Model 4 with biological data...... 86

Figure 26. SPOT Values with Synethic Strain-Specific Parameter Differences....................... 89

Figure 27. SPOT values from 0% to 100% of possible genomes for Aim 3, Model 4. ........... 91

Figure 28. Model 4 P-values of SPOT regression versus the fraction of all possible genomes analyzed can determine the significance of SPOT values. ........................................ 92

Figure 29. (A) MCMC Acceptance Rate. (B) Sum Squared Error. Note log scale abscissa. (B), Inset: Linear abscissa scale zoom-in on the 1 million iterations post burn-in. 101

Figure 30. MCMC Parameter Histograms............................................................................. 103

Figure 31. Correlation plots of MCMC parameter exploration.............................................. 105

Figure 32. Simulated [IFNe] exposure before infection protects cells from plaque formation. ......................................................................................................................... 116

Figure 33. Local sensitivity analysis of baseline simulation................................................. 117

Figure 34. Local sensitivity analysis with elevated paracrine signaling. ............................ 119

Figure 35. Local sensitivity with an elevated interferon diffusion coefficient (15x baseline or $540 \mu m^2 s^{-1}$). ................................................................................................................... 121

Figure 36. Standard deviation and standard error of simulations versus replicas for the baseline......................................................................................................................... 122

Figure 37. A parameter sweep of $\beta$ (rate of transition from uninfected (U) to eclipse phase (I1) cells) from 0.01x to 100x baseline reveals steady growth increases. .......... 123

Figure 38. Parameter sweep of viral replication rate ($k_{V,V}$) reveals dramatic changes to final plaque diameter......................................................................................................................... 124
Figure 39. Elevated IFN diffusion coefficient ($D_{IFN_e}$) leads to plaque arrest....................... 126

Figure 40. Figure 13 Plaque Growth Rate with all outliers visible. ................................. 127

Figure 41. Figure 13 Ve AUC with all outliers visible. ...................................................... 128

Figure 42. Figure 38 IFN$c_e$ AUC with all outliers visible. ................................................ 128

Figure 43. Figure 38 Ve AUC with all outliers visible. ...................................................... 129

Figure 44. Figure 39 IFN$c_e$ AUC with all outliers visible. ............................................... 130

Figure 45. Figure 39 Ve AUC with all outliers visible. ...................................................... 130

Figure 46. A) Model schemes of the four models considered in this work. B) Model structure variations. ................................................................. 133

Figure 47. Macrophage and MCP1 correlation. ................................................................. 135

Figure 48. AD and NSSD model 4 energy plots ................................................................. 136

Figure 49. Models 2 and 3 predictions. ............................................................................. 137

Figure 50. Macrophage Model 4 AD and NSSD Parameter Distributions ....................... 139

Figure 51. Macrophage Model 4 with unshared $d_I$. ....................................................... 140

Figure 52. Macrophage Model 4 with unshared $d_M$. ...................................................... 140

Figure 53. Macrophage Model 4 with unshared $d_V$. ...................................................... 141

Figure 54. Macrophage Model 4 with unshared $K_{M,I}$. .................................................. 141

Figure 55. Macrophage Model 4 with unshared $K_{V,I}$. ................................................... 142

Figure 56. Macrophage Model 4 with unshared $n$. ......................................................... 142

Figure 57. Macrophage Model 4 with unshared $r_{M,I}$. .................................................... 143

Figure 58. Macrophage Model 4 with unshared $r_{V,I}$. .................................................... 143

Figure 59. Macrophage Model 4 with unshared $r_{V,V}$. .................................................. 144

Figure 60. Characteristic Curves for Model 1................................................................. 146
Figure 61. Model 4 Characteristic Curves

Figure 62. Model H5N1:H1N1 Parameter Ratio Sweeps

Figure 63. Model 4 H5N1:H1N1 Parameter Ratio Sweeps
Preface

I am grateful for so many things in my life. I want to first acknowledge my fiancé, my emotional support, my chaos, my reason to get out of the office and see the world. Thank you to my family, in all its twisted branches, those I was born into and those that have grown to support me. I’m lucky for the amazing mentors, in undergraduate (Dr. David Dixon) as well as graduate (Dr. Jason E. Shoemaker), and my supportive lab members Dr. Emily Ackerman, Dr. Robert Gregg, Lauren Luciani, Caroline Larkin, and Tatum McGeary. I am eternally grateful for the opportunities I’ve been given.
1.0 Introduction

1.1 Viral Respiratory Infections

Respiratory infections are a constant threat to public health with deadly infections often characterized by cytokine storms, i.e., overly aggressive innate immune responses that result in severe, unnecessary lung tissue inflammation. Both influenza and SARS-CoV-2 are RNA viruses, and the data to date suggests that cytokine storms are a common feature of both viruses during severe infections\(^1\). Typical seasonal influenza virus strains are responsible for 290,000 – 650,000 annual deaths globally\(^2\), and occasional, highly pathogenic pandemic strains, such as the 1918 Spanish Flu\(^3\), 1957 H2N2\(^4\), 1968 H3N2\(^5\), and 2009 H1N1\(^6\) flus result in significantly higher mortality rates. As of May 27th, 2022, the SARS-CoV-2 virus, which causes COVID-19, has caused over 527 million recorded infections and 6.28 million deaths worldwide\(^7\).

Immune responses can help or hinder an organism’s ability to overcome an infection, and excessively, inflammatory responses, like cytokine storms, can cause greater tissue damage, higher mortality, and slow recovery\(^8,9\). In highly pathogenic infections, an aberrant inflammatory response – specifically a prolonged, elevated inflammatory state and a high level of type-I interferons in the bloodstream, clinically called hypercytokinemia (colloquially known as a cytokine storm)\(^1\) – is believed to be a significant driver of mortality\(^10,11\). Excessive inflammation also exacerbates tissue damage and hinders clinical recovery\(^9,12\).

Vaccination is effective for protecting public health against seasonal influenza; however, when new strains unexpectedly emerge, such as the 2009 novel pandemic H1N1 virus or the 2019 SARS-CoV-2 virus, new treatment strategies that can be implemented rapidly, and preferably
independently of the specific virus, are needed. Immunomodulatory treatments that aim to reduce inflammation while still managing virus growth are a promising approach to protecting against emergent disease, but several fundamental questions on how unnecessarily aggressive immune responses emerge remain unknown. Mathematical modeling can help quantify the kinetics of the interactions that define the immune system, revealing the interactions that are most likely to be responsible for unnecessarily aggressive responses and potential targets, to interfere with immunity to ensure healthy virus clearance.

1.2 Mathematical Modeling

Modeling is tool to compress and translate observations, real-world data, and the growing behemoth of established literature into a form tractable to human interpretation. All models are wrong, but if careful and articulate questions are asked, models can reveal useful truths. Data itself is nothing more than an abstraction of real events, given form in numbers, categories, measurements, even natural language. Models simply construct another abstraction – a tool that can predict the same real event as the data it is built upon, given the same conditions. It is not necessarily useful to predict a past event, but the utility of models lies in proposing new conditions and making predictions which have not happened. A model which could predict last week’s lottery numbers is not exciting, but if predictions of next week were correct for even one number it would be advantageous.

This thesis is centered around data from viral respiratory infections, and the models developed here are limited to the same scope. However, the model types, parameter optimization, statistical analyses, and algorithm development are broadly applicable to anywhere that real events
can be made abstract and recorded. This section serves to broadly introduce some of these concepts.

1.2.1 Ordinary and Partial Differential Equations

Ordinary differential equations (ODEs) are equations which model the progression of different states (populations, cytokine levels, income streams, velocity) over time or space. Partial differential equations are like ODEs, but the equations typically allow states to change over both time and space. This type of model is a common approach in systems biology, after their demonstrable success in analyzing the robustness of biological signaling\textsuperscript{13–15}, the highly dynamic behaviors of NF-kB, and ultrasensitive cell fate binary responses\textsuperscript{16}. ODEs allow for interpolation of the dynamics between a finite number of time points at which data has been measured, based on hypotheses of the mechanisms regulating the system’s components. From this, predictions can be made at extended time frames and under novel conditions.

1.2.2 Agent Based Model

Agent based, or multiscale models usually consist of individual agents like flocks of birds or cells in a lung, with a sense of “internal” and “external” modeling scopes. Internal models govern the agents – mathematical equations which could dictate the next movement of the bird within a flock based on its neighbors’ positions, or the amount of inflammation a cell is experiencing after detecting a virus. External models govern the environment, like wind currents and seasonal factors for our migrating birds or the physical process of inflammatory molecules diffusing through mucus to neighboring cells. Agent based models (ABM) can produce rich
temporal and spatial information but require both data types during construction to be defensible. Additionally, the number of decisions, equations, agents, parameters, and dimensions scales the computational and developmental cost of an ABM.

1.2.3 Parameter Exploration and Optimization

Models of any kind rely on parameters. These are completely abstract numbers which describe aspects of the system. A parameter can describe as broad a scope as the rate at which a virus can replicate itself within the cell, or as specific as the rate at which a single nucleotide is bonded with a small step in this process. Parameter values are based on data and estimations alike, and in any realm with noisy or biological data, one must not assign any numerical value too much importance. Parameters are descriptions, unique to the equations in which they act. Finding values for parameters which allow the model to replicate data is an integral step in model development, called parameter optimization. For simple models of < 5 parameters, simply guessing values by hand or in an automated way can be sufficient to find these parameter values. Once the number of parameters increases, both human intuition and the algorithms based on these approaches may start to falter. For larger models, stochastic (random) approaches are used with a numerical quality of the model’s ability to fit the data. More information can be found in each Aim’s Materials and Methods, or broadly within the field of stochastic methods.
1.3 Definition of Terms

Due to the nature of this work, an overview of the vocabulary used is necessitated. Some terms may have unique definitions compared to their typical use in literature: what follows is a best attempt to use unified language within this work.

**Model:** A model is a set of equations or rules which yield time-course or other predictions. A model can be applied to multiple sets of data (or conditions) by having an instance of the model for each dataset. This can consist of Ordinary Different Equations (ODE), Partial Differential Equations (PDE), agent-based (ABM or multiscale), Boolean, network, pharmacokinetic and pharmacodynamic, game engines, or other parameter-dependent functions.

**Energy:** Energy is the value of the objective function value from an optimizer such as Basin Hopping parameterization or Markov Chain Monte Carlo parameter exploration. Smaller energies indicate more suitable parameters and a closer fit to the biological data. Energy does not consider model complexity but is based on the comparison of model predictions to data. Energy typically results from a user-supplied cost function.

**Likelihood:** The likelihood function, L, is the exponential of negative energy, after Metropolis et al.\(^\text{17}\). Lower energies (better fits) give higher likelihoods for a model to be correct. Likelihood functions are the basis of both Basin Hopping’s and Markov Chain Monte Carlo’s stochastic steps.
**Akaike Information Criterion (AIC):** AIC\textsuperscript{18,19} (Equation 1-1) is a penalty function designed to consider both the model fit, as defined by Metropolis-like energy\textsuperscript{17} or statistical error\textsuperscript{20} functions, and model complexity, in this application using Degrees of Freedom (DOF), $k$. Unlike traditional statistical error functions, such as Sum of Squares Error (SSE), lower AIC should indicate a model which is more likely to explain the underlying data while using the simplest model possible. The weighting of error and model complexity is both arbitrary and important; this weighting determines the sacrifices made in the model’s predictive ability for the sake of simplicity. Bayesian Information Criterion (below) leverages this.

$$AIC = 2k - \ln (L)$$  \hspace{1cm} (1-1)

**Bayesian Information Criterion (BIC):** BIC, or Schwarz Information Criterion\textsuperscript{21}, utilizes a different relative weighting of error and complexity (Equation 1-2) and is dependent on the number of underlying data points, $n$. BIC performs a similar role to AIC, although there is ongoing debate between the criterions\textsuperscript{22}.

$$BIC = k \ln(n) - 2 \ln (L)$$  \hspace{1cm} (1-2)

**Parameter:** A parameter is a value which controls various aspects of model equations or outcomes. Parameters can represent rate constants, kinetic orders, probabilities, physical attributes, and similar concepts. Parameters are identified with italics (and subscripts as needed) such as $K_{vv}$. Parameters with values which can reasonably be assumed constant, such as decay rates, can be neglected from SPOT analysis to improve the computational tractability of larger models.

**State:** A model state is a species, equation, or group whose dynamics are represented by the model, such as a chemical concentration or count of predators. States are identified with square brackets and italics, such as $[V]$. 


**Genome:** A genome succinctly identifies which parameters are shared and independent between multiple data sets. Genomes consist of a series of Boolean bits (1 or 0), with a length equal to the number of parameters in the model. 0 indicates a parameter which is shared between all datasets. 1 indicates a parameter which is fit independently to each dataset. The degree of freedom for a genome is thus given in Equation 1-3.

\[
\text{DOF} = \text{Length}(\text{Genome}) + (\text{Number of Datasets} - 1) \times \sum \text{Genome}
\]  

(1-3)

**Model Structure:** A model’s structure is the combination of a genome which defines the parameter sharing rules, and a model, which defines the dynamics, combined in a process referred to as translation. The model structure is parameterized using data and Basin Hopping, resulting in a BIC value and a parameter set. This BIC value, or genome performance, is variable since repeating the stochastic parameterization may result in different parameter sets.

**Dataset:** Datasets are a collection of data, typically consisting of measurements at discrete time intervals from two or more distinct sets of conditions. The data is assumed to share a common underlying model, e.g., identically collected data on H5N1 and H1N1 viral infections *in vitro*. A single model may be capable of recreating both strains’ data, given an appropriate genome to define parameter sharing rules. Datasets can also refer to testing cohorts, treatment groups, age, species, or other grouping variables between which parameter differences are expected. A dataset may be real (coming from observations) or synthetic (computer-generated measurements with intentional differences).
**Basin Hopping (BH):** Basin hopping is a two-phase parameter estimation method, consisting of a stochastic global search method (hops, like Markov Chain Monte Carlo steps, to rapidly traverse parameter space) and a local optimization (basins, a gradient descent method used to rapidly find the minima associated with each hop). Generates a single, maximum likelihood value for each parameter.

**Markov Chain Monte Carlo (MCMC):** A stochastic method which samples parameter space in a probabilistic way. While slower to reach a global minimum than Basin Hopping, MCMC provides a true *exploration* of parameter space, returning distributions of parameters and a characterization of the loss function’s space rather than a single parameter set and BIC value. Based on the Metropolis algorithm\(^\text{17}\).

**Shared Parameter:** A shared parameter has the same value in all instances of equations, regardless of which dataset the equation relates to. It contributes only 1 Degree of Freedom to the model.

**Independent or Unshared Parameter:** An unshared, or independent parameter has a different value in each dataset’s model instance. It contributes one Degree of Freedom for each dataset.

**Degrees of Freedom (DOF):** represents the number of parameters which a model uses to fit to data. Shared parameters contribute 1 DOF, while unshared parameters contribute 1 DOF per dataset present. See Equation 1-3.
2.0 Determining How Paracrine Signaling Impacts Epithelial Cells’ Response to Influenza Infection

2.1 Introduction

The first step to occur during an immune response is the detection of the pathogens, leading to the early, localized, innate immune response\(^{23}\). This response to viral infection leads to the production of Type-I interferon (IFN). Interferon serves to establish an antiviral state by activating and inducing Mx proteins, RNA-activated protein kinase, and the 2-5A system\(^{24}\); they also regulate other immune responses, by acting on natural killer cells, T cells, B cells, dendritic cells, and phagocytic cells\(^{25}\). The presence of influenza virus is primarily sensed by cytoplasmic retinoic acid-inducible gene 1 (RIG-I) and endosomal Toll-like receptors 7 and 9 (TLR\(^{26,27}\)). RIG-I senses viral RNA in the cytoplasm\(^{28}\) but is antagonized by many influenza A viruses’ nonstructural protein I (NS1) to varying, strain-specific magnitudes\(^{29–31}\). TLR7 is free of this antagonism\(^{32,33}\), and is activated after the influenza envelope has been degraded by endosomal proteases. SARS-CoV-1’s N protein has been implicated in the inhibition of Type-I interferon production via antagonism of RIG-I dynamically similar to influenza\(^{31,34,35}\). This suggests RIG-I as a common viral sensor protein and a common target of antagonism for RNA viruses. The activation of either sensor leads to the phosphorylation of interferon regulatory factor 7 (IRF7, IRF7P) and the production of IFN to act as a signaling cytokine. IFN induces secondary messenger molecules, ultimately leading to the induction of immune modulation and antiviral genes\(^{36}\). IFN is secreted from the infected cell and sensed through the Janus kinase/signal transducer and activator of transcription pathway (JAK/STAT), in both an autocrine and paracrine manner. The JAK/STAT
pathway leads to the induction of a broad family of IFN-stimulated antiviral genes\textsuperscript{37}, as well as the supplementary (autocrine) or novel (paracrine) production of IFN. These IFN-stimulated genes (ISGs) cause cell death through apoptosis, necroptosis and pyroptosis\textsuperscript{38}, slow viral replication within the cell\textsuperscript{39}, regulate the infiltration and activity of key innate immune cells to clear the infection, and help initiate the adaptive immune response\textsuperscript{40}. There are at least 12 distinct mechanisms of cell death\textsuperscript{41}. IFN stimulation alone does not seem sufficient to trigger cell death, however, virus and immune cell triggered death mechanisms are enhanced in the presence of IFN and ISGs. While there is significant crosstalk between death mechanisms, two main categories exist: lytic and non-lytic. Lytic mechanisms such as proptosis and necroptosis result in a significant release of cytokines upon cell death. Non-lytic mechanisms such as apoptosis result in the intracellular contents being sequestered upon cell death. Influenza and SARS viruses do not seem to release virions on cell death, while HIV might\textsuperscript{42}.

The current models of innate immune response to RNA virus infection lack major intracellular components or lack important biological interactions, limiting their applicability to understanding how severe inflammation emerges. Extensive molecular pathway maps exist\textsuperscript{23,43}, but they currently lack mathematical description to support simulating the immune response. Some models of the intracellular innate immune response have incorporated RIG-I and TLR activity, but consider their effect to be constant, independent of the viral load, and the models are inherently unstable, complicating their use\textsuperscript{44,45}.

In this study, we construct a novel ODE model to simulate the intracellular innate immune response of human bronchial epithelial cells (HBECs) to influenza A infection and use the model to determine the interactions that most affect cytokine production. This model was constructed with computational expense for parameterization and eventual ABM implementation in mind, with
a minimum number of ODE’s and parameters that capture the dynamics of interest. The model is numerically stable under realistic conditions and non-stiff, enhancing its reproducibility and reducing computational cost. The model incorporates a viral growth model\(^46\) and the proportionality of sensor protein activity to viral RNA levels in the cytoplasm, the first such integration of cell dynamics and viral replication. The feedback of interferon production on viral replication through the interferon stimulated gene (ISG) family\(^47\) is included. A literature search for data to perform parameterization produced viral titers\(^48,49\) and the time-series of RNA data in HBECs\(^50\). RNA data originated from Shapira et al.’s 2009 work elucidating a network of viral-host interactions via genome wide expression profiling\(^50\). Viral titers came from Ramos et al.’s 2013 work on the polyadenylation stimulating factor 30 (CPSF30) binding function of the NS1 protein\(^48\). These consist of subsets, in which competing, parallel pathways were inhibited, allowing for improved identifiability of the model parameters; first, a wild-type A/Puerto Rico/8/1934 Influenza A (PR8) infection, in which RIG-I is assumed fully antagonized and TLR is fully active; and second, an NS1 knockout PR8 strain which has both TLR and RIG-I activity\(^50\). The antagonism of RIG-I in wild-type PR8 infection is shown to drastically alter infection outcomes. Additionally, some parameters were sourced from or bounded by their respective values in previous models. This work establishes the first cell-level model of interferon signaling induced by influenza infection that can be used to compare host responses between infections with different influenza viruses, antagonism motifs and different RNA viruses. Paracrine signaling is demonstrated to produce the majority of HBEC’s cytokine response to influenza infection, while the initial sensor protein pathways are shown to serve as an ignition for said paracrine signaling.
2.2 Materials and Methods

Figure 2. Schematic of intracellular innate immune signaling (Weaver model).

Straight black arrows represent positive interactions while circle-capped black arrows represent antagonism through Interferon Stimulated Genes (ISGs) or Nonstructural Protein 1 (NS1). Cells is normalized from 1 (initial infection) to 0 (complete death of culture). Virus is normalized by dividing by PR8’s peak viral titer at all time points, resulting in a peak of 1 and an initial virus concentration of 6.9E-8 for an MOI of 5. The highlighted Paracrine Signaling pathway is responsible for both cell to cell (paracrine) and cell to self (autocrine) production of IFN. The complete model description and equations are given in Appendix A.1.
2.2.1 Data Sources

Three primary literature sources were used for data to estimate model parameters. First, micro array gene expression data\textsuperscript{50} of two influenza strain (PR8 and an NS1-knockout PR8) time-course experiments in human bronchial epithelial cells (HBECs) were used to fit IFN, STATP, and IRF7 gene expression. Second, viral titers of wild-type PR8 influenza in human lung adenocarcinoma epithelial (A549) cells\textsuperscript{49} and NS1-knockout PR8\textsuperscript{48} were used to fit viral titers. Viral titers would ideally be obtained with the same cell type, time points, and infection methodology as the micro-array data; however, the immune response similarity of A549 cells and HBECs\textsuperscript{51} justifies this approach.

2.2.2 ODE Simulation and Sensitivity Analysis

Julia v1.3 was used to simulate the ODE model with the DifferentialEquations v6.11.0, ParameterizedFunctions v4.2.1, and DiffEqParamEstim v1.12.0 packages. The ODE system is solved with the non-stiff solver VERN7. A Sobol method global sensitivity analysis was conducted to determine the degree of control that each parameter exerted on the system using DiffEqSensitivity v6.7.0.

2.2.3 Model Parameterization

Since the ODE system relies on 15 parameters, simple regression methods are insufficient to successfully parameterize the model. Instead, a stochastic, parallel tempering Markov Chain Monte Carlo method (PT MCMC) was implemented in Julia v1.3. To initialize the parameters, a
literature search and manual fitting methods were conducted. The literature search provided estimated decay rates for STATP\textsuperscript{52}, IRF7\textsuperscript{53}, and IRF7P\textsuperscript{54}. A manual fitting gave estimates and stability-based bounds for cell death ($k_{61}$), viral replication ($k_{71}$), and nonspecific viral clearance ($k_{73}$) rates. Estimates from the Qiao model\textsuperscript{45} were used to initialize the remaining parameters. Since parallel tempering results in faster convergence than single-chain methodologies\textsuperscript{55}, and a thorough exploration of parameter space was desired, a PT MCMC optimization algorithm was run with 1 million iterations with three parallel chains, for a total of three million samples per fitting attempt.

The sum squared error minimized by the MCMC is given in Equation 2-1.

$$SSE = \sum_{i=1}^{S_m} \sum_{j=1}^{T_1} (O_{ij} - N_{ij})^2 + \sum_{k=1}^{T_2} (E_k - V_k)^2$$  \hspace{1cm} (2-1)

The left-hand portion of the objective function determines the error of the system dynamics, where $O_{ij}$ is the experimentally observed log-fold change of Species $i$ at time $j$\textsuperscript{50}, relative to a control RNA level of the same species and time point. Two points from biological replicates are available for each $i,j$ pair. Intracellular interferon, [IFN], and [IRF7] were directly tracked by their corresponding RNA levels. Because phosphorylated STAT cannot be measured using microarray, the RNA levels of interferon-induced GTP-binding Protein Mx1 (MX1), which is induced primarily by the action of STATP\textsuperscript{23}, were used as a proxy for [STATP] concentration. $E_k$ is the normalized literature viral titer estimate at time $k$\textsuperscript{56}, and $V_k$ is the normalized calculated viral titer at time $k$.

### 2.2.4 Structural Identifiability

A common problem in ODE representations of biological systems is a lack of identifiable parameters, leading to non-unique parameter sets\textsuperscript{57}. Models with imperfect structural identifiability must be carefully interpreted, as fixed values for parameters cannot always be
obtained. Structural identifiability analyses were carried out with structural identifiability taken as extended-generalized observability with lie derivatives and Decomposition (STRIKE-GOLLD) in MATLAB R2019a\textsuperscript{58}. These analyses were done under two sets of conditions – perfect identifiability and practical identifiability. Under perfect identifiability conditions, all seven species are assumed to be perfectly observed, i.e., measured directly by experiment. Under practical identifiability, only [IFN], [STATP], [IRF7], [Cells], and [Virus] were observable, which reflects the availability of data under which the model was trained. Full Structural identifiability results are available in Appendix A.2.

2.2.5 Interparameter Correlation

Parameter correlation in MCMC training results was tested using Pearson’s correlation coefficient from SciPy v1.4.1\textsuperscript{59} in Python 3.6.8. Significant correlations were considered as those with a correlation coefficient $> \pm 0.5\textsuperscript{60}$. Correlated parameters indicate mechanisms with interdependent behaviors, such as the parameters of a Hill-like kinetic or the birth and death rates of a population. Parameter correlation results are presented in Appendix A.3.
2.3 Results

2.3.1 MCMC Parameterization

During parameterization, the system was simulated to 48 hours post infection (HPI); only the first 36 hours are shown here for clarity. All accepted parameter fits require the system to be stable, returning to zero for all species after the infection has run its course. In this system, the multiplicity of infection is $5^{50}$, leading to 99.3% of all cells being initially infected $^{61}$. Shapira et al.’s $^{50}$ experiments were carried out in vitro, without any immune cell presence. Thus, the model does not incorporate immune cell presence or phagocytosis, and it was assumed that all cells will die solely because of viral effects. The steady state for the model is complete cell culture death and the eventual decay of all species. The parameter set with the lowest SSE is shown in Figure 3.

![Figure 3. Innate immune model simulations show excellent fit to biological data.](image)
The solid black lines are the trajectories that best fit the training data, as quantified by the MCMC Likelihood. The shaded area is ±1 standard deviation of the best 1,000 parameter sets’ simulations. LFC = Log 2 fold change of gene expression versus control samples with no virus. Broader predictions are seen in species without data present. The model’s stability is ensured as complete population death occurs during this infection. The model can also be interpreted as an intracellular model for a single cell, since this was a high MOI infection without significant spatial effects (Aim 2).

2.3.2 Model Validation by Predicting Response to Infection Using a NS1 Knockout

Influenza Virus

Once the model was parameterized, a validation study on a nonstructural protein 1 (NS1) knockout strain of PR8 influenza (dNS1PR8) was conducted. Since NS1 was assumed to be fully antagonizing the action of the sensor protein RIG-I during the initial training, the parameter \(k_{11}\) associated with RIG-I was unfit and assumed to be zero (unitless). It was manually estimated at 1E5 (unitless) for the dNS1PR8 simulation, based on the same likelihood function as the MCMC parameterization. All other parameters maintained their values from the wild-type PR8 training. An ensemble of the top 1,000 model parameterizations from the wild-type training were simulated with the nonzero \(k_{11}\) term and plotted against the NS1 knockout data. The validation model results are shown in Figure 4.
Figure 4. NS1 knockout simulation shows much greater cell survival and sustained cytokine production.

Solid black lines are the best fitting parameter set. Shaded grey region is ±1 standard deviation of the best 1,000 parameter sets’ simulations (predicted). Dashed lines show wild-type PR8 best-fit from Figure 3 (fit). LFC = Log 2 fold change of gene expression versus control samples with no virus.

The validation case lends credence to the underlying model structure; a different combination of sensor proteins and a modified strain of influenza can be modeled only by introducing a new term for the previously antagonized RIG-I. The assumption of complete and instantaneous RIG-I antagonism utilized for the PR8 parameterization is likely worse than biological reality, but it is useful to establish parameterizations at both extremes of RIG-I activity. Thus, the validation simulation has high RIG-I activity. This new simulation shows several key differences. First, the interferon and extracellular interferon peaks are larger in magnitude and
occur faster than in the wild type, in good agreement with microarray data. Second, the viral load is 96% lower than the wild type, which agrees with the viral titer data. Finally, the simulation predicts that only 20% of infected epithelial cells die, regardless of the simulation’s time frame. While this qualitatively agrees with the much lower lethality of the dNS1PR8 strain, it does not reflect the biological expectation of certain cell death after the infection of all epithelial cells, regardless of viral strain. Overall, the validation study suggests that the model structure is sound and is capable of novel predictions. Further, the PR8 and dNS1PR8 parameterizations of the model may closely represent worst- and best-case viral infections for immune competent Human Bronchial Epithelial Cells.

2.3.3 Sensitivity Analysis Reveals IRF7 Phosphorylation as Critical Step

A Sobol global sensitivity analysis was conducted, allowing each parameter to vary over the same parameter space that the MCMC algorithm explored in the initial training. The resulting system sensitivities are shown in Figure 5.
Figure 5. System Sensitivity Analysis carried out on MCMC results reveals $k_{42}$ and $k_{51}$ as the most sensitive parameters.

White bars represent First Order Sobol sensitivity indices. Black bars represent Total Order Sobol sensitivity indices. Retinoic acid-inducible gene 1 (RIG-I)’s sole parameter, $k_{11}$, is not shown, as it was not fit via the Monte Carlo method.

The most sensitive parameters are those which exert the most control over the innate immune response. Notably, the initial sensing of the virus’ presence via TLR ($k_{12}$) contributes a relatively small proportion of the overall system response; most of the immune response originates from the paracrine signaling pathway. Parameters $k_{42}$ and $k_{51}$ correlate to the phosphorylation of [IRF7] to [IRF7P], and the induction of additional [IRF7] by the action of [IRF7P], respectively, which dominate the paracrine signaling pathway. Thus, the sensitivity analysis suggests that the model’s outcomes are strongly controlled by these two interactions.
2.3.4 Simulating Varying RIG-I Antagonism Reveals Robust Sensor Protein Action

Next, an *in silico* knockdown study of RIG-I was performed. This study is comparable to varying the production and effectiveness of the NS1 protein across several influenza virus strains\(^{29,31}\). This was done by varying the \(k_{11}\) parameter in 25% increments from dNS1-PR8’s lack of antagonism (0% RIG-I knockdown, \(k_{11} = 1E5\)) to wild-type PR8’s total antagonism (100% RIG-I knockdown, \(k_{11} = 0\)). As shown in Figure 6, any RIG-I activity above zero (0% to 75% knockdown) showed significant reductions in viral load and target cell lethality. 0% knockdown yielded the largest magnitude immune response, as quantified by \([IFN_e]\), however, 25% through 75% knockdowns showed a robust immune response of near equal magnitude despite the reduction in RIG-I activity. This suggests that RIG-I is robust against viral antagonism and plays a vital role in initializing the host’s cytokine response to viral infections.

Figure 6. Simulations at varying levels of RIG-I knockdown show robust virus sensing capability.
Percent knockdown, or reduction in value, of $k_{11}$ parameter. Here, 0% knockdown means zero NS1 antagonism, matching the dNS1PR8 strain results, with $k_{11} = 1E5$. Moreover, 100% knockdown is equivalent to total antagonism via NS1, matching the wild-type PR8 results, with $k_{11} = 0$. PR8 and dNS1PR8 (green and black lines) represent minimum and maximum interferon response, respectively. Counter intuitively, the virus which has the highest mortality also results in the lowest inflammation in this model. This can be interpreted as over fitting for the highly inflammatory state, which has an unbounded positive feedback loop through paracrine signaling when cells survive infection. The addition of death mechanisms (Aim 2), removal of target cells entirely and the use of model ensembles (Aim 3), and machine learning guided model ensembles (Aim 4) aim to address the shortcomings of this initial model.
2.3.5 Sensor Protein and JAK/STAT Originated Interferon Production

This model incorporates the production of IFN through both sensor protein action and the paracrine JAK/STAT pathway. These contributions to IFN production were isolated in silico, to determine the relative contribution of each pathway under different conditions. Figure 7 demonstrates these contributions for simulated wild-type PR8 and dNS1PR8 influenza infections. PR8 simulation showed almost complete antagonism of the sensor protein signaling, while dNS1PR8 simulation revealed a dynamic interplay between sensor proteins and paracrine IFN production, based on the infection stage. This analysis suggests that paracrine signaling is the major contributor to IFN production, especially in the presence of RIG-I antagonism, and that TLR activity alone is insufficient to trigger a strong immune response to infection.

![Figure 7](image)

**Figure 7.** (A) Total, Sensor Protein, and Paracrine [IFN] production simulations in dNS1PR8 influenza. (B) Total, Sensor Protein, and Paracrine [IFN] production simulations in wild-type PR8 influenza.

Total, Sensor Protein, and Paracrine [IFN] production in (B) are overlapping. Sensor Protein and Paracrine do not visually sum up to Total since these plots have a log fold change ordinate.
2.4 Conclusion

The data used to inform this model was collected from human bronchial epithelial cell infections, with two influenza strains in vitro. Both strains induce an innate immune response, but the sensor protein RIG-I’s activity is antagonized by only one strain, leading to vastly different interferon production and peak viral load. The model presented here provides estimates of the rates that regulate intracellular responses to RNA virus infection. The model can be used to assess how distinct RNA viruses’ impact IFN production, a key early step in activating the immune system, and is a valuable platform for determining how intracellular immune signaling may be distinctly regulated between influenza and coronaviruses.

A sensitivity analysis of the model revealed that $[IRF7]$ and $[IRF7P]$ have significant control over the innate immune response. This strong single-protein control of the system suggests an area of further experimental investigation; viruses may try to impede these reactions as well as initial sensor protein action, to limit the innate immune response mounted against the invader. Moreover, pre-stimulation of the TLR4 pathway has been shown to lead to an earlier induction of IRF7P production and increased protection from deadly influenza infection. The systems-level analysis here suggests that IRF7 is a potent target for the immune-targeted treatment of severe respiratory infection, both as a means of increasing host immune response and as a target for interference for the mitigation of cytokine storms.

By isolating paracrine and sensor protein originating production of IFN in silico, paracrine signaling is revealed to be responsible for most cytokine production and, thus, immune response. Early and strong sensor protein action serves to ignite this feedback loop in a dNS1PR8 strain. In both PR8 and dNS1PR8 strains, TLR’s activity is too slow and of insufficient magnitude to significantly alter infection trajectory. When active, RIG-I has a profound effect on the peak viral
titer. Effects would be more profoundly distinguished in low multiplicity of infection or in vivo cases, where uninfected neighboring cells would produce IFNs solely because of paracrine signaling. RIG-I is robust against antagonism, with 25 – 100% activity all reducing viral titer by about 97%. Full antagonism via NS1 is necessary for the virus to reach maximal peak. This level of antagonism is not achieved immediately, leaving a window for RIG-I to act before enough NS1 is produced inside the cell and its antagonism sets in. An area of future work is the relaxation of this simplifying assumption and the incorporation of NS1 production, based on viral load and the exploration of different RNA viruses’ antagonism mechanisms.

Developing a minimal ODE model that minimized the computational expense of performing MCMC optimization in high dimensional space necessitated several simplifying assumptions. All type-I interferon species were grouped together in a single $[IFN]$ equation. Several species’ decay rates were estimated from literature$^{52-54}$, rather than being included in the MCMC optimization. RIG-I was assumed to be completely antagonized by the wild-type PR8 strain, which is supported by investigations of the NS1 protein$^{29,31}$. The wild-type simulation is always stable since all simulated cells die. A validation study was performed by predicting a dNS1PR8 strain of influenza, using the same parameter set from training in the wild-type infection. The predictions of the validation case lend support to the model’s capability to capture the interactions of interest without overfitting and suggest that the model can predict responses outside of the training data. This validation case predicts only 20% cell lethality, despite total initial infection, which leads to a mathematically unstable chronic inflammation state. This places limits on model interpretation beyond 24 hours for NS1 knockout, or other reduced severity strains since an infection in vitro is expected to be fully lethal in the absence of immune response and cellular regeneration. However, simulated cell death proportion acts as an indicator for virus severity.
3.0 Identifying the Major Regulators of Viral Plaque Growth

3.1 Introduction

Plaques are visible areas of infected and dead cells that occur in cell cultures infected with a virus. Pre-stimulation of toll-like receptors to induce earlier interferon production protects against highly pathogenic influenza strains in mice\(^6^2\), while cell culture pre-stimulation with type-I interferons prevents viral plaque growth by SARS-CoV (the original 2003 SARS virus)\(^6^3\), SARS-CoV-\(^2\)\(^6^3\), and influenza\(^6^4\). Nebulized interferon α2b and interferon β are being investigated as an early treatment and preventative measure for COVID-19\(^6^5,6^6\). Collectively, these studies demonstrate that immune response regulation must balance tissue damage from inflammatory responses against efficient viral clearance. Computational modeling may reveal how complex responses emerge during infection and aid in identifying immune-targeted treatments.

Recent computational models have considered many aspects of inflammatory responses to viral infection\(^6^7–7^0\). ODE based models assume either homogeneity or a compartment-based quasi-spatial structure and typically ignore the diffusion of virus, local cytokine signaling, heterogeneity of cell responses to stimuli, and stochasticity of individual cells’ responses\(^6^7,7^1\). Recent models\(^6^8,7^2,7^3\) of interferon response to viral infection commonly invoke a generic virally resistant cell type. A cell of this type is either immune to viral infection or stops ongoing viral replication completely. This all-or-nothing response does not capture the dynamics of interferon stimulated genes’ (ISGs’) effects on viral growth\(^7^2,7^3\) or the nuances of partially resistant cells. A spatial model of influenza viral spread and plaque growth\(^7^1\) replicates the linear growth of viral plaques \textit{in vitro} and explores the impact of diffusion coefficients on viral plaque formation but did not incorporate
the cells’ interferon signaling response to the infection. Recent studies of DNA virus infection (Herpes simplex virus 2; HSV-2) used agent-based models to examine the role of adaptive immune cells in restricting plaque growth\textsuperscript{74} while another study found that the degree of stochastic signaling minimized the amount of interferon needed to restrict cell death\textsuperscript{75}. However, HSV-2 is a DNA virus that activates different signaling pathways from RNA viruses, such as influenza, and, as described above, severe respiratory infection often involves strong inflammatory signaling responses. This paper extends these approaches to explain plaque growth arrest due to ISGs for respiratory viral infections.

Plaque growth assays seed the virus at low multiplicity of infection (MOI) and allow it to replicate and form plaques across a monolayer of host cells in cell culture. We developed a multicellular spatial interferon signaling model (which we will call the MSIS model) of the early inflammatory response to RNA viral respiratory infections \textit{in vitro} using CompuCell3D\textsuperscript{76} (CC3D). MSIS simulations can replicate observed plaque growth, cytokine response, and plaque arrest. The MSIS model allows us to determine conditions that lead to either arrested or persistent plaque growth during a simulated infection of a monolayer of lung epithelial cells with an RNA virus. Plaque growth assays are commonly used to compare virus growth rates across cell lines\textsuperscript{71,77}, to quantify the concentration of infective agents\textsuperscript{78,79}, and to observe the effects of drugs and compounds on virus spread\textsuperscript{80–83}. Simulation of \textit{in vitro} experiments \textit{in silico} allows for cheaper, faster, higher-throughput hypothesis generation than experiments. The MSIS model replicates familiar biological plaque growth assays and cell staining experiments, making its simulation methodology and results readily accessible to wet-lab biologists.

Our model focuses on two interacting processes: viral replication and the host cells’ early interferon response. The modeled virus is produced in infected cells, released into the extracellular
environment, and diffuses in this environment. The modeled inflammatory response includes interferon production, export, diffusion and decay, and the induction of virally resistant cell states via ISGs. The model represents a monolayer of immobile human bronchial epithelial cells (HBECs). Each cell contains a separate model of epithelial cell interferon signaling, viral replication and release, and cell death, which is an ODE model\textsuperscript{84} calibrated to data from influenza infected HBECs, that has been modified to include species release or export to the extracellular environment. We adapted a standard model of cell types during viral infection\textsuperscript{85}, with cells transitioning from uninfected, to eclipse phase, virus releasing, and dead cell types. The extracellular environment allows for diffusion of both virus, which leads to the formation of viral plaques, and type-I interferons, which are responsible for paracrine interferon signaling. The MSIS model gives insight into the mechanisms of IFN regulation and the arrest of viral plaques.

### 3.2 Materials and Methods

#### 3.2.1 ODE Model

The MSIS model simulates the replication and spread of an RNA virus infection in a monolayer of epithelial cells and the interferon response induced by the infection. Using CompuCell3D, we created simulations of the MSIS model that represent a diffusive extracellular environment above a square grid of discrete cells, each of which incorporates an ODE representation of epithelial cell interferon production in response to infection by an RNA virus. A schematic of this system is shown in Figure 8.
Figure 8. Conceptual diagram of the MSIS model.

The MSIS model consists of an Intracellular sub-model, which describes intracellular interferon signaling during infection, and a Cellular sub-model, which defines changes in cell types and extracellular molecular diffusion. Uninfected cells (U, blue) produce [IFN] via paracrine signaling alone since no virus is present in these cells. Eclipse-phase cells (I1, yellow) produce [IFN] via viral sensor proteins (RIG-I and TLR7) and paracrine signaling (through the STAT pathway). I1 cells also export [IFN] into the extracellular environment. I1 cells allow virus replication but do not release virus into the extracellular environment. Virus-releasing cells (I2,
red) produce [IFN] in a manner identical to I1 cells and export [IFN] and release virus into the extracellular environment. Dead (D, purple) cells do not interact with their surroundings and have no Intracellular sub-model. Cell type and chemical species colors are conserved throughout the sections of this work dedicated to the MSIS model. Each cell contains an instance of the Intracellular sub-model representing interferon signaling; RIG-I, TLR, [IFN], [IRF7], [IRF7P], and [STATP], viral infection, replication, and release; [V], and cell health, [H]. Type-I interferons, [IFNe], exported by U, I1, and I2 cells, and virus, [Ve], released by I2 cells diffuse and decay in the extracellular environment. Paracrine interferon signaling occurs through the JAK/STAT pathway, indicated by the arrow from [IFNe] to [STATP] across the Intracellular/Cellular border.

3.2.2 Spatial Considerations of the MSIS model

During virus infection, lung epithelial cells produce and export virus and anti-viral type-1 interferon proteins. In cell culture, these extracellular species diffuse freely in the medium above the apical surface of cells.

The conceptual model is that the apical surface of the epithelium interacts with the bottom surface of the medium in which extracellular [IFN], [IFNe], and virus, [Ve], diffuse and decay. We represent the cells and the chemical species in the extracellular medium as a cell lattice next to two chemical field lattices, one for [IFNe] and one for [Ve]. Cells export [IFN] and release virus from their apical surface into the adjacent domain in the chemical field. The CompuCell3D model is a 2D lattice model with the side of each voxel representing 3.0 microns. Unless otherwise specified, the simulation domain is a 300 by 300 lattice, representing a tissue patch of 900 by 900 μm. We represented the layer of epithelial cells using a 100 by 100 array of square cells, each occupying 3 by 3 voxel sites. The cells are infected by [Ve] and respond to [IFNe] in the same adjacent domain.
The area of the cell in the cell lattice represents the interface between the extracellular space and the cell’s apical surface.

Due to the spatial aspect of the model, the concentrations of extracellular species, \([V_e]\) and \([IFN_e]\), can be reported at specific lattice sites, averaged over the area of a cell, or averaged over the enter lattice. \([V_e]\) and \([IFN_e]\) indicate the concentration at a specific lattice site while \([V_e]_{per\ cell}\) and \([IFN_e]_{per\ cell}\) indicate the average concentration over a specific cell for extracellular virus and extracellular \([IFN]\), respectively. The model assumes no spatial variability within the cell.

3.2.3 Cell Types and Rationale

During an RNA virus infection in lung epithelial cells, cells go through four distinct stages. Lung epithelial cells are interferon-competent and produce interferon in response to infection by a virus. During an infection, both infected and healthy cells can respond to changes in extracellular IFN\(^{23}\) via the JAK/STAT pathway. After infection, cells enter an eclipse phase for about 6 hours, during which they produce, but do not release, virus\(^{56,86}\). After the eclipse phase, cells begin to release virus and continue to do so until the cell’s resources are depleted, resulting in death.

To model the four stages of infection a cell experiences, the CC3D-based MSIS model has cells (agents) with 4 distinct types: uninfected (U), eclipse phase (I1), virus releasing (I2), and dead cells (D). Figure 8 provides a conceptual overview of the MSIS model. Uninfected cells, U, contain no virus but can produce and export \([IFN]\) in response to \([IFN_e]\) via the STAT pathway. Paracrine signaling occurs when interferon external to the cell induces the phosphorylation of STAT (\([STATP]\) in Figure 8). U cells transition to the eclipse phase (I1) immediately after a successful infection event. Eclipse-phase (I1) cells can produce and export \([IFN]\), and replicate, but not release virus\(^{87}\). \([IFN_e]\) (via paracrine signaling activation of the JAK/STAT pathway) and
viral sensor protein (RIGI and TLR7) activation both stimulate cells to produce and export [IFN]. When an I1 cell transitions to the virus-releasing type (I2), all properties of the cell remain the same except that the cell can now release the intracellular virus into the extracellular virus field. When an I2 cell transitions to dead (D), it ceases to produce and export [IFN] or release virus but continues to occupy space in the simulation.

3.2.4 Plaque Growth Metrics

Viral plaques are visible areas of dead or damaged cells that occur where a virus has spread across a continuous patch of cells in cell culture. At early times, a growing plaque consists of a central domain of I2 cells surrounded by a ring of I1 cells. At later times, the plaque consists of a domain of dead cells surrounded by a ring of I2 cells, in turn, surrounded by a ring of I1 cells. We measure the radial growth speed of the outer border of the domain of eclipse (I1), virus releasing (I2), and dead (D) cell types. In the simulations, we determine these speeds by seeding a single I1 cell in the center of a simulated sheet of cells and measuring the total area of each cell type over time. We assume the plaques are circular to estimate their radius. The change in the outer radius of the domain of each cell type over time gives the plaque growth velocities. Simulations involve probabilistic infection events and stochastic cell type transitions. We averaged plaque growth metrics over 20 simulations for each parameter set (Figure 36, Appendix B.8). In experiments, plaque-plaque interference occurs when two or more plaques grow into the same spatial region, slowing the radial growth of the colliding plaques. This paper simulates only the growth of isolated plaques.
3.3 Results

3.3.1 Multicellular Spatial Model of RNA Virus Infection and IFN Signaling (MSIS model)

Reproduces ODE Model Dynamics for High MOI infection

We first checked whether the MSIS model reproduced the dynamics of the Weaver model (Aim 1) for the same simulated experimental conditions. The Weaver model was fitted to data from HBECs that were uniformly infected with an influenza virus at MOI = 5. For such high MOI initial conditions, the spatial inhomogeneity of the multiscale model should have a negligible effect on the population-level dynamics, because all cells are infected simultaneously.

The average concentrations of the intracellular species and viral titers of the MSIS model are like those of the Weaver model under high MOI conditions. For MOI = 5, more than 99% of cells are expected to be infected. The Weaver model has two cell types, alive and dead, and does not include eclipse phase cells. To replicate the Weaver model simulations for an MOI = 5 infection, we initialized the MSIS model with only virus releasing (I2) cells and no eclipse phase (I1) cells. A non-uniform cell type distribution (Figure 9 A) and local $[IFN_e]$ concentration field (Figure 9 B) emerge in the MSIS model simulations due to the stochastic cell transitions, which lead to spatially varying $[IFN_e]$ and $[V_e]$, which in turn lead to non-uniform rates of death of I2 cells.

Figure 9 C compares the average fraction of live cells and average levels of chemical species in the MSIS and Weaver models. The fraction of live cells vs time has the same shape in the two models, but dead cells start accumulating slightly later in the MSIS model than in the Weaver model. A major distinction between the MSIS model and the Weaver model is that MSIS cells are discrete. Dead cells have no intracellular chemical species and do not release virus or
export $[IFN]$. These distinctions mean that we must compare the levels of intracellular chemical species (or Health) in live cells (i2 cells in this case) to the Weaver model outputs. However, the levels of extracellular species reflect production by all cells over time and thus we compare the $[IFNe]$ averaged over all lattice sites to the Weaver model outputs. For homogeneous, high MOI starting conditions, all concentrations grow rapidly after the onset of viral release, reach a maximum, and then decay nearly exponentially on a slower time scale. For each variable, the MSIS model value is always greater than or equal to the Weaver model value. Relative errors are largest at times when the values are near their maxima and are always less than 15%. Cell death begins slightly later in the MSIS model than in the Weaver model and the cell death rate increases slightly faster, so that all cells die at nearly the same time. Since the MSIS model produces dynamic responses like those of the Weaver model under high MOI, we will assume that differences between the dynamics of the two models at low MOI result from spatial effects, not from differences in parameters or errors in spatializing the Weaver model. This paper focuses on simulated spatially heterogeneous low MOI initial conditions, which more closely resemble *in vitro* plaque growth assays than high MOI.
A. Cell Types

B. Extracellular Interferon Field

C. Numerical Results

Figure 9. Comparison of time series for key variables between the multicellular spatial interferon signaling (MSIS) model and Weaver model for high MOI demonstrates validity in similar conditions.
All cells are initially infected with 6.9E-8 (unitless) virus, matching the original data to which the Weaver model was fit\textsuperscript{50}. (A) Snapshots of the cell field showing cell type (virus releasing [I2] in red and dead [D] in purple) at different times in a representative MSIS simulation. (B) Snapshots of the concentrations of extracellular interferon (high concentrations in red, low concentrations in blue) at different times in a representative MSIS simulation. (C) Time series for key variables for the Weaver and MSIS models. MSIS simulations are averaged over 20 replicas at matching times (Figure 36, Appendix B.8). Error bars are included but are too small to be visible. For the MSIS model, average concentrations for intracellular species and Health are calculated over all live (I2) cells at each time point while $[IFN_e]$ is averaged across the entire simulation domain.

3.3.2 MSIS Model Recapitulates Experimentally Observed Plaque Formation and Growth Dynamics

High MOI experiments are useful for determining the time course of viral titer and how long cells survive a viral infection, but, unlike plaque assays, they do not provide information about viral spread and the spatial aspects of cytokine responses. We explored low MOI plaque assay experiments \textit{in silico}. Figure 10 (left) shows multiple plaques that formed in a culture of cells infected with an H5N1 influenza virus. We first evaluated if the MSIS model produced plaque-like structures beginning with a single point of infection, like those in experiments for low MOI. We created a simulation with two I1 cells seeded in similar locations to a subset of the plaques shown in Figure 10’s left image. Figure 10 right shows the simulation at 80 hours. The MSIS model reproduces the circular geometry of experimental plaques. The length scales differ
between the experimental and simulated plaques because the MSIS model is parameterized for an H1N1 virus, while the experiment shown used a faster replicating H5N1 virus.

Figure 10. Comparison of an experimental plaque assay for influenza (H5N1; left) with an MSIS model plaque simulation (for H1N1; right).

The simulation seeded two plaques in a simulation domain to replicate a subset of the experimental area. The simulated plaques have a similar structure to the experiment. Outlined area in the experimental image corresponds roughly to the area of the simulation domain.

Next, we explored plaque growth dynamics in the MSIS model. Figures 11 A, B show experimental plaque radii vs time (data reproduced from 71). While the increase in viral load during infection is typically exponential, plaque radius grows linearly in time. The experiment measured the radius of the outer edge of the domain of dead cells (equivalent to D in the model) and the outer edge of the domain of infected cells (equivalent to I1 in the model). The MSIS model distinguishes the eclipse phase (I1) from virus releasing (I2) cells, which normally cannot be distinguished in experimental plaque growth assays. For simulations beginning with a single I1 cell, Figure 11 C shows that the MSIS model replicates several experimental observations. Both experiments (Figure 11 A, B) and the MSIS model (Figure 11 C) show a lag phase with no plaque
growth. During the lag phase, the cells are not releasing virus and no new cells are being infected. Figure 11 E shows snapshots of the cell types at 17 hours, 33 hours, 50 hours, and 67 hours in a single replica simulation. The plaque consists of a central nearly round disk of dead cells, surrounded by a concentric ring of I2 cells, in turn, surrounded by a concentric ring of I1 cells. A few I1 and I2 cells are scattered in the dead cell disk, dead and I1 cells in the I2 ring, and I2 cells in the I1 ring. Figure 11 D shows snapshots of the \([V_e]\) field at corresponding times in the same simulation replica. The virus concentration is maximal over the ring of I2 cells and decreases rapidly at larger and smaller radii. Figure 11 F shows the \([IFN_e]\) concentration field at corresponding times for the same simulation replica. Figure 11 F shows that the \([IFN_e]\) level is high in a very narrow ring over the boundary between the I1 and I2 rings in the plaque. Figure 11 G shows the cell type composition of the culture over time. Dead cells first appear after 20 hours, after which the radius of plaque’s central, circular domain of dead cells increases linearly in time. The radial growth rate of the plaque remains constant until the plaque reaches the edge of the simulation domain. Around 18 hours post-infection, \([V_e]\) (Figure 11 H) and \([IFN_e]\) in the culture (Figure 11 I) decrease briefly because the initially infected cell has died and stopped releasing virus and exporting \([IFN]\). During this time, the second generation of infected cells (those infected by the virus released by the initially infected cell) are primarily I1 phase and not yet releasing virus.
Figure 11. Plaque growth simulations replicate experimentally observed linear radial plaque growth.

(A, B) Radius vs time of outer boundaries of the domains of infected and dead cells for wild-type and H275Y mutant A/Miss/3/2001 (H1N1) infection-induced plaques, respectively. Data reconstructed from 71. Squares indicate the radius of the outer edge of the plaque (the boundary between infected cells and uninfected cells) and circles indicate the radius of the boundary between
dead cells and infected cells in the plaque. Dotted lines show a linear regression for visualization of plaque radius vs time. (C) Simulated plaque growth shows the lag phase and linear growth of the experimental plaques. The solid line indicates the median for 20 simulation replicas (Figure 36 includes additional information on the simulations standard deviations) and the shaded areas indicate the 5th and 95th percentiles of observed values. Panels D, E, and F show sequential snapshots (at 17, 33, 50 and 67 hours) of the \([V_e]\) field (D), cell type (E), and \([IFN_e]\) field for a single simulation replica of a growing plaque. Time progresses from left to right. Panels G, H, and I show the median (solid line) and 5th to 95th percentile (shaded areas) of the simulated cell types, average \([V_e]\), and \([IFN_e]\), respectively, calculated for an ensemble of 20 simulation replicas (Figure 36).

The MSIS model recapitulates the experiments’ linear radial growth of viral plaques. The MSIS model’s ability to simulate both high and low MOI experiments and reproduce phenomena seen experimentally, without additional parameter fitting to these conditions, gives confidence in its predictive capabilities in novel circumstances. The next five sections of Results are based on simulations of plaque growth assays, which give insights into the significance of spatial inhomogeneity to the mechanisms regulating plaque growth.

### 3.3.3 Increased STAT Activity Leads to Arrested Plaque Growth and Reduces Final Plaque Diameter

The JAK/STAT pathway triggers an inflammatory reaction via auto/paracrine signaling and inhibition of this pathway has been implicated in improved H1N1 influenza survival in mice. We wished to assess the impact of JAK/STAT activity on plaque growth dynamics in the MSIS model. We simulated plaque growth while altering the ability of extracellular interferons to
activate the JAK/STAT pathway in the MSIS model from its baseline value, 45.9 μM hours$^{-1}$, up to 125.89x this value. For three values of $k_{\text{STATP,IFNe}}$, we show the plaque size and shape at 80 hours post-infection (Figure 12 A) and the cell type dynamics over time (Figure 12 B). The baseline value leads to unconstrained plaque growth. Values of $k_{\text{STATP,IFNe}} \geq 459.22$ μM hours$^{-1}$ (10x baseline value) led to the arrest of plaque growth. Increasing $k_{\text{STATP,IFNe}} \geq 4592.2$ μM hours$^{-1}$ (100x baseline value) reduces the time to plaque growth arrest, resulting in smaller plaques. These simulations use the same initial conditions and parameters except for the modified values of $k_{\text{STATP,IFNe}}$. Increasing the degree to which $[\text{IFNe}]$ promotes $[\text{STATP}]$ production arrests plaque growth and reduces the final plaque size.

Figure 12 C shows the rate of change of plaque radius at the end of the simulation as a function of $k_{\text{STATP,IFNe}}$, which controls the degree to which a given level of $[\text{IFNe}]$ leads to active $[\text{STATP}]$. For $k_{\text{STATP,IFNe}}$ multipliers of 15.85 and above, the plaque growth rate is always zero at the end of the simulation, indicating plaque arrest. Arrest occurs earlier for higher $k_{\text{STATP,IFNe}}$ (Figure 12 B, 2nd 3rd panels). $k_{\text{STATP,IFNe}}$ multipliers above 6.31 reduce the area under the curve (AUC) for average $[V_e]$ (Figure 12 D), while multipliers between 1.0 and 6.31 have little to no effect on viral AUC. The AUC of average $[\text{IFNe}]$ (Figure 12 E) increases with increasing $k_{\text{STATP,IFNe}}$, with a dramatic increase in the range of multipliers of 6.31 to 10.0. Note logarithmic ordinate scale for both average $[V_e]$ and $[\text{IFNe}]$ AUC. Larger $k_{\text{STATP,IFNe}}$ would correspond to a stronger interferon response and reduced viral titer. $k_{\text{STATP,IFNe}} \geq 4592.2$ μM hours$^{-1}$ leads to non-physiological unbounded production of $[\text{IFN}]$, due to the lack of an interferon-mediated cell death mechanism in both the Weaver and MSIS models.
Figure 12. Elevated [$STATP$] activity (larger $k_{STATP,IFNe}$) leads to arrested plaque growth.

(A) Images of the simulated plaques at 80 hours post-infection for a single simulation replica when $k_{STATP,IFNe}$ was 1x, 10x, or 100x larger than its baseline value. Arrested plaque growth occurs when $k_{STATP,IFNe}$ is 10x or 100x larger than baseline. (B) The median (solid line) and 5th and 95th percentiles (shaded regions) for 20 simulation replicas of the cell types over time for $k_{STATP,IFNe}$ at 1x, 10x, or 100x larger than its baseline value of 45.9 μM hours⁻¹. (C) The plaque radius’ linear growth rate at 80 hours, (D) the area under the curve (AUC) of the average $V_e$, and (C) the AUC of the average [$IFNe$] when $k_{STATP,IFNe}$ is changed between its nominal value to 125.98x nominal.
3.3.4 Elevated RIG-I Activity Delays Cell Death and Increases IFN Production

In influenza infection, greater viral inhibition of RIG-I signaling via NS1 protein often increases viral infection severity\cite{29,31,89}. We wished to investigate the effects of decreasing this antagonistic strength on plaque growth dynamics in silico, now with spatial considerations, after the demonstrated importance of this mechanism in Aim 1. In our simulations $k_{\text{IFN},V(RIGI)}$ controls the strength of the RIG-I response, with larger values corresponding to a stronger response (more IFN produced per unit of virus). Our simulations so far assumed that the invading virus completely inhibited the RIG-I pathway ($k_{\text{IFN},V(RIGI)} = 0$, Aim 1 model). Previous work used data from cells infected with an NS1-knockout influenza virus (A/Puerto Rico/8/1934 [dNS1PR8]) to estimate the rate of IFN production via RIG-I virus sensing ($k_{\text{IFN},V(RIGI)} = 1E5 \, \mu\text{M h}^{-1}$)\cite{84}. We ran single-plaque growth simulations for 14 values of $k_{\text{IFN},V(RIGI)}$ between 0% and 100% of this estimate. These simulations use the same initial conditions and parameters except for the value of $k_{\text{IFN},V(RIGI)}$.

At 80 hours post-infection (Figure 13 A) the plaque radius is nearly the same for all cases, shown for 0%, 50%, and 100% activity. However, the cell type composition of the plaque (Figure 13 B) differs significantly, with significantly less cell death and thus a higher fraction of I2 cells, for $k_{\text{IFN},V(RIGI)}$ multipliers greater than 50%. Higher levels of RIG-I signaling (larger values of $k_{\text{IFN},V(RIGI)}$) only slightly reduce the radial plaque growth at the end of the simulations (Figure 13 C). The AUC of the average $[V_e]$ decreases steadily with increasing RIG-I activity (Figure 13 D), decreasing more rapidly for $k_{\text{IFN},V(RIGI)}$ multipliers greater than 25%. The AUC of average $[IFN_e]$ increases dramatically for parameter multipliers less than 0.03x nominal and more gradually thereafter (Figure 13 E).
Figure 13. Increased RIG-I activity ($k_{IFN,V(RIGI)}$) lowers plaque growth rates and viral titers, slows cell death, and increases interferon production.

(A) Images of plaques at 80 hours post-infection for a representative simulation replica for three values of $k_{IFN,V(RIGI)}$ (0, 0.5E5 μM hours$^{-1}$ and 1E5 μM hours$^{-1}$) and (B) the median (solid line) and 5th and 95th percentiles (shaded regions) of the plaque radius over time for 20 simulation replicas for $k_{IFN,V(RIGI)}$ equal to 0x, 0.5x, or 1x its nominal value of 1E5 μM hours$^{-1}$. (C) The plaque growth rate at 80 hours, (D) the area under the curve (AUC) of the average $V_e$, and (E) the AUC of the average $IFN_e$ for different values of $k_{IFN,V(RIGI)}$. Full data with 5 additional outliers for the plaque growth rate and the average $[V_e]$ AUC are available Figures 40 and 41, respectively.
Increasing levels of RIG-I activity for a given level of virus (larger $k_{IFN,V(RIGI)}$) increases the intracellular production of $[IFN]$. This higher intracellular $[IFN]$ leads to higher $[IFNe]$ due to cell export. Higher $[IFNe]$ leads to a reduction of intracellular viral levels. Since the rate of decrease of cell health, $[H]$, is linear with respect to intracellular virus level and the death rate of cells is proportional to both the virus level and $[H]$, higher values of $k_{IFN,V(RIGI)}$ increase the survival time of infected cells both by decreasing the intracellular virus level and by slowing the decrease of $[H]$. Overall, the model predictions are consistent with the expectations that greater RIG-I activity leads to reduced virus production, i.e., reduced virus titers.

### 3.3.5 Interferon Prestimulation Arrests Plaque Growth

In experiments, prestimulation of cell cultures with type-I interferons reduces the amount of virus produced in cells infected with SARS-CoV, SARS-CoV-2, or influenza. We simulated prestimulation experimental conditions in the MSIS model to explore these protective effects by exposing uninfected (U) cells to $[IFNe]$ at 0.04 μM at 12 hours pre-infection (-12 hours, since infection is referenced as time = 0), using the values of the parameters in Appendix B. All cells were exposed to the same concentration of $[IFNe]$. Since cell type transitions do not occur in the absence of virus, after 12 hours, all cells had identical intracellular chemical concentrations shown in Table 1. At 0 hours, $[IFNe]$ is set to zero to simulate washing $[IFNe]$ out of the cell culture, and a single cell is infected in silico by setting it to the I1 type. We then assessed the impact of $[IFN]$ prestimulation on plaque growth.
Table 1. Intracellular chemical concentrations in cells 12 hours after in silico exposure to 

<table>
<thead>
<tr>
<th>Species</th>
<th>Initial Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>[IFN]</td>
<td>0.035 μM</td>
</tr>
<tr>
<td>[IRF7]</td>
<td>0.097 μM</td>
</tr>
<tr>
<td>[IRF7P]</td>
<td>0.028 μM</td>
</tr>
<tr>
<td>[STATP]</td>
<td>0.714 μM</td>
</tr>
</tbody>
</table>

Simulated prestimulation entirely arrests plaque growth after 35 hours (Figure 14 A), while the same initial infection in a field of naïve, unstimulated cells resulted in the infection and eventual death of all simulated cells (Figure 14 C, G). Only the initially infected cell dies. The proportion of eclipse phase (I1) cells steadily decreases after 20 hours, indicating a cessation of new infections (Figure 14 B). The average $[V_e]$ concentration (Figure 14 C) also decreases after 20 hours. The average $[IFNe]$ concentration (Figure 14 D) is higher than in the baseline simulation. Time series for the intracellular variables, akin to Figure 14 C, are available in Appendix B.6, Figure 32.
Figure 14. Prestimulating cells with type-I interferon led to plaque growth arrest in simulations.

We simulate an experiment with 0.04 μM $[IFN_e]$ prestimulation for 12 hours, which is removed immediately before infection. (A) Sequential snapshots (at 10-, 15-, 25- and 35-hours post-infection) of plaques for a representative simulation replica. (B) Cell type fractions vs time. (C) Average $[V_e]$ vs time and (D) the average $[IFN_e]$ vs time. The solid lines indicate medians and shaded areas represent the 5th and 95th percentiles over 20 replicas.

3.3.6 Faster Interferon Diffusion Promotes Plaque Growth Arrest

Diffusion coefficients for the virus and interferon will depend on virion diameter and the viscosity and chemistry of the medium in vitro. We varied virus and interferon diffusion coefficients simultaneously. Because the actual diffusion coefficient of $[IFN_e]$ is likely to be 11x
to 17x greater than that of the virus, we varied the interferon diffusion coefficient from 54 μm² s⁻¹ to 2160 μm² s⁻¹ (1x to 40x the baseline interferon diffusion coefficient) and the virus diffusion coefficient from 54 μm² s⁻¹ to 216 μm² s⁻¹ (1x to 4x the baseline virus diffusion coefficient). Simulations used the same initial conditions and parameters except for the revised diffusion coefficients. We calculated the median growth rate of the plaque radius at the end of the simulation over 20 replicas. If the median linear growth rate was 0 at the end of the simulation, we classified the parameters as leading to plaque arrest (orange); otherwise, we classified the parameters as leading to continued growth (blue). In our simulations, an interferon diffusion coefficient of 8x to 10x the viral diffusion coefficient led to plaque growth arrest (Figure 15). The curved boundary between the domains suggests that for high viral diffusion coefficients, virus diffusion ceases to be the rate-limiting factor in plaque growth. In summary, there is a broad range of values for both diffusion coefficients in which plaque arrest and continuous growth may occur. Better estimates of these diffusion coefficients can help clarify the relative importance of intracellular versus extracellular processes in viral infection.
Figure 15. Dependence of plaque growth rate and arrest on viral and $[IFN]$ diffusion coefficients.

Each box shows 20 replica simulations’ cell type progression over time for the indicated diffusion coefficient multiplier combination. The solid-colored lines indicate the medians of the radii, and the shaded regions indicate the 5th and 95th percentile radii over 20 replicas. In the orange
shaded region (above the bold line) plaques arrest by 80 hours. In the blue shaded region (below the bold line), plaques continue to grow until the end of the simulation.

3.3.7 Sensitivity Analysis Reveals that the Main Parameters Controlling Radial Plaque Growth Differ Between Regimes

To determine how individual parameters affect the growth of plaques, we performed local sensitivity analyses around parameter sets in three distinct regimes in parameter space; the baseline parameter set (Appendix B.5), the High JAK/STAT regime \((k_{\text{STATP,IFNe}} = 688.5 \, \text{μM hours}^{-1}, 15\times \text{baseline value})\), and the High IFN Diffusion regime \((D_{\text{IFNe}} = 540.0 \, \text{μm}^2 \, \text{s}^{-1}, 10\times \text{baseline value})\). For each regime, we ran 20 simulation replicas using the regime’s nominal parameter values. Then, we perturbed each parameter individually ±25%, ran 20 simulation replicas for each perturbed parameter set, and performed statistical analyses on several sensitivity metrics derived from the simulated trajectories. Sensitivity metrics include the percent change from the average of the baseline simulations of the plaque radius growth rate, the maximum value of \([V_e]\) and \([IFNe]\) that occurred over time, and the AUC of average \([V_e]\) and \([IFNe]\). We determined the statistical significance of the change in each metric from its unperturbed value using a student’s t-test. Statistical test results and sensitivity metrics are reported in the Appendix B.7 (Figures 33 – 35). Increasing and decreasing the parameter values primarily led to directionally consistent changes in the sensitivity metrics, e.g., if the metric increased when the parameter increased by 25%, then the metric also decreased when the parameter decreased by 25%. The top row of Figure 16 shows the cell type progression and variability for plaque growth assays for each regime.
Figure 16. Local single-factor sensitivity analysis varies greatly for three simulation regimes.

(A) ‘Baseline’ corresponds to the baseline parameters in Appendix B.5. (B) ‘High JAK/STAT’ corresponds to a 15x increase in the phosphorylation rate of STAT to $[STATP]$ via the JAK/STAT pathway (parameter $k_{STATP,IFNe}$). (C) ‘High IFN Diffusion’ corresponds to a 10x increase in the diffusion coefficient of $[IFNe]$. Sensitivity analyses varied each parameter one-at-a-time ± 25% around its unperturbed value and quantified the average plaque radius growth rate.
at the end of the simulation, the maximum extracellular virus, $[V_e]$, and interferon, $[IFN_e]$, levels that occurred, and the area under the curve (AUC) for both average $[V_e]$ and $[IFN_e]$. The sensitivity metrics average the absolute values of the metric for increased and decreased parameters over 20 replicas for each parameter set.

Previous sections demonstrated that variation in multiple parameters could lead to either continuous or arrested plaque growth. The baseline parameter set leads to the continuous growth of the plaque. In this regime, the rate of $[STATP]$ dephosphorylation (determined by $\tau_{STATP}$, Appendix B.2), the strength of induction of $[IRF7]$ by $[STATP]$ and $[IRF7P]$ (determined by the $k_{IRF7,STATP}$ and $k_{IRF7,IRF7P}$ rate parameters in Appendix B.2), the maximal rate of viral replication (represented by the rate parameter $k_{V,V}$), the extracellular virus diffusion coefficient ($D_{Ve}$), and the rate of nonspecific extracellular viral clearance ($\tau_{Ve}$) have the largest effects on the metrics. For example, increasing the maximal viral replication rate ($k_{V,V}$, effect shown in Appendix B.9, Figure 38) or the extracellular virus diffusion coefficient ($D_{Ve}$, effect shown in Appendix B.9, Figure 39) leads to faster plaque growth, whereas increasing the virus release rate to the extracellular environment ($Q_V$) would slow plaque growth. The High JAK/STAT and High IFN Diffusion regimes both have arrested plaque growth. In these regimes, the parameters associated with the activation of paracrine signaling have statistically significant sensitivity to perturbations tested. The magnitude of the effects is higher in the High JAK/STAT regime than in the High IFN Diffusion regime. This, combined with the difficulty in selective diffusion length modification, suggests that paracrine signaling is a more feasible target for immunomodulation. The increase in parameter sensitivity in arrested plaque growth regimes also suggests that experimental conditions leading to arrested growth could improve the parameterization of future models and investigations into the interferon signaling response to viral infection.
3.4 Conclusion

The MSIS model produced plaque-like structures (Figure 10). The MSIS model includes parameters that were fit to data from H1N1-infected cell culture experiments, but several parameters estimate also come from the literature (see Appendix B.5). Without additional parameter training, we showed that the model produced plaque growth dynamics (Figure 11 C) like those observed in cells infected with two different H1N1 influenza viruses (Figure 11 A, B). We then focused on using the MSIS model to evaluate how altering intracellular signaling rates and/or diffusion rates might impact plaque growth and performed sensitivity analyses to determine the experimental conditions under which the model’s parameter values can best be estimated.

One of the most significant outcomes of this study is that the sensitivity analysis of the MSIS model suggests that experiments should be performed in conditions that lead to plaque growth arrest rather than unlimited growth to improve the identifiability of interferon signaling parameters (Figure 16). Often, cell culture experiments of virus growth dynamics employ cell lines or conditions that promote virus plaque growth. For example, Vero cells are frequently used in studies because they do not produce interferon and therefore support robust virus replication. However, our sensitivity analysis shows that performing experiments in cells with more robust IFN responses will provide more informative data to estimate 19 of the interferon signaling parameters, compared to only 9 parameters being significantly sensitive in regimes leading to unconstrained plaque growth. And 8 of these 9 parameters significantly affect the model outputs under both constrained and unconstrained plaque growth. In all, the model suggests that experiments performed in IFN-competent cells under conditions that lead to plaque arrest are best for accurately inferring interferon signaling-associated parameter values.
We also used the MSIS model to evaluate the effects of increased paracrine activity via \([STATP]\), increased intracellular virus detection via RIG-I, and prestimulating cells with \([IFN_c]\). All three changes lead to increased concentrations of \([IFN_c]\) but only elevated paracrine signaling resulting from enhancing \([STATP]\) activation (Figure 12) and interferon prestimulation (Figure 14) led to plaque growth arrest. Both \([IFN]\) prestimulation and enhanced \([STATP]\) production via \([IFN_c]\) resulted in a reduced concentration of extracellular virus and an increased concentration of \([IFN_c]\). Enhanced \([STATP]\) activation and \([IFN]\) prestimulation leading to suppressed virus production and plaque growth are consistent with known biology and experimental observations\(^{140,141}\). However, the model’s predictions on the effects of enhancing intracellular detection of virus via the RIG-I pathway differs significantly from experimental observations. Experiments show that enhanced RIG-I binding of viral RNA leads to increased IFN production, reduced virus production, and smaller plaques\(^{142}\). Increasing RIG-I activity \textit{in silico} increased \([IFN]\) production and decreased virus production but did not significantly change the plaque size at the end of the simulation (Figure 13 A). It did affect the cell type demographics, leading to significantly fewer dead cells and many persistent I2 cells (Figure 13 B). Future work could investigate the effect of intracellular IFN and viral load on the rate of cell death, which is independent of these factors in the MSIS model.

We then considered how diffusion coefficients impact plaque growth (Section 3.3.6). A Stokes-Einstein estimate of diffusion coefficients for virus particles (with an effective radius of 80 – 120 nm)\(^{143}\) and interferon proteins (with an effective radius of 7 nm)\(^{144}\) predict 11x – 17x larger diffusion coefficients for interferon in most media. While diffusion coefficients can vary over several orders of magnitude during a single cell culture experiment (due to cell secretion of molecular species like collagen which increase medium viscosity or proteases which can decrease...
it)\textsuperscript{71}, we used a constant, equal, diffusion coefficient for both species (54.0 μm\(^2\) s\(^{-1}\)) in our baseline simulations, resulting in continuous radial plaque growth. The decay rates (\(\tau_{Ve}\) and \(\tau_{IFNe}\)) yield effective diffusion lengths for \([Ve]\) and \([IFNe]\) of 0.09 μm and 0.23 μm, respectively. In Figure 15, we explored how changing the diffusion coefficients impacts plaque growth, identifying a clear boundary between regimes of arrested and continuous plaque growth. Figure 15 shows that even when \(D_{IFNe}\) is significantly larger than \(D_{Ve}\) both arrested and unconstrained plaque growth can occur for different values of \(D_{IFNe}\). In summary, we show that the model can produce unconstrained and constrained plaque growth over a wide range of diffusion coefficient combinations. Future work will focus on refining these values and may consider time-dependent diffusion coefficients.
4.0 Comparing H5N1 and H1N1 infection dynamics in murine hosts

4.1 Introduction

Infections with different influenza A viruses reveal distinct trends in the observed timing and magnitude of immune system dynamics, which correlate to the severity of clinical outcomes\(^{90}\). Occasionally, high pathogenic subtypes emerge, which can result in deadly, worldwide pandemics such as the 1918 Spanish Flu and 1968 Flu pandemics. Of particular concern is the threat that avian H5N1 influenza viruses pose to public health\(^{91}\). An estimated 60% of human H5N1 infections end in death, the majority of which unexpectedly occur in those under 65\(^{92}\). Infections with H5N1 viruses are characterized by higher viral loads, longer viral clearance times, and increased levels of inflammation and tissue damage in comparison to low pathogenic influenza viruses\(^{93}\).

While it remains unclear how H5N1 and other highly pathogenic viruses induce a more severe inflammatory response, there are several potential explanations. One possibility is that H5N1 viruses replicate more quickly, and that observed differences in the immune response are driven primarily by the viral replication rate\(^{94}\). Another possibility is that H5N1 viruses may antagonize the immune system differently during the early stages of infection. A specific candidate mechanism involves the influenza virus’ nonstructural protein 1 (NS1). NS1 is well-established as an antagonist of intracellular immune signaling through the inhibition of retinoic acid-inducible gene I (RIG-I) activity, which leads to a delayed type-I interferon response\(^{95,96}\). The importance of RIG-I has further been demonstrated in this work\(^{97,98}\). By introducing mutations to the NS1 protein, some studies have shown that the NS1 protein of H5N1 viruses may more strongly antagonize cellular antiviral responses\(^{99,100}\). Another factor that may contribute to the H5N1 virus’ enhanced...
pathogenicity is that H5N1 can more readily infect lung resident macrophages, though there is conflicting evidence on whether infected macrophages lead to enhanced inflammation\textsuperscript{101} or not\textsuperscript{102}. Given the many factors contributing to H5N1’s pathogenicity, there is an opportunity to use dynamical mathematical modeling to analyze time-course infection data and identify the processes (factors) that differ between infections with different viruses.

Dynamic mathematical models have been used to better understand the mechanisms driving \textit{in vitro} and \textit{in vivo} immunodynamics observed during influenza infection (\textsuperscript{103,104} are reviews of select relevant models). To date, most mathematical models of influenza infection consist of ordinary differential equations (ODEs) that systematically link virus replication and the availability of host target cells (cells that can be infected) to intracellular immune signaling (interferon responses) and/or immune cell activity. These models have been used to explore a variety of areas: target cell refractory periods as an explanation for double viral peaks\textsuperscript{85}, prioritizing therapeutic targets to optimally reduce inflammation while controlling viral load\textsuperscript{105–108}, providing evidence that interferon paracrine signaling is the primary factor regulating hypercytokinemia\textsuperscript{109}, and determining why viral titers rebound during bacterial co-infection\textsuperscript{110}. Separately, agent-based models (ABMs), a rule-based approach that treats each cell as an individual entity while considering spatial effects and stochasticity, have been used to reveal the optimal experimental conditions for examining infection-induced interferon production, to quantify the benefits of noisy intracellular immune signaling\textsuperscript{75}, and to elucidate the effect of spatial aspects on infection outcomes\textsuperscript{111}. An engineering-based approach that employed a reduced ODE model of virus replication and treated measurements of key immune factors as system inputs suggested that increased levels of interferon-\textalpha/-\textbeta promoted slower viral growth and limited immune cell stimulation in aged mice\textsuperscript{112}. As in these previous studies, mathematical modeling is a
knowledge-driven, integrative approach well suited to explore the regulatory mechanisms responsible for the differences observed between mild and severe influenza infections.

To elucidate the biological mechanisms that contribute to the distinct immunodynamics observed between H1N1 and H5N1 influenza virus infections, we developed mechanism-based, dynamic mathematical models of the innate immune response and performed several parameterizations to identify the biological processes (parameters) that are most likely to be differentially regulated between the two infections. The model parameters were fit to viral load and immunologic data from mice that had been infected with either an H1N1 or H5N1 virus. Using parameterization to test the contribution of macrophage activity in interferon production and viral suppression reveals that the inclusion of these mechanisms may negatively impact model quality (Section 4.3.1). Global sensitivity analyses of the models (Section 4.3.2) reveal distinct system control for each structure. Comparing model fits to the data (Sections 4.3.3 – 4.3.5) using the Akaike information criteria (AIC) suggests that the optimal model is achieved when the production rate of interferon \( r_{I,V} \) is distinct between the two infections. In total, this modeling-based approach determines that the distinct rate of interferon induction in H5N1 infections is the most likely candidate mechanism for explaining the distinctive immune response observed in H5N1 infections.
4.2 Materials and Methods

4.2.1 Model development rationale and equations

Studies have established that many innate immune processes are differentially regulated in mild and severe influenza infections\textsuperscript{113–118}. As such, we focused on developing simple models of the early immune response. We first describe the relevant immunology and then describe how virus replication and innate immunity are mathematically modeled.

Lung epithelial cells as well as lung-resident innate immune cells, i.e., macrophages and neutrophils, display pattern recognition receptors to detect viral RNA at the site of infection\textsuperscript{119,120}. Pattern-recognition toll-like receptors 7 and 9 (TLR7/9)\textsuperscript{121}, retinoic acid-inducible gene I (RIG-I)\textsuperscript{30}, and the pro-inflammatory NF-kB\textsuperscript{122} pathway work in concert to activate the type-I interferon response\textsuperscript{96}. These pattern recognition receptors are antagonized by the invading virus to strain-specific degrees\textsuperscript{29,31}. Interferons induce the transcription of interferon stimulated genes (ISGs) that are responsible for establishing an antiviral state in the cells near infected cells\textsuperscript{123} and activating several components of the immune system. Studies suggest that the timing of the type-I interferon response is key in limiting viral replication and recruiting an appropriate pro-inflammatory response\textsuperscript{124–126}. Induction of interferon production is also partially responsible for regulating the activity of innate immune cells such as macrophages and neutrophils. The precise role of these immune cells in viral clearance is still debated: while macrophages can engage in several inflammatory processes\textsuperscript{127} and are important for enhancing interferon induction\textsuperscript{128}, they may also be targeted for infection by highly pathogenic viruses such as H5N1 influenza, altering their overall activity\textsuperscript{101,102}.
A significant body of contradictory predictions of macrophages’ role in infection was reviewed\textsuperscript{101,102,127,129–132}. To explore all this space of mechanistic possibilities, four distinct model structures were created. Literature supporting the removal of each mechanism is given as super-text in Figure 17, while the parameters controlling each process are indicated only in Model 1 for simplicity.

### 4.2.2 Experimental data collected from literature and relating the data to the model

Measurements of the viral load, interferon concentration, and a surrogate measurement of macrophage counts were collected and organized from Shoemaker et al.\textsuperscript{90}. Briefly, female C57BL/6J mice were infected with a low pathogenic A/Kawasaki/UTK-4/09 H1N1 virus (H1N1) or high pathogenic A/Vietnam/1203/04 H5N1 virus (H5N1) at 1E5 PFU. A control group was mock-infected with PBS. At 14 time points spanning the first week of infection, three animals per infection group were sacrificed. Their lungs were harvested and analyzed by a variety of techniques to quantify the viral load and the state of the immune system. The H5N1 infected animals died between days 5 and 7. As such, only the first 13 measurements spanning days 0 – 5...
are included in this work. In all, 234 measurements (78 for each model state) were collected and organized for model parameterization.

The specific measurements and their relationship to the mathematical models are as follows: Viral titers were determined via plaque assay resulting in units of plaque-forming units per mg of lung tissue (PFU/mg). In Appendix C.1, $[V]$ is the log10 of PFU/mg. To represent the change in interferon concentration over time, $[I]$, log2 fold change of the gene expression of $Ifnb1$ relative to mock-infected, time-matched samples (unitless) was used. Full details on normalizing the gene expression can be found in the original work. Whole lung macrophage counts were determined at only four time points in the original work, spread across several days. As a result, the concentration of MCP1 (measured using ELISA assay) was selected to act as a surrogate measurement of macrophage cell count, $[M]$. Appendix C.3, Figure 48 shows a linear regression of the log10 macrophage cell count and log2 MCP1 concentration ($R^2 = 0.98$, with a slope of 0.613). The conversion between macrophage and MCP1 is, therefore, given by Equation 4-1:

$$\log_{10}[M] = 0.6301 \log_2[MCP1]$$ (4-1)

where $[M]$ is the macrophage cell count in the lung and $[MCP1]$ has units of pg/mL. During parameter training, the macrophage state is fit to the log2 of MCP1 measurements. Equation 4-1 is then used to transform MCP1 predictions into estimates of macrophage counts in the lung.

4.2.3 Parameter training

Basin Hopping (BH) via SciPy and Parallel Tempering Markov chain Monte Carlo (PT MCMC) were employed as global optimization algorithms to train parameter values. BH rapidly identifies a single estimate of parameter values, while PT MCMC characterizes the parameter space over an extended number of samples. The objective function used (Equation 4-2) is akin to
a weighted sum of squared error and is referred to hereafter as energy, following Metropolis et al.\textsuperscript{17}. This function was developed to equally consider error contribution from all ODE states ($[V]$, $[I]$, $[M]$) by normalizing over the data. Nearly any custom error function which quantifies the difference between model predictions and available data could be used in place of Equation 4-2 to accommodate different model and data structures.

\[
\text{Energy} = \sum_{x=1}^{X} \sum_{t=0}^{T} \frac{(M_{x,t} - O_{x,t})^2}{2O_{x,t}}
\]  

(4-2)

$M_{x,t}$ and $O_{x,t}$ are the model output and the average of triplicate observed data points, respectively, for each state, $x$, and time point, $t$, across all states, $X$, and time points, $T$. Each time point was divided by the corresponding data point, $O_{x,t}$, to normalize energy values. The initial condition values (at time=0) are not used in data fitting, since the ODEs will always start with these values. All MCMC simulations were run across six chains of temperature (0.99, 0.9, 0.8, 0.4, 0.2, and 0.05) to ensure adequate exploration of parameter space. Parameters were bounded by the numerical stability of the system; and priors were defined as uniform between zero and the stable upper limit. Optimized parameter values near this limit were evaluated again, until the result was both stable and well within bounds. Visualization code for the MCMC parameterizations can be found at github.com/ImmuSystems-Lab/Macrophage_Model. Additional information on parameterization results can be found in Appendix C.4.

4.2.4 Model and scenario prioritization

While an energy function conveys the quality of the fit achieved by parameterization for a given model, it is incapable of comparing models with varying numbers of parameters (differing degrees of freedom). Akaike Information Criterion (AIC) was used to compare models with
different numbers of parameters and determine the superior model based on a tradeoff between the model’s fit to training data (energy) and the number of free parameters used to achieve the fit. The optimal model is the model that reports the lowest AIC value. While Bayesian Information Criterion is argued to best suited to select a ‘true model’ out of an ensemble, AIC is better at removing incorrect models. For the current work, only a small subset of all possible model structures was investigated. Thus, we could not assume that the true model would be present, and AIC was used for model selection. As AIC is relative, a difference greater than 2 was considered significant when comparing two outcomes. AIC is defined as:

\[
AIC = -2 \ln (MLE) + 2 \cdot P_{\text{free}}
\]

(4-3)

where MLE is the maximum likelihood estimate and \( P_{\text{free}} \) is the number of parameters being fit. The number of free parameters in a model depends on the scenario being considered, which is described in Tables 2 and 3 in Sections 4.3.4-5.

4.2.5 Sensitivity analysis

An extended Fourier Amplitude Sensitivity Testing (eFAST) global sensitivity analysis was performed in Python Version 3.8.10 with Sensitivity Analysis Library (SALib) Version 1.4.5 to determine the output variance of each state (Appendix C.1) as a function of input variance to each parameter. The output of the method is First-Order indices that represent the outcome variance of each system state that can be attributed to the perturbation in a single parameter, \( p \). High First-Order indices imply that a single parameter has a significant role in controlling system outcomes, while low values indicate a less significant impact. While sensitivity of infection outcomes (viral and IFN peaks, etc.) were previously used, we aimed to determine the overall system sensitivities or the output variance of each state; \([V], [I], [M]\), for each
parameter, \( p \). This was determined at 100 time points between 0 and 5 days. The average of these sensitivity indices over all time points is reported as a metric of overall system control during the infection.

4.3 Results

4.3.1 In silico screenings of candidate innate immune models find that H5N1 and H1N1 viruses induce interferon production at different rates in vivo

The goal of this work is to determine the innate immune processes that are differentially regulated in animals infected with a moderate H1N1 or severe H5N1 influenza virus. These processes can be represented as differences in the values of a parameter of the mathematical model. To identify differentially regulated processes, four biologically informed mathematical models with structural differences surrounding macrophage activity were developed, and a series of parameter fittings was performed to determine which parameter(s) distinguish an optimal fit to experimental data derived from H5N1 and H1N1-infected animals.

The models are also shown in Appendix C.2, wherein four different regulatory structures link the concentration of virus, \([V]\), the level of interferon, \([I]\), and the number of macrophages, \([M]\), in the lung. The primary distinction between the four models involves the role of macrophages. In Model 1, containing 12 parameters, macrophages can induce interferon production and suppress virus replication. However, experimental evidence suggests that macrophages may not play a major role in suppressing virus replication\(^{132}\). As such, we constructed four models of the innate immune response. In Model 2, macrophage induction of interferon
production is removed. In Model 3, macrophages do not directly suppress virus replication. Models 2 and 3 both contain 11 parameters. In Model 4, containing 10 parameters, both macrophages’ ability to induce interferon and suppress virus replication are removed.

All four models were compared under three scenarios using the AIC as the discrimination metric. The overall strategy of the approach is illustrated in Appendix C.2. In the ‘No strain-specific differences’ (NSSD scenario, or All Shared [AS] Condition within this Aim), parameters have equal values in both infections, with their optimized values resulting in trajectories between both sets of data. A single copy of each model is trained to the H5N1 and H1N1 data, resulting in one trained (parameterized) model. In the ‘One strain-specific difference’ (OSSD) scenario, we assume that a single interaction or process may be differentially regulated in the two infections. To consider this, we train two copies of a model to the data, one copy for the H5N1 data and another for the H1N1 data, but only allow one parameter to take on different values between each copy (referred to as independent parameters, like the approach used in 108). All other parameters must maintain the same value. This results in an H1N1 and an H5N1-specific parameterized version of a model, each of which has identical parameter values except for the strain-specific parameter under consideration. And lastly, we considered the ‘All different’ (AD) scenario in which all parameters can take on different values when training a model to the H1N1 or H5N1 data, resulting in an H1N1 and an H5N1-specific parameterized version of the model in which all parameters have different values. AD provides a benchmark of the equations’ ability to capture the dynamics of each strain individually, while NSSD benchmarks the goodness of fit for when each infection is mechanistically identical. All four models were parameterized under each scenario (AD, NSSD, and an OSSD for each parameter in each model structure) using a basin hopping
algorithm. AIC scores were used to determine which model and scenario results in the best fit to the data.

![Figure 18. Energy versus AIC values for all four model structures under different parameterization scenarios.](image)

All Different (AD), One Strain-Specific Difference (OSSD), and No Strain-Specific Difference (NSSD). The Model 4 OSSD $r_{I,V}$ scenario yields the global minima. Each shape represents a maximum of 500 Basin Hopping stochastic jumps. Each jump is followed by up to 500 gradient descent steps. If 250 jumps pass without an improvement in the global minima, Basin Hopping exits. Although this does not prove convergence, this practice reliably finds the minima from the proposed initial parameters. Most (97%+) parameterizations do not require full iterations.

Comparing the AIC results after training each model to the experimental data under each scenario suggests that H5N1 and H1N1 viruses induce the production of interferon at different rates. Figure 18 shows the energy and the AIC for all tested combinations of model and scenario. Generally, all four models can attain similar goodness of fits to the immunologic data. Model 4 tends to have the lowest AIC, a result of both low energy fits and the fact that Model 4 has the
fewest parameters. The lowest energy is achieved by all four models under the AD scenario, which is expected as this scenario has the highest degree of freedom for fitting the models to the data. However, the lowest AIC values are not achieved under the AD scenario. The minimum AIC occurs in the OSSD scenario where the parameter representative of interferon production rate, $r_{I,V}$, takes on H5N1 and H1N1-specific values. All four models achieve their lowest AIC under this condition (noted in Figure 18) with Model 4 achieving the lowest AIC overall. This suggests that virus-induced interferon production is regulated in a strain-dependent manner, a proposition that is independent of the model, and therefore, macrophage activity, employed. These findings also suggest that Model 4 is the best model for regressing against the H5N1 and H1N1 immunologic data. However, these conclusions are drawn from Basin Hopping, which converges when no new minimum can be found within 50 stochastic jump steps. To test the validity of these results, parallel Tempering Markov Chain Monte Carlo methods were used for a deeper exploration of parameter space in Section 4.3.3.
4.3.2 Strain-specific interferon production is not an artifact of parameter sensitivity

A challenge associated with this type of *in silico* screen is to determine if the screening methods have merely identified the most sensitive model parameter as the best parameter to take on different values and provide the best fit to the data. We next investigated the parametric sensitivity of the candidate models to determine if $r_{I,V}$ was the most sensitive model parameter. We conducted a sensitivity analysis of all the models to each of their constituent parameters using the eFAST algorithm\textsuperscript{136,137}. The sensitivity of each state is reported in the form of fractional variance that can be explained by the variance of a single parameter, $p$. These indices are shown in Figure 19.

![Figure 19. First-order indices of the eFAST sensitivity analysis of the Macrophage models.](image)

Indices are reported as the normalized change for each model state, for each parameter. Parametric sensitivity analysis for each model shows that the most sensitive parameters differ across the candidate models. In Model 1, the concentration of interferon (I) and number of lung
macrophages, \([M]\), are most sensitive to macrophage associated parameters \((r_{M,I} \text{ and } r_{I,M})\), while the concentration of virus \([V]\) is primarily dependent on the rate of interferon induction by the virus, \(r_{I,V}\). This trend holds for Models 2 and 3. In Model 4, the concentration of interferon, \([I]\), and \([M]\) are most sensitive to the rate of interferon induction by \([V]\), \(r_{I,V}\), while \([V]\) is most sensitive to \(r_{V,V}\). This establishes that the four model structures have a unique control scheme, i.e., the most sensitive parameters differ between the different models. This also demonstrates that the minimum AIC values of \(r_{I,V}\) OSSD models during the \textit{in silico} screen were not simply the result of \(r_{I,V}\) being the most sensitive parameter. Thus, the remainder of this work is comprised of further analyses using Model 4 to understand the parameter space associated with the model fitting to H5N1 and H1N1-specific data. Aim 4 builds on this observation that models with slightly different structures have significant control differences, applying it to strain-specific parameters.

\[4.3.3 \text{ Exploration of Model 4’s parameter space using PT MCMC}\]

Preliminary \textit{in silico} screens and sensitivity analyses establish that Model 4 provides the best fit to the immunologic data when \(r_{I,V}\) is allowed to take on H5N1 and H1N1-specific values. However, further exploration of the parameter space using Parallel Tempering Markov Chain Monte Carlo (PT MCMC) parameterization was needed to determine the breadth of the parameter space that supported Model 4’s best fit to H5N1 and H1N1 data. Using PT MCMC, we re-evaluated all the scenarios described in Appendix C.2, Figure 47 B for Model 4. For each MCMC optimization, 2 million iterations were run.

Figure 20 shows the fits of Model 4 under the ‘all different’ (AD) and ‘no strain-specific differences’ (NSSD) scenarios plotted against the H5N1 and H1N1 \textit{in vivo} mouse data. Standard deviation intervals of the top 1,000 solutions, i.e., the 1,000 lowest energy parameter sets that were
identified, are narrow for the model’s fits under both the AD (black) and NSSD (blue) scenarios, indicating a range of possible model trajectories with similar energy. The resultant trajectory for the NSSD scenario is the average of the two strains’ data sets and, expectedly, fits neither strain. The AD scenario fits reproduce the observed dynamics for each strain very well, showing that the Model 4 equations can produce known \textit{in vivo} behavior and strain-specific parameterizations can improve model energy at the cost of higher degrees of freedom.

Figure 20. The top 1,000 fits of Model 4 to the H1N1 (top row) and H5N1 data (bottom row) when using PT MCMC parameterization.

The top fits under the AD scenario (all parameters allowed to independently estimate across strains) are shown in black and NSSD results (all parameters shared between strains) are shown in blue. Intervals represent the standard deviation of the 1,000 lowest energy parameter sets. Data from Shoemaker et al.\textsuperscript{90} are shown as circles with bars indicating the standard deviation. Within the AD scenario, almost all parameters have strain-specific values except for $r_{V,I}$. Of note between
H1N1 and H5N1 are the rate of host interferon production ($r_{I,V}$, 1.66 vs. 2.83 days$^{-1}$) and the rate of viral replication ($r_{V,V}$, 0.957 vs. 1.19 days$^{-1}$).

### 4.3.4 MCMC-based parameter exploration again finds that H5N1 and H1N1 viruses induce interferon production at different rates in vivo

We next considered Model 4’s goodness of fit to the H5N1 and H1N1 data under the OSSD scenarios using PT MCMC. The energy and AIC for all scenarios tested are reported in Table 2. For completeness, we show the time course trajectories of the best fit achieved for Model 4 under all OSSD scenarios in Figure 21. Energy per iteration for both AD and NSSD scenarios are shown in Appendix C.4, while best-fit parameter values and units are provided in Appendix C.6.

**Table 2.** The minimum energy, degrees of freedom (DoF), and AIC values achieved by Model 4.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Independent Parameter</th>
<th>Energy</th>
<th>DoF</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSSD</td>
<td>None</td>
<td>15.04</td>
<td>10</td>
<td>50.08</td>
</tr>
<tr>
<td>AD</td>
<td>All</td>
<td>3.33</td>
<td>20</td>
<td>46.66</td>
</tr>
<tr>
<td>OSSD</td>
<td>$r_{V,I}$</td>
<td>10.83</td>
<td>11</td>
<td>43.66</td>
</tr>
<tr>
<td></td>
<td>$r_{V,V}$</td>
<td>9.37</td>
<td>11</td>
<td>40.74</td>
</tr>
<tr>
<td></td>
<td>$K_{V,V}$</td>
<td>9.79</td>
<td>11</td>
<td>41.59</td>
</tr>
<tr>
<td></td>
<td>$d_{V}$</td>
<td>9.65</td>
<td>11</td>
<td>41.31</td>
</tr>
<tr>
<td></td>
<td>$\eta_{V}$</td>
<td>6.65</td>
<td>11</td>
<td>35.30</td>
</tr>
<tr>
<td></td>
<td>$d_{I}$</td>
<td>10.3</td>
<td>11</td>
<td>42.61</td>
</tr>
<tr>
<td></td>
<td>$r_{M,I}$</td>
<td>12.36</td>
<td>11</td>
<td>46.73</td>
</tr>
<tr>
<td></td>
<td>$k_{M,I}$</td>
<td>12.29</td>
<td>11</td>
<td>46.57</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>12.28</td>
<td>11</td>
<td>46.57</td>
</tr>
<tr>
<td></td>
<td>$d_{M}$</td>
<td>12.37</td>
<td>11</td>
<td>46.75</td>
</tr>
</tbody>
</table>
The lowest AIC is achieved when the rate of virus induction of interferon, $r_{I,V}$, is allowed to have strain-specific values. Minimum energy values fall between 9 and 13 except in the case where the rate of interferon production, $r_{I,V}$, is independently estimated, which yields a minimum energy of 6.65. While this is closest to the minimum AD energy for Model 4 (3.33), AIC calculations reveal that the resulting value of 35.30 for $r_{I,V}$ is not only lower than the results of the other nine OSSD parameterizations of Model 4 but is lower than that of the high degree of freedom AD results. Overall, using MCMC instead of basin-hopping for data fitting did not lead to a different conclusion with regards to the optimal solution occurring when $r_{I,V}$ is independently estimated for H5N1 and H1N1. The independent parameter column identifies the parameter allowed to take on different values while training two copies of the model to the H5N1 and H1N1 data. These fits are demonstrated in Figure 21.

Figure 21. Model 4 output for the minimum energy parameter set (lines) for OSSD parameterizations and corresponding training data (markers) for H1N1 (top row) and H5N1 (bottom row).
Data is from Shoemaker et al.\textsuperscript{90} are shown with the standard deviation associated with triplicate data points per time point. In total, the model fits in Figure 21 capture some trends of the \textit{in vivo} data, but many scenarios result in steady-state dynamics after just 1 day ($d_I, r_{I,V}$). However, the $r_{I,V}$ OSSD scenario has a distinct numerical improvement in fit over the NSSD results, although this result may be hard to visually determine. When each parameter is allowed to differ between strains, histograms can inform whether the strains’ parameter distributions are unique. Focusing on the $r_{I,V}$ OSSD scenario histograms, a comparison of the resultant top 1,000 parameter distributions across strains yields a significant difference between distribution means [Mann-Whitney test $p < 0.001$ for $r_{I,V}$ between H1N1 (blue) and H5N1 (red), Figure 22] indicating that the strains have unique values for this parameter. All other parameter distributions for OSSD models overlap significantly (Figure 22), except for $d_I$. This is to be expected when these parameters are shared between strains. The clear separation in distributions for the unshared parameter is not always present. When other parameters are allowed to vary between strains, the unshared parameter distributions often have significant overlap (Appendix C.7, Figures 51 – 60). Combined with the AIC results in Table 2, these results highlight that $r_{I,V}$ OSSD achieves the most statistically defensible fit to the datasets.
Figure 22. Posterior density distributions for all parameters for Model 4 with $r_{I,V}$ varying between strains.

Only $r_{I,V}$ can have strain-specific values. All other parameters have the same value when fitting the model to H5N1 and H1N1 data. The x axis is given in $\log_{10}$ Parameter Value. Distributions result from the 1,000 lowest energy solutions identified using PT MCMC. Narrow posterior distributions indicate that the parameter had a small range of values under which the model optimally fit the data, while broad distributions indicate that a range of values would yield fits of the same energy.

4.3.5 Independent estimation of virus parameters per strain does not improve model AIC

Because it would be computationally intractable to fit all possible combinations of parameter values, this study focused largely on observing the effect that differences in single parameters while training to two infection data sets has on model quality. However, we
hypothesized that disparate immune dynamics between viral strains may be related to all virus-based rates such as growth rate, $r_{V,V}$, or death rate, $d_V$. To test this, Model 4 was parameterized such that the viral state parameters, $r_{V,V}$, $K_{V,V}$, $r_{V,I}$ and $d_V$ (Denoted $\{V\}$) could take on different values when training to the H5N1 and H1N1 data, while all other parameters remained shared between strains. Six additional ‘Virus-Host’ parameterizations were performed with the addition of one of the non-viral state parameters, $\{V\} + OSSD$ (DOF: 15). Independent parameter identifies the parameters allowed to take on different values when training to the H5N1 and H1N1 data.

Table 3. The minimum energy, DOF, and AIC values for all seven Viral subset, $\{V\}$.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Independent Parameter</th>
<th>Energy</th>
<th>DoF</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>${V}$</td>
<td>${V}$</td>
<td>9.34</td>
<td>14</td>
<td>46.68</td>
</tr>
<tr>
<td>$({V} + OSSD)$</td>
<td>${V} + r_{L,V}$</td>
<td>5.55</td>
<td>15</td>
<td>41.11</td>
</tr>
<tr>
<td></td>
<td>${V} + d_I$</td>
<td>8.38</td>
<td>15</td>
<td>46.75</td>
</tr>
<tr>
<td></td>
<td>${V} + n_I$</td>
<td>8.66</td>
<td>15</td>
<td>47.72</td>
</tr>
<tr>
<td></td>
<td>${V} + K_{M,I}$</td>
<td>8.89</td>
<td>15</td>
<td>47.79</td>
</tr>
<tr>
<td></td>
<td>${V} + n$</td>
<td>8.92</td>
<td>15</td>
<td>47.85</td>
</tr>
<tr>
<td></td>
<td>${V} + d_M$</td>
<td>8.89</td>
<td>15</td>
<td>47.78</td>
</tr>
</tbody>
</table>

The model solutions for each parameterization are found in Figure 23. Qualitatively, the resulting fits are more indicative of the expected dynamic trends, including during late infection. Model fits follow similar trends, with $\{V\} + r_{L,V}$ achieving the best fit to data. Corresponding minimum energy and AIC values are found in Table 3. A comparison of the top 1,000 parameter distributions per strain yields significant differences between distribution means, except for $r_{V,I}$ in $\{V\} + K_{M,I}$ (Mann-Whitney test $p < 0.001$ for all independently estimated parameters). This indicates that virus-related kinetic parameters likely vary between strains. Minimum energies associated with the $\{V\} + OSSD$ parameterizations are lower than that of $\{V\}$ alone, with a minimum energy of 5.55, associated with the independent fitting of $\{V\} + r_{I,V}$. Compared to the AD and NSSD scenarios, $\{V\} + r_{I,V}$ results in a lower AIC value, reiterating the role of interferon
production rate in strain-specific infection dynamics. Although strain-specific viral parameters are

demonstrably present in the datasets, \( V + r_{I,V} \) has a higher AIC than the \( r_{I,V} \) OSSD scenario. This

attributes great importance to strain-dependent interferon production rate over simple strain-

dependent viral kinetics and implies that increased degrees of freedom are detrimental to model

quality. Investigations with higher degrees of freedom were not performed due to the

computational time required for each MCMC fit to run 2 million samples per study.
Figure 23. Model 4 output for minimum energy parameter set (line) for virus-related parameter independent fits.

Corresponding training data (markers) for H1N1 (top row) and H5N1 (bottom row). \{V\} is representative of four viral parameters: \(r_{v,v}\), \(K_{v,v}\), \(r_{v,l}\), and \(d_v\). Data from Shoemaker et al.\(^90\) are shown with the standard deviation associated with triplicate data points per time point. The model trajectories in Figures 22 and 23 display an interesting property – a bias towards simplicity. Metrics other than AIC which have relative weighting of complexity and predictive accuracy would create different dynamics, lying between optimization entirely for simplicity (predict the mean of the data) and optimization entirely for data fitting (error-based model evaluation). Likely since lower DOF eventually favors predicting the mean of the data, all scenarios achieved stable steady states.

The selection of error function is arbitrary if the resulting relative weighting of error and complexity in AIC are considered. For this application, both metrics contribute similar scalars to AIC (total error ranging from ~1 to 10 per strain for typical parameterizations, and 10 – 24 DOF). Thus, the model is greatly (sometimes dominantly) influenced by the DOF penalty, especially when compared to an error-only model parameterization.
4.4 Conclusion

Several mechanisms have been hypothesized to explain differential immunoregulation between low and high pathology infections. Three specific mechanisms are that high pathogenic viruses may simply replicate more quickly, high pathogenic viruses may differently interact with antiviral signaling pathways (i.e., interferon signaling), or high pathogenic viruses may infect and/or alter the behavior of macrophages (see introduction for further details). Prior modeling efforts implicated the infection of macrophages as a driving factor for strain-dependent pathogenicity\textsuperscript{108}, however, the study did not consider alternative mechanisms and further exploration was needed\textsuperscript{146,147}. The \textit{in silico} screen used here is an unbiased approach that allows several candidate mechanisms to compete, with the most likely candidate mechanism being selected based on the model’s goodness of fit to H5N1 and H1N1 training data. The infection data originates from identical lineage, age, and gender-matched murine subjects, minimizing inter-individual variability and increasing the likelihood that differences observed between infections are due to strain-specific immunoregulation or virus replication.

Of the three hypotheses for why H5N1 viruses induce distinct immune responses, the primary finding from the \textit{in silico} screen (Section 4.3.1) is that the rate of interferon production by infected lung cells is likely different in H1N1 and H5N1-infected animals. The lowest AIC was achieved when the interferon-associated parameter, $r_{I,V}$, was allowed to take on different values while training to each infection cohort – regardless of the model employed. The robustness of this finding is further supported by the wide distribution of parameter values which optimally fit the data, quantified by the MCMC analysis, and by the results of the sensitivity analysis. One concern about our \textit{in silico} screening approach, and indeed in model-based analysis in general, is that the most sensitive parameters are often identified as the most important for maintaining phenotypes.
as they are the easiest to use for tuning system dynamics. Across the four models considered here, the top parameters to which the model outputs are sensitive differed (Section 4.3.2). Nevertheless, $r_{I,V}$ was identified as the most likely candidate across all four models. Finally, in Section 4.3.4, MCMC analysis showed that the best fit for the scenario with strain-specific $r_{I,V}$ values could be achieved for a wide range of parameter values. It was found that the rate of interferon production, $r_{I,V}$, is approximately 2 – 3 times larger in H5N1 infected lung cells. Additional analyses were performed to consider strain-specific virus replication rates combined with strain-specific immune rates. Our work demonstrates that the magnitude of the interferon response is strain-dependent, and that these differences arise primarily from the host cells’ rate of interferon production in response to viral presence rather than strain-specific viral replication behavior.

It is important to note that while the in silico screen identifies strain-specific interferon production as the key mechanism for differential immunodynamics, this does not fully negate the possibility of other mechanisms. Each of the three mechanisms discussed are supported by some studies and contradicted by others. For example, with regards to macrophages, studies have shown that macrophages are susceptible to high pathogenic viruses and H5N1 viruses can replicate in human macrophages cultured from monocytes. However, it has also been shown that macrophages collected from human donors can be infected by H5N1 viruses, but do not produce virus nor inflammatory cytokines. With regards to strain-specific regulation of interferon, there is evidence that H5N1 viruses may upregulate interferon production early in infected cells in vitro. Our in silico screen considers several possible mechanisms for why H5N1 and H1N1 immunodynamics may differ, and while we conclude that strain-specific interferon production is the most likely mechanism, we only considered two or more possible mechanisms occurring
simultaneously. Given the complexity of the immune system, future efforts will focus on considering more complex candidate mechanisms.
5.0 SPOT

5.1 Introduction

This Aim does not represent a publication, but the development of a toolkit which it is hoped will have novel insights on multiple condition or mixed effect modeling in the future. Thus, each section of the Results for this aim may be thought of as “Future Work” more than complete analysis. A model’s parameter identifiability, mathematical stability, and sensitivity greatly influence the control and dynamics of the system, and any analyses performed on results from the model. Any singular result could be an artefact from particulars of the model construction. An alternative approach, called the ensemble approach, has applications for strain-specific viral dynamics, as demonstrated in Aim 3 using multiple model structures to overcome some of these limitations which arise from using a single model. This approach could similarly be applied to sex-based hormonal differences, age differences, treatment groups, or other multi-condition data sources. However, this approach has been deployed in a limited fashion due to both the computational cost of repeated MCMC parameter explorations and the need for human curation of the model structures and parameter sharing. This Aim proposes a flexible, computationally optimized method to enumerate and evaluate large ensembles of models with different parameter sharing between multiple data sets. The methodology and code are called the Shared Parameter Optimization Toolbox (SPOT).

This toolbox contains several analysis tools for model ensembles, which are detailed in Results. SPOT highlights the host production of interferon as the main difference between H5N1 and H1N1 dynamics across 1,024 parameter sharing combinations, directly supporting the findings
from Aim 3. This result achieved a p-value of < 0.01. However, any comparative analysis method requires at least two sources of data, limiting applications unless synthetic datasets are utilized.

5.2 Materials and Methods

5.2.1 Methodology

At its core, SPOT evaluates the overall fitness of models with a semi-stochastic gradient descent Basin Hopping algorithm tasked to find optimum parameters for each model instance. An instance of the model is parameterized using data from two or more sources. This optimization method is not an exhaustive search, or parameter exploration, like MCMC, instead returning a minimum energy parameter set with 1 – 3 orders of magnitude reduction in computational time and memory footprint. This allows for most machines to evaluate as many model structures in parallel as available CPU cores. The fitness is the Bayesian Information Criterion, Equation 5-1, which considers both the model’s ability to capture the dynamics present in data (model fit) and the number of parameters required to achieve this (Degrees of Freedom). The fitness of many models is used in meta-analyses to find features in model structures which result in consistently improved model fitness. These analyses will be detailed along with results from reanalyzing Aim 3 using SPOT.

\[ BIC = k \ln(n) - 2 \ln(L) \]  

(5-1)

A more complete mathematical description of SPOT is given in Appendix D.1. For convention, each model structure has a genome, which consists of a number of entries equal to the
number of parameters in the original model. The value for each parameter can be in two states. When the value is non-strain specific for all instances, referred to as 0 within the genome, the parameter is Shared. When the value is strain-specific, 1 within the genome, each dataset has an individual value of the parameter. This adds model complexity but may improve BIC for features which have sufficient justification present in the data. These parameters are called Unshared. The results in a small array of Booleans which can be combined with a model and multiple data sets, Sharing or Unsharing parameters as laid out by the genome.

5.2.2 Software

All programming for this work was performed using Python 3.8.10. The full code is available at github.com/ImmuSystems-Lab/shared-parameter-optimization-toolbox. Relevant packages and versions are provided in Table 4.

<table>
<thead>
<tr>
<th>Package Name</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEAP</td>
<td>1.3.1</td>
</tr>
<tr>
<td>itertools</td>
<td>8.11.0</td>
</tr>
<tr>
<td>matplotlib</td>
<td>3.5.0</td>
</tr>
<tr>
<td>multiprocess</td>
<td>0.70.12.2</td>
</tr>
<tr>
<td>numba</td>
<td>0.54.1</td>
</tr>
<tr>
<td>pandas</td>
<td>1.3.4</td>
</tr>
<tr>
<td>scipy</td>
<td>1.7.2</td>
</tr>
<tr>
<td>seaborn</td>
<td>0.11.2</td>
</tr>
<tr>
<td>sklearn</td>
<td>0.0</td>
</tr>
<tr>
<td>statsmodels</td>
<td>0.13.2</td>
</tr>
</tbody>
</table>

5.2.3 Data Sources

This work involves the parameterization of various model structures, using both synthetic and biological datasets. A brief description of each biological dataset and their respective
application within the current work is provided below, followed by an overview of the methods used to create synthetic datasets. Additional information on the dataset methodologies and structures can be found in Aim 3. Briefly, the biological datasets used in conjunction with the Ackerman MCP Model were sourced from Shoemaker et al\textsuperscript{139}. This in vivo data is derived from triplicate C57BL/J mice, which were infected with either A/Kawasaki/UTK-4/09 H1N1 virus (H1N1) or A/Vietnam/1203/04 H5N1 virus (H5N1). $[M]$ and $[I]$ states are in log\textsubscript{2} space, while $[V]$ is in log\textsubscript{10} space.

5.3 Results

5.3.1 Characteristic Curves from Brute Force Evaluation

The number of possible model structures is combinatorial, i.e., Aim 3’s model with 10 parameters and 2 data sources has 1,024 possible parameter sharing structures varying between 10 and 20 degrees of freedom. SPOT’s evaluation of these structures takes roughly 4 hours real-time, with 56 evaluations in parallel. It is important to note that these results are not based on the sensitivity or identifiability of any single model; instead, they are based on features present in the dataset. Figure 24 shows Models 1 and 4 from Aim 3, demonstrating the separation in model structures according to common parameterization features.
Figure 24. Models 1 and 4 Characteristic Curves for a SPOT-significant parameter ($r_{I,V}$) and a non-significant parameter ($r_{V,V}$).

Characteristic curves for Models 1 and 4, separated by Shared (orange) and Unshared (blue). Each dot represents a complete parameterization of a genome via Basin Hopping. Despite unique model structures and sensitivities, both Models strongly indicate that strain-specific $r_{I,V}$ values consistently justify the increased degrees of freedom. This can be seen as clear visual separation of Shared and Unshared model fits in the left column. This implies that at least one of the biological mechanisms which $r_{I,V}$ represents causes the infection dynamic differences between H5N1 and H1N1. The viral replication rate, $r_{V,V}$, surprisingly does not have a significant SPOT value, i.e., model structures with strain-specific viral replication parameters do not improve data fits sufficiently to justify the increased degrees of freedom. This can be seen as a lack of visual separation between Shared and Unshared model fits in the right column and indicates similar viral replication rates during the experiment.
5.3.2 Categorical Linear Regression Highlights Strain – Specific Rates

For every model structure, the genomes (parameter sharing rules) are separable into Shared and Unshared categories for each parameter (0 and 1, respectively). By applying a multiple categorical regression to the BICs of each genome, a slope is found between the Shared and Unshared categories. This slope represents the degree of distinction between model fits which Share or Unshare this parameter and are influenced by the BIC of both groups. This makes a single value, the slope of the BICs between these groups, encompass both model fit and complexity between two groups of models. Each group has a shared feature, i.e., a parameter is Shared within all model instances of Group A. The same parameter is all Unshared within Group B. These slopes fall into one of several categories, and the resulting value of categorical slopes for each parameter in Aim 3, Model 4 are given in Figure 25.

![Figure 25. SPOT value for each parameter from Aim 3, Model 4 with biological data.](image)

The SPOT value (y-axis) represents a parameter’s normalized change in BIC when said parameter is allowed to fit in a strain-dependent manner. Parameters with no strain-dependent mechanisms typically have a positive SPOT value, comprised of equal model fit quality (energy)
but increased complexity (DOF). Those parameters with strain-dependent values may have lower negative SPOT values, indicating an improvement in model fit with a magnitude larger than the DOF penalty imposed.

**SPOT Value of 1:** A SPOT value of exactly 1 usually indicates a parameter which has a global sensitivity of zero, i.e., the parameter has no control over system dynamics. Parameters of this type have no ability to improve model energy and thus only increase DOF when Unshared. These parameters cannot be identified as significant. Identifying these earlier and removing them from the genome enumeration would save time for parameter sweeps, if they can’t be eliminated any other way.

**SPOT Value Near 1:** A SPOT value near 1 indicates a parameter which should most likely be shared between all present datasets, since independently fitting the parameter (a genome value of 1 for the parameter, increasing the DOF by fitting a new parameter value for each data set) correlates with an increased BIC. A SPOT value of about 1 demonstrates that the additional DOF is the main cause of increased BIC when the parameter is Unshared, and that the model energy is typically similar to genomes with the same parameter being Shared.

**SPOT Value Significantly Above 1:** A SPOT value significantly above 1 (and/or extremely wide confidence intervals in SPOT values, outliers, and large spread in the DOF:BIC characteristic curve) may indicate poor model fits. This can usually be rectified with increased Basin Hopping or gradient descent steps, a finer time resolution during ODE solving, modification of the model equations or solver methods, parameter bounds, initial guesses, or data transformations. A few high-BIC genomes are typical for real models, simply due to the mathematical stability issues when separate ODE model instances share parameters and random
initial parameter guesses. If the minimum number of acceptable model fits for the desired resolution and p-value are attained, simply discarding these outliers is acceptable.

**SPOT Value Near 0:** A value near 0 shows that model fits improve, but the DOF penalty from the additional parameter is equal in magnitude to the improvement in energy. Note that Categories 3 and 4 (SPOT ≤ 0) often occur with structurally unidentifiable parameters like decay rates. Other model structures with different identifiability can be used to probe these aspects of the datasets (see Aim 3).

**Negative SPOT Value:** A negative SPOT value indicates that the associated parameter should most likely be fit independently for each dataset. The improvements in model fit (energy) have a large enough magnitude to overcome the DOF penalty for additional parameters, and thus justify the unshared values. This can occur with none, one, or several parameters, depending on the model structure, data, energy function, parameter identifiability, and global sensitivity. Thus, SPOT results must be utilized in the context of a well-understood model environment and interpreted with these limitations in mind.
5.3.3 Sensitivity of SPOT to Dataset-Specific Parameter Values

The SPOT value for a parameter depends strongly on the presence of a difference present in the underlying data. An increased magnitude of this difference leads to more significant (lower or more negative SPOT values). By generating synthetic datasets using a model using intentional strain-specific parameter differences, the relative change in SPOT value to these differences can be elucidated. This is given in Figure 26.

![Figure 26. SPOT Values with Synethic Strain-Specific Parameter Differences.](image-url)
Since the model evaluations have been performed *a priori*, this analysis is a matter of visualizing the models’ features. In the top panel, the parameter \( r_{I,V} \) was increased only for the H5N1 synthetic dataset, while remaining constant in the H1N1 dataset, in ratios from 1.25 to 5.0. As this ratio increases, \( r_{I,V} \)’s SPOT value decreases, trending towards 0. However, other parameters (\( K_{V,V}, r_{V,V}, \) and \( d_V \)) also experience lower SPOT values. This is due in part to parameter identifiability within the model, but also indicates a dataset which has not yet achieved a statistically significant difference between H5N1 and H1N1 for any parameter. It is expected that the first parameter which achieves a negative SPOT value should be considered significantly different between data sets. In the bottom panel \( d_V \) is varied in the same strain-specific ratios to no effect. Unidentifiable parameters do not have the same effect on the system, meaning that fully probing parameter ratios in this way can narrow down which parameters are significant, similar to a global sensitivity analysis.

### 5.3.4 SPOT with Reduced Samples

In the case of large models or early screenings, it may not be necessary to evaluate all possible model genomes. Here, random sets of genomes from Aim 3, Model 4 were evaluated in SPOT. 100 such random samplings were performed for each Fraction of All Possible Samples. The results of this down sampling are given in Figure 27.
Figure 27. SPOT values from 0% to 100% of possible genomes for Aim 3, Model 4.

SPOT was applied on Aim 3, Model 4, with H5N1 and H1N1 infected mice providing two biological strains. Solid lines indicate the average of 100 resamples at 100 fractions, resulting in 10,000 total data points. Shaded areas are the 95% confidence intervals. At a fraction of 0.1, the SPOT predictions have settled close to the values obtained with all genomes evaluated. Most parameters have SPOT values near 1, suggesting that these parameters should be shared between H5N1 and H1N1 strains. $r_{i,v}$’s SPOT value is -0.7, meaning that this parameter is tied to the most likely mechanism to explain the strain differences seen in the data. $d_I$’s SPOT value is close to zero, suggesting either that independently fitting this parameter to each strain could partially explain the difference between strains, or parameter identifiability issues with $r_{i,v}$. When applying resampling, it is important to consider the loss of statistical power incurred. The number of samples required to meet any given p-value vary by parameter, as demonstrated in Figure 28.
Figure 28. Model 4 P-values of SPOT regression versus the fraction of all possible genomes analyzed can determine the significance of SPOT values.

A. Overall behavior of P-values. An area of interest (p-values from 0.01 to 1) is given in B. While most parameters have reached significance (p < 0.01) by 10% of all genomes, $r_{1,1}$, the only parameter with a negative SPOT value in this model, takes 80% of genomes evaluated before the trend reaches statistical significance. The use of these p-values can set confidence levels on comparative analyses, which are difficult when comparing parameter values alone.
6.0 Discussion and Future Work

Together, this work makes several attempts to explain the different dynamic regimes of the early interferon signaling during viral respiratory infections. Each Aim targets different aspects of this critical immune signaling step. Aim 1 establishes best- and worst-case inflammatory states with an ODE model and shows the robustness of RIG-I activity to viral antagonism. Aim 2 extends this with spatial considerations and demonstrates that RIG-I can prolong the survival of infected cells. Aim 3 highlights that H5N1 and H1N1 do not have significantly different viral replication dynamics, and indeed primarily differ by the host immune response. This could be from the unique nonstructural protein (NS1) of each influenza strain differentially antagonizing RIG-I. Finally, Aim 4 proposes a method to elucidate these strain-specific differences in a rapid, unbiased manner.

From these aims, a robust picture of the innate immune response to viral respiratory infection emerges. The establishment of a robust, early, pro-inflammatory (high IFN) response is critical to resisting infection, but dysregulation of this response may lead to cytokine storm or prolonged inflammatory-mediated cell damage. Future work will focus on the role of the adaptive immune system and clinical interventions to modulate the overall course of an infection.

6.1 ODE Model of Interferon During Viral Respiratory Infection

In Aim 1, an ODE model of the intracellular innate immune system’s response to influenza infection was developed and used to evaluate system properties associated with viral lethality and IFN production dynamics. This model considers signaling components not available in previous
work, uses unique data for parametrization and validation that perturbs distinct aspects of the pathways, and is numerically stable and well suited for continued adaption in future multiscale modeling efforts. This model does not stand alone in literature, and indeed must serve as a basis of further advancements. The action of drugs can be simulated by modifying only the relevant term(s) which the drug is thought to affect. Similarly, other strains or species of viral infection can be modeled by changing only the viral replication and clearance parameters. Both possibilities would be significantly computationally cheaper than the reparameterization of the entire model for each new system. If other parameters must change to reasonably fit a new data source, this could indicate previously unknown effects of the drug or infection on the species modeled herein, like NS1’s well known antagonistic effects. The degree of this antagonism on RIG-I massively controls cytokine production trajectory and infection outcomes, as demonstrated here in silico by varying levels of RIG-I knockdown. This insight applies to all RNA viruses, with varying mechanisms of antagonism. Such predictions generated by the ODE model will provide insight into infection trajectory, disease outcome, and their manipulation by intervention.

6.2 Spatial Effects of Interferon Regulation and Viral Infections

In Aim 2, we developed a mechanistic, multicellular spatial model of interferon signaling (the MSIS model) that we used to evaluate how changes in select reaction rates impacted plaque growth in RNA virus-infected cell cultures. One major shortcoming of the MSIS model is the lack of additional mechanisms to support simulating cell death. During infection, cell death occurs via several mechanisms, including via programmed cell death (apoptosis) and pyroptosis, cell death induced via inflammasomes\textsuperscript{145}. Lacking these mechanisms, cell death only occurs in the MSIS
models as the intracellular concentration of virus increases and the cell health declines (Appendix B.1). Cells can become stuck in the I2 cell type as the reduced concentration of intracellular virus and the slow rate of health decline significantly reduces the likelihood of a cell transitioning to the dead type. The equation that defines how cell health declines (Appendix B.2) was directly translated from a population-level model where health translated to the fraction of uninfected cells, but as a model of the health of a single cell, having the rate of health decline be linearly dependent on the current health of the cell (i.e., health declines more rapidly for healthier cells) might not be reasonable. To improve the model's relatability to experimentation, future work will focus on including additional mechanisms of cell death as well as improving the kinetic description of how cell health impacts a cell’s transition to death.

The SARS-CoV-2 and influenza viruses for which this model was constructed have many similarities. Like influenza’s NS1 protein, SARS-CoV’s NSp1 antagonizes RIG-I signaling89, and genome analysis shows an 87% conservation of NSp1’s genome between SARS-CoV and SARS-CoV-263. This similarity suggests that the MSIS model could readily be adapted to model SARS-CoV-2-induced interferon signaling from measurements of SARS-CoV-2-specific virus kinetics. The MSIS model can also be extended to consider additional spatial aspects of infection. The modular architecture supports independent and collaborative development of extensions to account for additional immune response mechanisms in vitro such as IFN-mediated cell death. It also supports extending the model to include aspects of the immune response in vivo such as propagation of IFN signaling by local innate immune cells and recruitment of adaptive immune cells to the site of infection.
6.3 Strain-Specific Modeling of H5N1 and H1N1 Influenza

In Aim 3, four distinct, three-state ODE models of the early innate immune response to influenza virus were used to investigate the mechanistic roots of differential immunoregulatory behavior observed in vivo between low and high pathogenic H1N1 and H5N1 strains.

The caveats of this study primarily relate to the available data and parameter identifiability. Insufficient macrophage count data was available, and the concentration of MCP1 was used as a surrogate measurement. While data was available to assess the accuracy of using MCP1 as a surrogate, there remains the possibility that macrophage counts differed from our estimates. With regards to parameter identifiability, in highly connected systems such as Model 1, it is often difficult to reasonably estimate values for all parameters. This can be improved in future work by incorporating data from knockout mice studies wherein feedback in immune signaling can be removed. The shared parameter optimization framework is highly generalizable to other cohorts of data including age, race, and sex-specific studies, making it a highly valuable tool for investigating disparate kinetics between groups of interest and the drivers of observed clinical behavior\cite{151,152}. Additionally, conclusions from cohort-specific studies may prove useful for informing and simplifying future modeling work with additional cohorts.

6.4 Minimizing Model Size While Fitting Multiple Data Sets with Parameter Sharing

In Aim 4, we present the Shared Parameter Optimization Toolbox (SPOT), a comparative analysis toolkit which can demonstrably determine the smallest number of parameters required for a model to predict multiple sets of conditions by evaluating models with different structures.
SPOT’s use of multiple model structures allows for selection of a model regardless of individual parameter values or underlying data, i.e., the strain-specific interferon production in Aim 3 is highlighted via SPOT as the best parameter sharing scheme overall. This type of multiple condition comparative modeling has applications in many fields, allowing for unbiased comparison of features in a computationally efficient manner.

Down sampling or Latin hypercube sampling can be used to reduce the number of parameter optimizations, or genomes, to be evaluated for any given analysis. Fixing certain parameters, such as those determinable from experiments, those structurally unidentifiable, or those of little to no interest, to make models more computationally tractable.
6.5 Publications Resulting from This Work


Appendix A Innate Immune Model

Appendix A.1 ODE Model

Weaver et al.’s\textsuperscript{97} model of the innate immune response to respiratory RNA viral infection consists of seven ordinary differential equations (states) with three fixed parameters and 15 unknown parameters. Type-I interferon, $[IFN]$, and extracellular Type-I interferon $[IFN_e]$, represent the pro-inflammatory cytokine response mounted by the innate immune system. Phosphorylated signal transducer and activator of transcription, $[STATP]$, and interferon regulatory factor 7 and its phosphorylated form, $[IRF7]$ and $[IRF7P]$, track the cellular mechanisms which propagate this cytokine signal. The proportion of living cells, $[P]$, is a measure of remaining viable cells, while the virus, $[V]$ is a measure of viral load, normalized from 0 to 1 against the wild-type PR8 peak viral load. For a single set of equations (one strain), the model has 14 parameters (DOF: 14). The parameters $n$, $k_7$, $\tau_2$, $\tau_3$, $\tau_4$, and $\tau_5$ have fixed values sourced from the original work and its cited literature. These parameters are decay rates ($k_7$, $\tau_x$) and a Hill-like coefficient ($n$), whose values are not expected to change under any conditions investigated in this work. The Weaver model is given in Equations A.1-1 through A.1-7. Grey shaded parameters were to be included in SPOT analyses.

\[
\frac{d[IFN]}{dt} = [P]\left(\frac{k_1}{k_{13}}[RIGI][V] + \frac{TLRk_{12}[V]^n}{k_{14} + [V]^n} + k_{14}[IRF7P]\right) - k_{21}[IFN] \quad (A.1-1)
\]

\[
\frac{d[IFN_e]}{dt} = k_{21}[IFN] - [IFN_e]\tau_2 \quad (A.1-2)
\]
\[
\frac{d[STATP]}{dt} = \frac{k_{31}[P][IFN_e]}{k_{32} + k_{33}[IFN_e]} - [STATP] \tau_3 \tag{A.1-3}
\]

\[
\frac{d[IRF7]}{dt} = [P](k_{41}[STATP] + k_{42}[IRF7P]) - [IRF7] \tau_4 \tag{A.1-4}
\]

\[
\frac{d[IRF7P]}{dt} = k_{51}[P][IRF7] - [IRF7P] \tau_5 \tag{A.1-5}
\]

\[
\frac{d[P]}{dt} = -k_{61}[P][V] \tag{A.1-6}
\]

\[
\frac{d[V]}{dt} = \frac{k_{71}[P][V]}{1 + k_{72}[IFN_e]} - k_{73}[V] \tag{A.1-7}
\]

Equations A.1-1 through A.1-7 are measured in units of μM, thus, the right-hand side of each equation is in units of μM hour\(^{-1}\). All species except [P] and [V] are in units of μM. [TLR] and [RIGI] are set to unity and represent the presence of the sensor proteins. [P] represents the number of productive, infected cells present divided by the initial count of said cells; [P] = (live cells)/(initial cells). Since a multiplicity of infection (MOI) of 5\(^5\) results in 99.3% of target cells becoming infected\(^6\), it is assumed that the entire cell population is producing new viral particles at the start of the trial. After normalization of virus, the initial concentration is 6.9E-8. [P] will thus vary from 1 to 0 and is unitless. [V] represents a virus concentration normalized to the maximum amount observed; [V] = (virus concentration)/(max virus concentration). These values can take on molarity, PFU, or similar matching units; [V] is unitless and can vary from 0 to 1. The viral titers\(^48,49\) are in units of PFU mL\(^{-1}\), and the maximum measure was the 24-hour observation of wild-type PR8. A [V] value > 1 is possible for viral strains with higher peak viral loads than wild-type PR8 and at model timespans greater than 24 hours.

The parameterization of the ODE model was accomplished with a parallel tempering Markov Chain Monte Carlo (MCMC). The first 10\(^3\) iterations comprised burn-in, a period where the MCMC algorithm was searching parameter space for potential local minima. This can be seen in the acceptance ratio (Figure 29 A) and sum squared error (Figure 29 B).
A 19% acceptance ratio is held after $10^3$ burn-in samples. The MCMC algorithm would ideally accept new parameters 23% of the time; for this parameterization, the acceptance ratio of the primary chain was 19%. This was considered acceptable. Further hyperparameter tuning to obtain exactly 23% acceptance was possible, but not pursued. The overall fit of the model, as quantified by the energy, is shown in Figure 29 B. After burn-in, no trend towards fit improvement is observed, despite high-temperature chain exploration for other local minima, thus, the MCMC fitting algorithm is sufficiently converged. Two additional optimizations of the same length were run with randomized starting parameter values, which converged to parameterizations with the same SSE value and ranges as Run 1. The parameterization with the lowest SSE was used. The values for this parameterization are given in Table 5, along with their origin. All three MCMC runs converged to redundant parameter sets; these represent the minimum sum squared error (SSE) among all runs. Literature sourced values gave estimates for some species’ decay rates, which were not fit via MCMC.
Table 5. Best-fit model parameters for the Innate Immune Model.

<table>
<thead>
<tr>
<th>Process</th>
<th>Parameter</th>
<th>Units</th>
<th>Value</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>[IFN] induction via RIG-1</td>
<td>$k_{11}$</td>
<td>μM hours$^{-1}$</td>
<td>1e5</td>
<td>Fit manually</td>
</tr>
<tr>
<td>[IFN] induction via TLR7</td>
<td>$k_{12}$</td>
<td>hours$^{-1}$</td>
<td>9.746</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{13}$</td>
<td>[unitless]</td>
<td>12.511</td>
<td></td>
</tr>
<tr>
<td>[IFN] induction via [IRF7P]</td>
<td>$k_{14}$</td>
<td>hours$^{-1}$</td>
<td>13.562</td>
<td></td>
</tr>
<tr>
<td>Transport of [IFN]</td>
<td>$k_{21}$</td>
<td>hours$^{-1}$</td>
<td>10.385</td>
<td>Fit via MCMC</td>
</tr>
<tr>
<td>[IFN$_e$] degradation</td>
<td>$\tau_2$</td>
<td></td>
<td>3.481</td>
<td></td>
</tr>
<tr>
<td>[STAT] phosphorylation</td>
<td>$k_{31}$</td>
<td>μM hours$^{-1}$</td>
<td>13.562</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{32}$</td>
<td>μM</td>
<td>5.464</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{33}$</td>
<td>[unitless]</td>
<td>0.068</td>
<td></td>
</tr>
<tr>
<td>[STATP] dephosphorylation</td>
<td>$\tau_3$</td>
<td></td>
<td>0.3$^1$</td>
<td>52</td>
</tr>
<tr>
<td>[IRF7] induction via [STATP]</td>
<td>$k_{41}$</td>
<td></td>
<td>0.115</td>
<td>Fit via MCMC</td>
</tr>
<tr>
<td>[IRF7] induction via [IRF7P]</td>
<td>$k_{42}$</td>
<td>hours$^{-1}$</td>
<td>1.053</td>
<td>53</td>
</tr>
<tr>
<td>IRF7 mRNA degradation</td>
<td>$\tau_4$</td>
<td></td>
<td>0.3$^3$</td>
<td>54</td>
</tr>
<tr>
<td>IRF7 phosphorylation</td>
<td>$k_{51}$</td>
<td>hours$^{-1}$</td>
<td>0.202</td>
<td>Fit via MCMC</td>
</tr>
<tr>
<td>Cell death</td>
<td>$\tau_5$</td>
<td></td>
<td>0.3$^5$</td>
<td></td>
</tr>
<tr>
<td>Viral replication</td>
<td>$k_{61}$</td>
<td></td>
<td>0.635</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{71}$</td>
<td></td>
<td>1.537</td>
<td></td>
</tr>
<tr>
<td>[IFN] effect on virus</td>
<td>$k_{72}$</td>
<td>μM$^{-1}$</td>
<td>47.883</td>
<td>Fit via MCMC</td>
</tr>
<tr>
<td>Nonspecific viral clearance</td>
<td>$k_{73}$</td>
<td>hours$^{-1}$</td>
<td>0.17</td>
<td></td>
</tr>
</tbody>
</table>

1 Later work$^{98}$ determined a minimum value of 0.75 for mathematical stability
The parameter distributions are shown in Figure 30.

![MCMC Parameter Histograms](image)

**Figure 30. MCMC Parameter Histograms.**

Most parameters had a wide range of acceptable values which produced low-energy model fits. Narrower distributions are indicative of parameters which have few low-energy values. Note non-normal distributions for several parameters. The lack of normalcy is an indication of interparameter correlation within the MCMC samples. This correlation was tested using the Pearson correlation coefficient. The magnitude of correlations for all parameter combinations is summarized in Table 5.

**Appendix A.2 Structural Identifiability**

Only parameters $k_{31}$, $k_{32}$, and $k_{33}$ were structurally unidentifiable under Perfect Identifiability. This was expected, since the complexity of the Michaelis–Menten kinetic form precludes a unique set of constants from being identified within this ODE system without
additional information or a simplification of the kinetic term. Parameters from perfect identifiability, plus $k_{14}$, $k_{42}$, and $k_{51}$, were practically unidentifiable. Reparametrizing the ODE model presented in this work with time-course data for environmental IFN, cell population and an IRF7P proxy may permit unique solutions. Additionally, as the number of observed points approach a continuous data set, these parameters would become identifiable. [STATP]’s Michaelis–Menten parameters could only be uniquely identified with a more complex experimental design, a restructuring of the kinetic term to reduce fitted constants, or additional studies isolating its kinetic behavior. Structural identifiability results are summarized below.

\[
\frac{d[IFN]}{dt} = [P] \left( k_{11}[RIG][V] + \frac{[TLR]k_{12}[V]^n}{k_{13}+[V]^n} + k_{14}[IRF7P] \right) - k_{21}[IFN] \quad (A.2-1)
\]

\[
\frac{d[IFNe]}{dt} = k_{21}[IFN] - [IFNe]\tau_2 \quad (A.2-2)
\]

\[
\frac{d[STATP]}{dt} = \frac{k_{31}[P][IFNe]}{k_{32} + k_{33}[IFNe]} - [STATP]\tau_3 \quad (A.2-3)
\]

\[
\frac{d[IRF]}{dt} = [P](k_{41}[STATP] + k_{42}[IRF7P]) - [IRF]\tau_4 \quad (A.2-4)
\]

\[
\frac{d[IRF7P]}{dt} = k_{51}[P][IRF7] - [IRF7P]\tau_5 \quad (A.2-5)
\]

\[
\frac{d[P]}{dt} = -k_{61}[P][V] \quad (A.2-6)
\]

\[
\frac{d[V]}{dt} = \frac{k_{71}[P][V]}{1+k_{72}[IFNe]} - k_{73}[V] \quad (A.2-7)
\]

**BOLD** parameters are structurally unidentifiable under perfect observation. (All states directly observed). Grey boxed parameters are practically unidentifiable under the observation scheme available in the data50.
Appendix A.3 Parameter Correlation

Figure 31. Correlation plots of MCMC parameter exploration.

Nine parameter pairs were found to have a Pearson’s Correlation Coefficient > ±0.5. These significant pairs are plotted here. High correlation and non-normal distributions were not unexpected, and are a frequent challenge in fitting nonlinear, biological signaling systems. Several model features, namely multiple parameters affecting the same species and the presence of feedback loops by nature, necessitate a correlated random walk to maintain or improve the model’s
fit. This manifests as a lower-than-expected acceptance ratio, which can be overcome by hyperparameter tuning if undesirably slow parameter space exploration results.

Significant correlation pairs (Correlation coefficient > ±0.5) highlighted with Grey Box.

Table 6. Pearson Correlation Coefficients for ODE parameters found via MCMC.

<table>
<thead>
<tr>
<th>$k_{13}$</th>
<th>$k_{14}$</th>
<th>$k_{21}$</th>
<th>$\tau_{u2}$</th>
<th>$k_{31}$</th>
<th>$k_{32}$</th>
<th>$k_{41}$</th>
<th>$k_{42}$</th>
<th>$k_{51}$</th>
<th>$k_{61}$</th>
<th>$k_{71}$</th>
<th>$k_{72}$</th>
<th>$k_{73}$</th>
<th>Par</th>
</tr>
</thead>
<tbody>
<tr>
<td>.17</td>
<td>.77</td>
<td>.3</td>
<td>.02</td>
<td>.31</td>
<td>.53</td>
<td>.05</td>
<td>.01</td>
<td>.23</td>
<td>.15</td>
<td>.01</td>
<td>.09</td>
<td>.35</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>.09</td>
<td>.7</td>
<td>.1</td>
<td>.28</td>
<td>.3</td>
<td>.63</td>
<td>.02</td>
<td>.04</td>
<td>.03</td>
<td>.43</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-.06</td>
<td>-.04</td>
<td>.5</td>
<td>.39</td>
<td>.12</td>
<td>.17</td>
<td>.23</td>
<td>.34</td>
<td>.01</td>
<td>.1</td>
<td>0.28</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.16</td>
<td>0</td>
<td>.36</td>
<td>.2</td>
<td>.09</td>
<td>.01</td>
<td>.23</td>
<td>.01</td>
<td>.01</td>
<td>.06</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.13</td>
<td>.5</td>
<td>.66</td>
<td>.45</td>
<td>.01</td>
<td>.03</td>
<td>.05</td>
<td>.13</td>
<td>.1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.48</td>
<td>.44</td>
<td>.47</td>
<td>.14</td>
<td>.18</td>
<td>.03</td>
<td>.08</td>
<td>.09</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.47</td>
<td>.05</td>
<td>.03</td>
<td>.06</td>
<td>.03</td>
<td>0</td>
<td>.06</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.04</td>
<td>.17</td>
<td>.13</td>
<td>.06</td>
<td>.13</td>
<td>.29</td>
<td>.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.05</td>
<td>.05</td>
<td>.06</td>
<td>.1</td>
<td>.29</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.35</td>
<td>.01</td>
<td>.05</td>
<td>.07</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.03</td>
<td>.07</td>
<td>.06</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.44</td>
<td>.03</td>
<td>.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.17</td>
<td>.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

106
Appendix B Multiscale Immune Model

Appendix B.1 Cell Type Transition Probabilities

Transitions between cell types are stochastic, and the probability of a transition occurring depends on the simulation time step ($\Delta t=0.167$ hours), the cell’s local extracellular and intracellular concentration of virus ($[V_e]$ and $[V]$ respectively), the cell’s health ($[H]$; described under Appendix B.2 Intracellular Model Equations and Rationale), and a transition rate coefficient ($\beta$, $k$ or $\gamma_{H,V}$). $[V_e]$ per cell is measured as the local extracellular virus concentration each cell is exposed to over its entire cell area. We derived the rates in Equation B.1-1 through B.1-3 from the rate laws in $^85$, following literature transformation rules$^71$. When a cell is infected (transitions from $U$ to $I1$), the internal viral concentration changes from 0 to $6.9E-8$ (unitless), equivalent to a single virus particle entering the cell$^84$. This amount of virus was considered negligible compared to $[V_e]$ and is thus not removed from the extracellular virus. Table 8 gives all simulation parameter values. Within CC3D, cell-type transitions are implemented by sampling a random number for each cell between 0 and 1, inclusive, at each time step. The cell’s transition occurs when the probability, $P$, is greater than the random number. Each transition must occur in order.

\[
P(U \rightarrow I1) = 1 - \exp (-\beta [V_e]_{\text{per cell}} \Delta t), \tag{B.1-1}
\]

\[
P(I1 \rightarrow I2) = 1 - \exp (-\tau_{I1} \Delta t) \tag{B.1-2}
\]

\[
P(I2 \rightarrow D) = 1 - \exp (-\gamma_{H,V} [V] \{1 - H\} \Delta t) \tag{B.1-3}
\]
Appendix B.2 Intracellular Model Equations and Rationale

Figure 8 ‘Intracellular’ panel shows key molecules, species, and processes involved in an epithelial cell sensing RNA virus infection and producing IFN to suppress virus replication. The MSIS model adapts and extends to single cells the Weaver model of the dynamics of virus replication, interferon signaling, and cell health. The adapted Weaver model includes six ODEs (Equations B.2-1 through B.2-6) that define the rate equations for intracellular virus replication, interferon signaling, and cell health. Specifically, the equations define changes in the intracellular concentrations of virus ([\(V\)], unitless), interferon ([\(IFN\)] μM), phosphorylated STAT ([\(STATP\)], μM), IRF7 protein ([\(IRF7\)], μM), and phosphorylated IRF7 ([\(IRF7P\)], μM). Equation B.2-6 defines the dynamics of the health of the cell ([\(H\)], unitless). [\(IFN_{e}\)]_{per\ cell} is the average extracellular interferon concentration each cell is exposed to over its entire cell area.

Rate equations for intracellular species and health adapted from Weaver et al:

\[
\frac{d[V]}{dt} = \frac{k_{V,V}[H][V]}{1+[IFN_{e}]_{per\ cell}} - Q_{V}[V] \quad (B.2-1)
\]

\[
\frac{d[IFN]}{dt} = [H](k_{IFN,V}(RIGI)V + \frac{k_{IFN,Y(TLR)}V^n}{k_{IFN,Y(TLR)}+V^n} + k_{IFN,IRF7P}[IRF7P]) - Q_{IFN}[IFN] \quad (B.2-2)
\]

\[
\frac{d[STATP]}{dt} = \frac{k_{STATP,IFN_e}[H][IFN_{e}]_{per\ cell}}{k_{STATP,IFN_e}+[IFN_{e}]_{per\ cell}} - \tau_{STATP}[STATP] \quad (B.2-3)
\]

\[
\frac{d[IRF7]}{dt} = [H](k_{IRF7,STATP}[STATP] + k_{IRF7,IRF7P}[IRF7P]) - \tau_{IRF7}[IRF7] \quad (B.2-4)
\]

\[
\frac{d[IRF7P]}{dt} = k_{IRF7P,IRF7}[H][IRF7] - \tau_{IRF7P}[IRF7P] \quad (B.2-5)
\]

\[
\frac{d[H]}{dt} = -\gamma_{H,V}[H][V] \quad (B.2-6)
\]
Below, we provide a brief description of the Weaver model and then discuss how we modified the model to support its implementation in the MSIS model (see Figure 8 Intracellular panel). A thorough description of the rate equations is available.\textsuperscript{84}

The Weaver model groups interferon α and β into the single representative species interferon $[IFN]$. We modeled the inhibition of virus production in response to the cell’s spatially averaged level of extracellular interferon ($[IFNe]_{\text{per cell}}$) using non-competitive inhibition-like kinetics. We used mass-action kinetics to describe the induction of IFN (Equation B.2-2) by virus (via the RIG-I pathway) and $[IRF7P]$, and Hill kinetics to define the effect of the concentration of virus on IFN production via the TLR pathway. The rate of export of intracellular IFN into the extracellular environment obeys the concentration of IFN times a rate constant, $Q_{IFN}$. We model extracellular IFN’s ($[IFNe]_{\text{per cell}}$) activation of STAP with Michaelis–Menten kinetics (Equation B.2-3), and mass-action kinetics are used to model the effect of $[STATP]$ and $[IRF7P]$ on the rate of production of IRF7 (Equation B.2-4). We also use mass-action kinetics to describe the rate of $[IRF7P]$ production as a function of IRF7 (Equation B.2-5). In all equations, production terms are multiplied by the cell’s health, $[H]$, to represent the loss of production capacity in an infected cell. Heath is a relative metric bounded between 0 and 1, and the rate of the decay of health (Equation B.2-6) is proportional to the concentration of virus in the cell and the health of the cell. All these rate laws are consistent with the original Weaver model.

We made three changes to the Weaver model to employ it in the MSIS model. We reinterpreted the first-order virus degradation term in the original Weaver model to represent the release of virus into the extracellular environment in the MSIS model. The rate of release of virus to the extracellular environment is proportional to the concentration of virus times a rate constant, $Q_v$. The Weaver model was a population model, while Equation B.2-1 through B.2-6 represent the
intracellular regulation of a single cell. The Weaver model has a state, \( [P] \), which represents the fraction of live cells in the population. The mathematical equation for health is unchanged from the original Weaver model, but we have reinterpreted \([P]\) to represent the health, \([H]\), of each cell. All production terms are multiplied by the cell’s health (bound between 0 and 1) to represent the diminished production capacity of unhealthy, virus-infected cells. And, lastly, due to the spatial aspect of the MSIS model, we redefined the concentration of extracellular IFN in to be the average \([IFNe]\) over the area of a given cell; namely \([IFNe]_{\text{per cell}}\).

In the multicellular spatial MSIS model, each live cell (U, I1, I2 types) has a replica of the rate equations. For U and I2 cell types, the equations and their parameter values are unaltered. In I1 cells, the equations are the same and all parameter values are unchanged except for the parameter value that defines the rate of virus release into the extracellular environment, \(Q_v\), which is set to zero because eclipse phase cells (I1) do not release virus. Now that the mathematics of each cell’s dynamic responses have been defined, the spatial considerations must be accounted for when determining the concentrations of \([IFNe]\) and extracellular virus \([Ve]\) on a per-cell basis.

**Appendix B.3 Diffusion of Extracellular Species and Implementation in CC3D**

Virus releasing cells (I2) release intracellular virus into the extracellular environment. Uninfected, eclipse phase and virus releasing cells (U, I1, and I2) produce and export type-1 interferons in response to either virus sensing proteins or autocrine/paracrine signaling. In cell culture, these extracellular species diffuse freely in the medium above the apical surface of cells.

The MSIS model contains a cell lattice next to two chemical field lattices (described above) and the diffusion of extracellular species across either chemical field lattice is unaffected by the
presence of cells in the adjacent cell lattice. Equation B.3-1 models diffusion of extracellular interferons, where \( D_{IFNe} \) is the diffusion coefficient of interferon, \( Q_{IFN} \) is the rate constant for export of interferon by cells into the extracellular environment, and \([IFN]\) is the internal amount of interferon inside each cell. Cell types U, I1, and I2 can produce and export interferon.

Equation B.3-2 models diffusion of the extracellular virus, where \( D_{Ve} \) is the diffusion coefficient of virus and \( Q_v \) is the secretion rate constant for release of virus by late infected (I2) cells. Intracellular virus, \([V]\), is a normalized, unitless quantity representing the per cell viral load, while extracellular virus \([Ve]\) has units of PFU mL\(^{-1}\) and represents the concentration of infectious virus in the extracellular environment. The unit conversion is achieved via \( Q_v \)'s units of PFU mL\(^{-1}\) hours\(^{-1}\).

\[
\frac{\partial [IFNe]}{\partial t} = D_{IFNe} \nabla^2 [IFNe] + Q_{IFN}[IFN] - \tau_{IFNe}[IFNe]
\]  
(B.3-1)

\[
\frac{\partial [Ve]}{\partial t} = D_{Ve} \nabla^2 [Ve] + Q_v[V] - \tau_{Ve}[Ve]
\]  
(B.3-2)

CompuCell3D solvers use a simple time-slicing algorithm. Each CompuCell3D time step represents 0.167 hours. CompuCell3D first calculates the integrated amount of \([Ve]\) and \([IFNe]\) directly above each cell to calculate \([Ve]\)\(_{per\ cell}\) and \([IFNe]\)\(_{per\ cell}\) and passes these values to internal cellular ODE instances. It then integrates the diffusion and the intracellular species’ rate equations forward in time independently, using these fixed values of \([Ve]\)\(_{per\ cell}\) and \([IFNe]\)\(_{per\ cell}\) for the equivalent of 0.167 hours. It then calculates the amount of virus released and the \([IFN]\) exported from each cell over 0.167 hours and adds the amount released divided by the cell area into each voxel in the appropriate field at each position corresponding to a voxel of that cell. It then evaluates the probabilities for cell type transitions for each cell following to determine whether each cell experiences such a transition. For a more complete description of how CC3D implements simulations, please see 76.
Appendix B.4 Initial and Boundary Conditions

All simulations use periodic boundary conditions along the x and y axes. When simulating low MOI conditions, at time zero all cells are U type, except for one I1 cell at the center of the simulation (two I1 cells in Figure 10). To simulate high MOI conditions, all cells are initially I2 type. Table 7 gives the initial conditions for the intracellular variables of each cell type at time zero. In all simulations, the extracellular environment initially contains no $[V_e]$ or $[IFNe]$. To simulate interferon pretreatment, the simulation starts at 12 hours pre-infection, with all cells U type and exposed to $[IFNe] = 0.04 \, \mu$M. At time = 0 hours (12 hours after $[IFNe]$ exposure), we simulate washing of the cells by setting $[IFNe] = 0 \, \mu$M and initiate the infection by setting a cell at the center of the simulation’s lattice to the I1 type. Due to the IFN pretreatment, all cells have the same intracellular state at time zero except for the single I1 cell, for which $[V]$ is set to 6.9E-8 (unitless).

<table>
<thead>
<tr>
<th>Intracellular Species Initial Value</th>
<th>Cell Type</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U</td>
<td>I1</td>
</tr>
<tr>
<td>$[IFN]$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$[STATP]$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$[IRF7]$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$[IRF7P]$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$[H]$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$[V]$</td>
<td>0</td>
<td>6.90E-8</td>
</tr>
</tbody>
</table>

Appendix B.5 Parameter Determination

Many MSIS model parameters come directly from the Weaver model\textsuperscript{84}. The Weaver model was parameterized using the lowest sum-of-squares error (energy) resulting from a parallel tempering Markov chain Monte Carlo fit to data collected from HBECs infected with wild-type
A/Puerto Rico/8/1934 Influenza A\textsuperscript{50}. Each cell’s ODE model in the MSIS model is the Weaver model, modified as described previously. We adopted additional parameters from the literature\textsuperscript{53,54,78,154}. Table 8 gives a comprehensive list of model parameters and their origin. Virus diffusion coefficients can vary by several orders of magnitude depending on media type, based primarily on the medium’s viscosity\textsuperscript{71}. We set the diffusion coefficient for both $[V_e]$ and $[IFN_e]$ to 54.0 $\mu$m$^2$ s$^{-1}$, within the range of experimental measurements\textsuperscript{154,155} for both species. For these diffusion coefficients, the baseline parameter set led to continuous plaque growth. We rescaled the cell type transition parameter $\beta^{55}$ from units of median tissue culture infectious dose (TCID$_{50}$ hours$^{-1}$) to plaque-forming units (mL PFU$^{-1}$ hours$^{-1}$) for consistency with the Weaver model’s units for viral load.
Table 8. Baseline parameter values and sources.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
<th>Process</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{IFN,V(RIGI)}$</td>
<td>0.0</td>
<td>μM hours⁻¹</td>
<td>IFN production via RIG-I</td>
<td>84</td>
</tr>
<tr>
<td>$k_{IFN,V(TLR)}$</td>
<td>9.746</td>
<td>hours⁻¹</td>
<td>Maximal rate of IFN production</td>
<td></td>
</tr>
<tr>
<td>$K_{IFN,V(TLR)}$</td>
<td>12.511</td>
<td>[unitless]</td>
<td>Apparent dissociation constant</td>
<td></td>
</tr>
<tr>
<td>$k_{IFN,IRF7P}$</td>
<td>13.562</td>
<td>hours⁻¹</td>
<td>IFN production via $[IRF7P]$</td>
<td></td>
</tr>
<tr>
<td>$Q_{IFN}$</td>
<td>10.385</td>
<td>hours⁻¹</td>
<td>Export of IFN to the environment</td>
<td></td>
</tr>
<tr>
<td>$k_{STATP,IFNe}$</td>
<td>675.323</td>
<td>μM hours⁻¹</td>
<td>Max rate of $[STAT]$ phosphorylation via $[IFNe]$</td>
<td></td>
</tr>
<tr>
<td>$K_{STATP,IFNe}$</td>
<td>80.353</td>
<td>μM</td>
<td>Michaelis-Menten constant for $[STAT]$ phosphorylation via $[IFNe]$</td>
<td></td>
</tr>
<tr>
<td>$\tau_{STATP}$</td>
<td>0.3</td>
<td>hours⁻¹</td>
<td>Dephosphorylation rate of $[STATP]$</td>
<td>52</td>
</tr>
<tr>
<td>$k_{IRF7,STATP}$</td>
<td>0.115</td>
<td></td>
<td>Rate of IRF7 induction</td>
<td>84</td>
</tr>
<tr>
<td>$k_{IRF7,IRF7P}$</td>
<td>1.053</td>
<td></td>
<td>Rate of IRF7 induction</td>
<td></td>
</tr>
<tr>
<td>$\tau_{IRF7}$</td>
<td>0.75</td>
<td></td>
<td>Decay rate of IRF7</td>
<td>53,84</td>
</tr>
<tr>
<td>$k_{IRF7P,IRF7}$</td>
<td>0.202</td>
<td></td>
<td>Rate of IRF7 phosphorylation</td>
<td>84</td>
</tr>
<tr>
<td>$\tau_{IRF7P}$</td>
<td>0.3</td>
<td></td>
<td>Dephosphorylation rate of $[IRF7P]$</td>
<td>54</td>
</tr>
<tr>
<td>$\gamma_{HV}$</td>
<td>0.635</td>
<td></td>
<td>Rate of cell health loss</td>
<td>84</td>
</tr>
<tr>
<td>$k_{V,V}$</td>
<td>1.537</td>
<td></td>
<td>Rate of viral replication</td>
<td>72,84</td>
</tr>
</tbody>
</table>
Table 8 (continued).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
<th>Process</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{V,\text{IFNe}}$</td>
<td>0.020884</td>
<td>μM</td>
<td>$[\text{IFNe}]$ Michaelis-Menten constant</td>
<td>84</td>
</tr>
<tr>
<td>$Q_V$</td>
<td>0.197</td>
<td>PFU mL$^{-1}$ hours$^{-1}$</td>
<td>Rate of viral release</td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>3</td>
<td>[unitless]</td>
<td>Hill coefficient of TLR</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>1E3</td>
<td>mL PFU$^{-1}$ hours$^{-1}$</td>
<td>Uninfected to eclipse phase cells</td>
<td>85</td>
</tr>
<tr>
<td>$\tau_{II}$</td>
<td>0.167</td>
<td>hours$^{-1}$</td>
<td>Eclipse phase to late infected cells</td>
<td>(25)</td>
</tr>
<tr>
<td>$\tau_{Ve}$</td>
<td>0.542</td>
<td>hours$^{-1}$</td>
<td>Rate of extracellular virus decay</td>
<td></td>
</tr>
<tr>
<td>$\tau_{IFNe}$</td>
<td>3.481</td>
<td></td>
<td>Rate of extracellular $[\text{IFNe}]$ decay</td>
<td>84</td>
</tr>
<tr>
<td>$D_{Ve}$</td>
<td>54.0</td>
<td>μm$^2$ s$^{-1}$</td>
<td>Diffusion coefficient of $[V_e]$</td>
<td>71,72</td>
</tr>
<tr>
<td>$D_{IFNe}$</td>
<td>54.0</td>
<td>μm$^2$ s$^{-1}$</td>
<td>Diffusion coefficient of $[\text{IFNe}]$</td>
<td>154,155</td>
</tr>
<tr>
<td>$L_{Ve}$</td>
<td>0.09</td>
<td>μm</td>
<td>Diffusion length of virus</td>
<td></td>
</tr>
<tr>
<td>$L_{IFNe}$</td>
<td>0.23</td>
<td></td>
<td>Diffusion length of $[\text{IFNe}]$</td>
<td></td>
</tr>
<tr>
<td>Voxel Width</td>
<td>3</td>
<td></td>
<td>Width of lattice voxels</td>
<td>76,*2</td>
</tr>
<tr>
<td>Cell Size</td>
<td>9</td>
<td></td>
<td>Width of cells</td>
<td></td>
</tr>
</tbody>
</table>

*2 Specific to the CompuCell3D’s implementation of the simulation.
Appendix B.6 MSIS Model interferon prestimulation

Figure 32. Simulated [IFNe] exposure before infection protects cells from plaque formation.

The plate was washed with extracellular Type-I interferons, then a single cell was infected at the center of the plate. Since cell health is the median of all live cells’ health, the initially infected cell dying ~16 hours caused a brief increase in median cell health. No interferon-triggered death mechanism or resource limitations are present, leading to boundless amplification of the cytokine signal after the virus has been cleared. Bold lines are median of 20 replicas; shaded areas represent the 5th and 95th percentiles.
Appendix B.7 MSIS model local sensitivity analyses

Figure 33. Local sensitivity analysis of baseline simulation.

Down and up columns give average value change for each of the metrics when the parameter is varied -25% (down) and +25% (up) of their baseline value. These changes are shaded.
red for positive changes and blue for negative changes in the metric, with intensity normalized to the largest change within both columns of each metric. p-values are the statistical significance of the change, given the standard deviation of the stochastic simulations over 20 replicas. p-values < 0.01 are highlighted in yellow. Note that $\tau_{IRF7}$ has a large response in baseline [IFNe] Max because the baseline value for $\tau_{IRF7}$ lies near the stability criterion of $\tau_{IRF7} > 0.75$, so the 25% decrease leads to a numerically unstable system.
Figure 34. Local sensitivity analysis with elevated paracrine signaling.

This case corresponds to a 15x increase in the phosphorylation rate of [STATP] via the JAK/STAT pathway (parameter $k_{STATP,IFNe}$. Value changed from baseline of 45.922 μM hours$^{-1}$ to 688.83 μM hours$^{-1}$). Down and up columns give average value change for each of the metrics when the parameter is varied -25% (down) and +25% (up) of their baseline value. These changes are
shaded red for positive changes and blue for negative changes in the metric, with intensity normalized to the largest change within both columns of each metric. p-values are the statistical significance of the change, given the standard deviation of the stochastic simulations over 20 replicas. p-values < 0.01 are highlighted in yellow.
Figure 35. Local sensitivity with an elevated interferon diffusion coefficient (15x baseline or 540 μm² s⁻¹).

Plaques are arrested by the paracrine signal diffusion significantly faster than viral spread.

Down and up columns give average value change for each of the metrics when the parameter is
varied -25% (down) and +25% (up) of their baseline value. These changes are shaded red for positive changes and blue for negative changes in the metric, with intensity normalized to the largest change within both columns of each metric. p-values are the statistical significance of the change, given the standard deviation of the stochastic simulations over 20 replicas. p-values < 0.01 are highlighted in yellow.

Appendix B.8 MSIS model replicate justification

Figure 36. Standard deviation and standard error of simulations versus replicas for the baseline.

Used to justify n = 20 replicas for sensitivity analyses and parameter sweeps. Standard deviation (SD) is blue while standard error (SE) is orange.
Figure 37. A parameter sweep of $\beta$ (rate of transition from uninfected (U) to eclipse phase (I1) cells) from 0.01x to 100x baseline reveals steady growth increases.

Plaques still form at any nonzero value. A. Plaque growth over 80 hours post-infection. B. Tracking of cell types (U, I1, virus releasing, I2, and dead, D) for plaque growth dynamics corresponding to plaques in A. Center lines represent median over 20 replicas; shaded areas are the 5th and 95th percentiles. C – E. $\beta$ parameter multipliers versus growth rate of a single plaque at the end of the simulation at 80 hours (C) and the area under the curve (AUC) for both average $[V_e]$ (D) and average $[IFNe]$ (E) on log scales. C – E represent 20 simulation replicas. Full data
with 14 additional outliers for average extracellular interferon and 1 additional outlier for the average extracellular virus are available in Appendix B.10, Figures 42 and 43, respectively. Higher virus infectivity resulted in higher proportions of dead cells within the plaque. Note a non-mono
monic trend; natural virus infectivity leads to a minimum production of $[\text{IFN}_e]$. Decreases and increases in $\beta$ both led to higher $[\text{IFN}_e]$ production. Viruses have differing encapsulation proteins, genome sizes, and relative production of nonstructural proteins while replicating within a host cell. These differences lead to variable virus replication rates, represented in the model by $k_{V,V}$.

**Figure 38.** Parameter sweep of viral replication rate ($k_{V,V}$) reveals dramatic changes to final plaque diameter.
A. Plaque growth over 80 hours post-infection. B. Tracking of cell types (uninfected, U, eclipse infected, I1, virus releasing, I2, and dead, D) for plaque growth dynamics corresponding to plaques in A. Center lines represent median over 20 replicas; shaded areas are the 5th and 95th percentiles. C-E. Viral replication rate, $k_{V,V}$, multipliers versus growth rate of a single plaque at the end of the simulation at 80 hours (C) and the area under the curve (AUC) for both average $[Ve]$ (D) and average $[IFNe]$ (E, full data with 4 additional outliers available in Appendix B.10, Figure 39) on log scales. C – E represent 20 simulation replicas. Lowering viral replication below the nonspecific viral clearance rate prevents plaque formation. Higher replications have exponential changes in system metrics. Lowering viral replication speed slows, and can even prevent plaque growth, as virus is cleared from the extracellular environment more quickly relative to the rate of production and release. This limits the size of plaques in vitro and lesion size in vivo. Higher viral replication leads to exponentially faster growth and larger lesions since virus replication is self-amplifying. Viral titer growth follows an exponential growth curve; however, the radial growth of the plaques is linear. These replicate biological observations.
Figure 39. Elevated IFN diffusion coefficient ($D_{IFNe}$) leads to plaque arrest.

A. Plaque growth over 80 hours post-infection. B. Tracking of cell types (uninfected, U, eclipse infected, I1, virus releasing, I2, and dead, D) for plaque growth dynamics corresponding to plaques in A. Center lines represent median over 20 replicas; shaded areas are the 5th and 95th percentiles. C – E. Extracellular interferon diffusion, $D_{IFNe}$, parameter multipliers versus growth rate of a single plaque at the end of the simulation at 80 hours (C.) and the area under the curve (AUC) for both average $[V_e]$ (D, full data with any present outliers available in Appendix B.10) and average $[IFNe]$ (E) on log scales. C – E represent 20 simulation replicas. Plaque growth loses linearity and is arrested after 10x increase over baseline.
Appendix B.10  MSIS Model Parameter Sweep Outliers

Figure 40. Figure 13 Plaque Growth Rate with all outliers visible.

Five outlier simulations of increased $k_{IFN,V}$ (RIG-I) activity over baseline resulted in fully arrested plaques by 80 hours post-infection. These outliers were cropped out in the original figure to show the distribution of the remaining 275 data points more clearly. Data represents 20 simulation replicas.
The same five simulations which resulted in fully arrested plaques also result in dramatically lower average $[V_e]$ AUC. These outliers were cropped out in the original figure to show the distribution of the remaining 275 data points more clearly. Data represents 20 simulation replicas.
14 outlier simulations resulted in significantly higher average $[IFN_e]$ AUC. These outliers were cropped out in the original figure to show the distribution of the remaining 266 data points more clearly. Data represents 20 simulation replicas.

A single outlier simulation resulted in a much lower average $[V_e]$ AUC. This outlier was cropped out in the original figure to show the distribution of the remaining 279 data points more clearly. Data represents 20 simulation replicas.
Figure 44. Figure 39 IFNe AUC with all outliers visible.

4 outlier simulations resulted in much higher average \([IFN_e]\) AUC. These outliers were cropped out in the original figure to show the distribution of the remaining 276 data points more clearly. Data represents 20 simulation replicas.

Figure 45. Figure 39 Ve AUC with all outliers visible.
A single outlier simulation resulted in a much lower average $[ \bar{V}_c ]$ AUC. This outlier was cropped out in the original figure to show the distribution of the remaining 279 data points more clearly. Data represents 20 simulation replicas.

**Appendix B.11 Model and Raw Data Repository**

The code to run the model in CC3D and the raw data generated for all simulations discussed in this study are available at [https://github.com/ImmuSystems-Lab/Multicellular_Spatial_Model_of_RNA_Virus_Replication](https://github.com/ImmuSystems-Lab/Multicellular_Spatial_Model_of_RNA_Virus_Replication). All data is generated by the MSIS model’s code and has the same format.
Appendix C Macrophage Models

Appendix C.1 Model Equations

A three-state ODE model, referred to as Model 1, was developed using the immunological knowledge of the early innate immune response to a primary influenza A virus infection (i.e., the animal’s first exposure, no antibodies for the virus present) described above. The model contained in Equations C.1-1 through C.1-3 is illustrated in Aim 3. The units of each state (V, I, and M) are discussed in below.

\[
\frac{d[V]}{dt} = r_{V,V}[V] \left(1 - \frac{[V]}{K_{V,V}}\right) - r_{V,I}[V][I] - r_{V,M}[V][M] - d_{V}[V] \quad \text{(C.1-1)}
\]

\[
\frac{d[I]}{dt} = r_{I,V}[V] + r_{I,M}[M] - d_{I}[I] \quad \text{(C.1-2)}
\]

\[
\frac{d[M]}{dt} = \frac{r_{M,I}[I]^n}{K_{M,I}\left(1+[I]\right)^n} - d_{M}([M]) \quad \text{(C.1-3)}
\]

Virus production is modeled in Equation C.1-1. \([V]\), the concentration of virus in lung tissue, is modeled as logistic growth with a constant of proportionality, \(r_{V,V}\), and a carrying capacity, \(K_{V,V}\). This form of virus production was selected over target cell-based modeling approaches because data concerning the number of available target cells in the lung is not available, limiting the viability and accuracy of training a model. The effect of interferon-regulated inhibition of virus replication is modeled using mass action kinetics where \(r_{V,I}\) is the corresponding rate constant. The inhibition of virus production via macrophage is also modeled with mass-action kinetics where \(r_{V,M}\) is the rate constant. Virus degrades at a rate, \(d_{V}\).
Type-I interferon production is modeled by Equation C.1-2 where \( I \) is the concentration of interferon in the lung. Interferon is produced at a rate, \( r_{I,V} \), relative to viral load and decays at rate \( d_I \). Upregulation of interferon production via macrophages was modeled as a first-order mass-action kinetic with a rate, \( r_{I,M} \).

Macrophage production is modeled in Equation C.1-3 where \([M]\) is the number of macrophages in the lung. Interferon induction of macrophage production is modeled using a Hill kinetic with a production rate, \( r_{M,I} \), and an apparent dissociation constant, \( K_{M,I} \). Instead of the classic interpretation of the Hill coefficient, \( n \), as cooperativity in ligand binding\(^{156}\), it can be interpreted in this context as an activation threshold representing the threshold of interferon needed to induce macrophage production. This is like the activation threshold that must be exceeded to induce T cell cytokine production\(^{157,158}\). The parameter \( K_{M,I} \) is not raised to the Hill-like coefficient, \( n \), to improve parameter fitting. Macrophage decays at a rate of \( d_M \).

**Appendix C.2 Alternative Model Structures**

**Figure 46.** A) Model schemes of the four models considered in this work. B) Model structure variations.
\([V], [I], \) and \([M]\) represent virus concentration, interferon concentration, and macrophage cell count in the lungs of infected mice. Arrows represent activating interactions; lines ending in crosses represent inhibiting interactions. The parameters involved in each interaction are indicated in Model 1 (degradation reactions not shown). Model 1 is the fully connected model of the innate immune response model. Models 2 – 4 are reduced versions of Model 1, wherein select interactions were removed. (B) Each model is analyzed for its goodness of fit to experimental data under three different scenarios. Schemes of the model emphasize the different outcomes that occur under each scenario. Black arrows indicate parameters that retain the same value when fitting the model to H5N1 and H1N1 infection data. Red, broken arrows identify parameters that take on different values when training two copies of a model to the H5N1 and H1N1 infection data.

Equations C.1-1 through C.1-3 define the dynamic behavior of Model 1. We also developed reduced models, Models 2 – 4 (Figure 46), in which select interactions were removed to consider additional hypotheses on how the immune system in the lung might be regulated. For example, in CCR2-/- mice, there is conflicting evidence concerning whether inhibited macrophage infiltration into the lung of infected mice affects viral load\[^{146,147}\]. In addition, although macrophage upregulation of interferon is well justified, it is not guaranteed that parameters associated with this interaction can be estimated from the data. In Model 2, \([M]\) induction of \([I]\) is removed. In Model 3, \([M]\) inhibition of \([V]\) is removed. And in Model 4, both \([M]\) induction of \([I]\) and \([M]\) inhibition of \([V]\) are removed. These models were each fit to the experimental data to determine which model (and, therefore, which combination of biological processes) optimally fits the data based on the goodness of fit and the number of parameters estimated (DOF).
Figure 47. Macrophage and MCP1 correlation.

Log_{10} of macrophage cell count in the lung is highly correlated with the log_{2} of MCP1 gene expression. This regression is used to translate between MCP1 and Macrophage states.

\[
\log_{10}(M) = 0.6301 \log_{2}(MCP1)
\]
Appendix C.4 Energy Traces of Macrophage Model 4 MCMC Parameter Exploration

Figure 48. AD and NSSD model 4 energy plots.

A) H1N1 energy trace under the All Different (AD) condition. B) H5N1 energy trace under the same conditions. C) Combined energy trace for H1N1 and H5N1 under the All Shared (AS) condition. MCMC quickly completes burn-in and thoroughly explored parameter space for all three scenarios.
Appendix C.5 Macrophage Model Structures 2 and 3 Data Fits

Macrophage-dependent feedback mechanisms (Models 2 and 3 NSSD) exhibit worse fits than a model structure without these mechanisms (Model 4). Macrophage-based clearance of Virus ($r_{V,M}$, blue line) fits H5N1 data poorly. Data from Shoemaker et al are shown with the standard error associated with triplicate data points per timepoint.

Figure 49. Models 2 and 3 predictions.
Appendix C.6 Macrophage Model Structure #4 Optimal Parameters

The lowest energy model parameterization (Model 4, all parameters independently fit to each dataset) is summarized in Table 9.

Table 9. AD Minimum energy parameter values and units for each dataset in Model 4.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>H1N1</th>
<th>H5N1</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{V,V}$</td>
<td>1.22E+00</td>
<td>1.21E+00</td>
<td>$days^{-1}$</td>
</tr>
<tr>
<td>$K_{V,V}$</td>
<td>3.65E+01</td>
<td>7.80E+02</td>
<td>$log_{10}(PFU/mg)$</td>
</tr>
<tr>
<td>$r_{V,I}$</td>
<td>1.20E-01</td>
<td>1.07E-01</td>
<td>$days^{-1}$</td>
</tr>
<tr>
<td>$d_{V}$</td>
<td>1.61E-01</td>
<td>1.10E-05</td>
<td>$days^{-1}$</td>
</tr>
<tr>
<td>$r_{I,V}$</td>
<td>7.70E-01</td>
<td>3.06E+00</td>
<td>$[log_{10}(PFU/mg)\ hours]^{-1}$</td>
</tr>
<tr>
<td>$d_{I}$</td>
<td>9.59E-01</td>
<td>3.22E+00</td>
<td>$days^{-1}$</td>
</tr>
<tr>
<td>$r_{M,I}$</td>
<td>2.16E+07</td>
<td>9.71E+03</td>
<td>$Macrophage\ Cell\ Count/days$</td>
</tr>
<tr>
<td>$K_{M,I}$</td>
<td>1.90E+05</td>
<td>1.04E+09</td>
<td>unitless</td>
</tr>
<tr>
<td>$d_{M}$</td>
<td>8.80E+03</td>
<td>6.18E-01</td>
<td>$days^{-1}$</td>
</tr>
<tr>
<td>$n$</td>
<td>5.47E+00</td>
<td>9.98E+00</td>
<td>unitless</td>
</tr>
<tr>
<td>$r_{V,M}$</td>
<td>N/A</td>
<td></td>
<td>$[Macrophage\ Cell\ Count\ days]^{-1}$</td>
</tr>
<tr>
<td>$r_{I,M}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C.7 Macrophage Model 4 OSSD Parameter Distributions

This appendix contains AD, NSSD, and OSSD model 4 parameter posterior density distributions. Overlapping distributions with an OSSD parameter ($r_{M,I}$, etc.) likely indicate non-strain-specific mechanisms are likely, while distinct distributions ($d_i$, etc.) likely indicate strain-specific differences likely exist.

Figure 50. Macrophage Model 4 AD and NSSD Parameter Distributions.
Figure 51. Macrophage Model 4 with unshared $d_I$.

Figure 52. Macrophage Model 4 with unshared $d_M$. 
Figure 53. Macrophage Model 4 with unshared $d_V$.

Figure 54. Macrophage Model 4 with unshared $K_{M,I}$. 
Figure 55. Macrophage Model 4 with unshared $K_{V,V}$.

Figure 56. Macrophage Model 4 with unshared $n$. 
Figure 57. Macrophage Model 4 with unshared $r_{M,I}$.

Figure 58. Macrophage Model 4 with unshared $r_{V,I}$.
Figure 59. Macrophage Model 4 with unshared $r_{V,V}$. 
Appendix D SPOT Supplement

Appendix D.1 SPOT Methodology

**SPOT Value:** the result of a multiple categorical linear regression between the genome (Shared or Unshared categories, 0 or 1, for each parameter in the genome) and the BIC which results from each genome. Multiple categorical linear regression fits slopes, $k_n$’s, using Ordinary Least Squares (Equation D.1-1) and the BICs which result from evaluating the chosen population of genomes.

$$BIC \text{ predictions from parameterization} = \sum_{n=0}^{parameters} k_n G_n + BIC_{int} \quad \text{(D.1-1)}$$

The SPOT value for a parameter $n$ is given by Equation D.1-2. This value is the slope of the categorical regression related to the parameter of interest. Confidence intervals for SPOTs are computed according to Equation D.1-3.

$$SPOT_n = \frac{k_n}{ln (S_{BIC})} \quad \text{(D.1-2)}$$

$$SPOT CI_n = OLS CI_n / ln (S_{BIC}) \quad \text{(D.1-3)}$$

$k_n = \text{slope of categorical linear regression for parameter(s) } n$

$G_n = \text{Genome value for parameter(s) } n (0 \text{ or } 1 \text{ for shared or different, respectively})$

$BIC_{int} = \text{predicted BIC under the AS Condition. This is the intercept of the multiple categorical regression.}$

$S_{BIC} = \text{BIC Scaling Factor, the total number of datapoints present across all fit datasets.}$

This normalizes SPOT results into four categories with fixed values, regardless of the underlying model and data. The scale factor is a result of BIC’s penalty for increased degrees of freedom.
Appendix D.2 Model 1 Characteristic Curves

Figure 60. Characteristic Curves for Model 1.
Appendix D.3 Model 4 Characteristic Curves

Figure 61. Model 4 Characteristic Curves.
Appendix D.4 Model 1 Parameter Sensitivity Sweeps

Figure 62. Model H5N1:H1N1 Parameter Ratio Sweeps.

Note that $r_{V,M}$ and $r_{L,M}$ were not evaluated due to time limitations.
Appendix D.5 Model 4 Parameter Sensitivity Sweeps

Figure 63. Model 4 H5N1:H1N1 Parameter Ratio Sweeps.


92. CDC. Highly Pathogenic Asian Avian Influenza A(H5N1) in People | Avian Influenza (Flu).


